



# A survey on learning from graphs with heterophily: recent advances and future directions

Cheng-Hua GONG<sup>1</sup>, Yao CHENG<sup>1</sup>, Jian-Xiang YU<sup>1</sup>, Can XU<sup>1</sup>, Cai-Hua SHAN<sup>2</sup>, Si-Qiang LUO<sup>3</sup>, Xiang LI<sup>1</sup>✉

1. School of Data Science and Engineering, East China Normal University, Shanghai 200062, China

2. Microsoft Research Asia, Shanghai 200062, China

3. School of Computer Science and Engineering, Nanyang Technological University, Singapore 999002, Singapore

Received September 30, 2024; accepted March 18, 2025

E-mail: [xiangli@dase.ecnu.edu.cn](mailto:xiangli@dase.ecnu.edu.cn)

© The Author(s) 2025. This article is published with open access at [link.springer.com](http://link.springer.com) and [journal.hep.com.cn](http://journal.hep.com.cn)

## Abstract

Graphs are structured data that models complex relations between real-world entities. Heterophilic graphs, where linked nodes trend to have different labels or dissimilar features, have recently attracted significant attention and found many real-world applications. Meanwhile, increasing efforts have been made to advance learning from graphs with heterophily. Various graph heterophily measures, benchmark datasets, and learning paradigms are emerging rapidly. In this survey, we comprehensively review existing works on learning from graphs with heterophily. First, we overview over 500 publications, of which more than 300 are directly related to heterophilic graphs. After that, we survey existing metrics of graph heterophily and list recent benchmark datasets. Further, we systematically categorize existing methods based on a hierarchical taxonomy including GNN models, learning paradigms and practical applications. In addition, broader topics related to graph heterophily are also included. Finally, we discuss the primary challenges of existing studies and highlight promising avenues for future research.

## Keywords

graphs with heterophily; heterophilic graphs; graph neural networks; graph learning

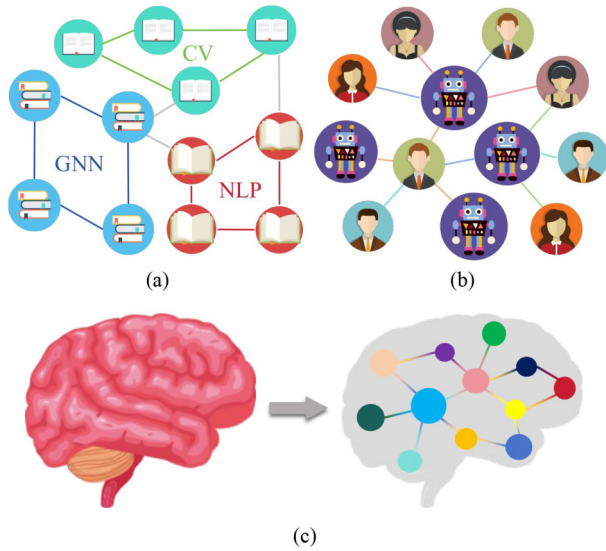
## 1 Introduction

Graph-structured data is ubiquitous in practical scenarios, which models entities as nodes and the complex relationships between entities as edges. Some graphs exhibit homophily [1], where linked nodes tend to have the same label or similar features, such as citation networks and friendship networks. As shown in Fig. 1(a), citation patterns show typical homophily, as papers tend to cite works within the same domain. In other cases, there also exist graphs with heterophily [2], where nodes with different labels or dissimilar features are more likely to be connected.

Graphs with heterophily have found numerous practical applications [3–5], underscoring the importance of this research area. For example, social bots have been extensively utilized to disseminate misinformation, instigate panic, and even influence elections, leading to significant negative social consequences [6–10]. In Fig. 1(b), we present a social network with automated bots, where bots tend to establish connections with users instead of other bots. Due to significant differences in characteristics and behaviors between bots and users, this network exhibits the typical graph heterophily. Moreover, the human brain can be conceptualized as a complex network, as depicted in Fig. 1(c), where different regions

serve as nodes and the connections between them are represented as edges. Given that various regions of brain support specific functions, each region displays unique structure and feature [11]. Therefore, the brain network is far from being homophilic but rather heterophilic [12]. Moving to the urban computing [13], the city is usually modeled as an urban graph where nodes are objects such as functional regions and edges are physical or social dependencies such as human mobility and traffic flow. Taking urban graphs constructed with human mobility as an example, heterophily usually exists as end nodes of an edge could be of different functionalities, such as residential area and workplace [14]. In summary, the heterophily inherent in graphs is widespread across numerous scenarios and is closely related to our human bodies, daily lives, and the environments we inhabit.

Recent advancements have seen the proposal of numerous Graph Neural Networks (GNNs) [15–17], which have achieved remarkable success in processing graph-structured data. Traditional GNNs implicitly assume that graphs are homophilic and follow the message passing mechanism [18,19], where each node updates the representation by aggregating messages from neighbors. However, this mechanism struggles with graph heterophily, where nodes may



**Fig. 1** The toy examples of homophilic and heterophilic graphs in real-world applications. (a) A citation network; (b) a social network with bots; (c) a brain network with community structure

incorrectly aggregate information from dissimilar neighbors, leading to suboptimal performance [3,5,20]. More precisely, message passing mechanism fails to discriminate uninformative local nodes and explore informative global nodes under a heterophilic setting. Simply aggregating neighborhood information without discrimination can easily introduce noise, resulting in indistinguishable representations. Moreover, message passing mechanism is naturally constrained to local topology and fails to reach distant but informative nodes. To dive into graph heterophily, learning topics such as metrics, benchmarks, models and learning paradigms have emerged recently. Due to the growing interest, we have recently witnessed some relevant surveys on this topic [3–5]. Early surveys [3,4] are limited in scope, focusing primarily on GNN models while neglecting other learning topics such as learning paradigms and practical applications. The recent handbook [5] compiles existing works related to heterophilic graphs, but it simply lists for easy reference but lacks a systematic categorization. Therefore, there is an urgent need for a more comprehensive and systematic review in this field.

In this survey, we start with the metrics of graph heterophily and benchmarks, elaborating on the sources and characteristics of each dataset. Next, we categorize existing heterophilic GNN models based on their architectures and underlying mechanisms, focusing primarily on GNN models and other advanced frameworks designed to handle graph heterophily. After that, we introduce recent advances on self-supervised learning and prompt learning in this field, and also provide more related topics to broaden the research scopes. Finally, we summarize the related practical applications, and give an outlook on the future development of this field. Our main contributions can be summarized as follows:

- To our best knowledge, this survey is currently the most comprehensive one in the area of heterophilic graph learning.
- This survey introduces a systematic taxonomy for learning

from graphs with heterophily and categorizes existing works from diverse aspects.

- This survey reveals the current challenges faced by existing works related to heterophilic graph learning and present promising directions and future insights.

## 2 Preliminaries

### 2.1 Notations

Let  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  denote a graph with a set of nodes  $\mathcal{V}$  and a set of edges  $\mathcal{E}$ , where  $N = |\mathcal{V}|$  is the number of nodes. The adjacency matrix of  $\mathcal{G}$  is denoted as  $\mathbf{A} = [a_{ij}] \in \{0, 1\}^{N \times N}$ , where  $a_{ij} = 1$  if there exists an edge  $e_{ij} = (v_i, v_j)$ . The degree matrix  $\mathbf{D}$  is a diagonal matrix with each diagonal element  $d_i = \sum_{j=1}^N a_{ij}$  being the degree of node  $v_i$ . The neighbor set of node  $v$  is denoted as  $\mathcal{N}(v) = \{v_j : (v_i, v_j) \in \mathcal{E}\}$ . The node feature matrix is denoted as  $\mathbf{X}$ , where the  $i$ th row  $x_i$  is the feature vector of node  $v_i$ . The node representation matrix is denoted by  $\mathbf{H}$ , where the  $h_i$  is the representation of node  $v_i$ . The label matrix is denoted by  $\mathbf{Y}$ , where the  $i$ th row  $y_i$  is the one-hot encoded label vector of  $v_i$ . For nodes  $v_i, v_j \in \mathcal{V}$ , if  $y_i = y_j$ , they are viewed as intra-class nodes; otherwise, they are inter-class nodes. Equally, an edge  $e_{ij} \in \mathcal{E}$  is taken as an intra-class edge if  $y_i = y_j$ , or an inter-class edge if  $y_i \neq y_j$ .

### 2.2 Message passing framework

The message passing framework [18,19] stands for a broad category of GNNs, where each node aggregates information from neighbors and then combining the aggregated information. The process can be formulated as:

$$h_i^{(l)} = \text{COM}\left(h_i^{(l-1)}, \text{AGG}\{h_j^{(l-1)} : v_j \in \mathcal{N}(v_i)\}\right), \quad (1)$$

where  $0 \leq l \leq L$  and  $L$  is the number of GNN layers,  $h_i^{(0)} = x_i$  and  $h_i^{(l)} (1 \leq l \leq L)$  denotes the node representation of  $v_i$  at the  $l$ th layer. The choice of aggregation  $\text{AGG}(\cdot)$  is flexible (e.g. mean, sum, max pooling), and the combination  $\text{COM}(\cdot)$  in each layer can also be customized. Message-passing GNNs are simple yet powerful, but they still encounter challenges such as over-smoothing [21], over-squashing [22], limited propagation [23], and the heterophily issue [20].

### 2.3 Graph transformer

Recently, Transformer [24] has rapidly advanced and revolutionized the field of NLP [25–27] and CV [28–31]. Inspired by that, Graph Transformers (GTs) has emerged as a prominent approach in graph learning [32,33]. Here, we briefly introduce the core components: self-attention mechanism, positional, and structural encodings.

#### 2.3.1 Self-attention mechanism

Self-attention is the fundamental mechanism of GTs. Given the node feature matrix  $\mathbf{X}$ , a GT layer first projects  $\mathbf{X}$  into the Query, Key, and Value matrices:

$$\mathbf{Q} = \mathbf{XW}_Q, \mathbf{K} = \mathbf{XW}_K, \mathbf{V} = \mathbf{XW}_V,$$

where  $\mathbf{W}_Q, \mathbf{W}_K, \mathbf{W}_V$  are three trainable weight matrices. After that, the node representation  $\mathbf{H}$  with a single self-attention head is

computed as follows:

$$S = \frac{QK^T}{\sqrt{d}}, H = \text{softmax}(S)V,$$

where  $d$  denotes the dimensionality of attention head, and  $S \in \mathbb{R}^{N \times N}$  denotes the self-attention matrix. Subsequently, multi-head attention can be implemented by concatenation, followed by MLP module, residual connection, and normalization [24]. The main difference between GTs and message-passing GNNs is that GTs disregard the topology, treating the input graphs as fully-connected.

### 2.3.2 Positional and structural encodings

To incorporate graph topology into GTs, Positional Encoding (PE) and Structural Encoding (SE) have been proposed. SEs focus on topological information on local, relative, or global levels, such as node degree, triangle, cycle counting, and subgraph context. More advanced methods include DSE [34], RWSE [35], and TCSE [36]. Different from SEs, PEs perceives the relative positions towards other nodes and absolute positions within the graph. For example, LapPE [37], RWPE [38], JaccardPE [39] are all typical PEs.

### 2.4 Graph laplacian and filters

Graph Laplacian [40] is defined as  $L = D - A$ , and its normalized version is  $\tilde{L} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ . Its eigen-decomposition gives  $L = U\Lambda U^T$ , where  $U$  is the eigenvector matrix, also called graph fourier basis. Given a graph signal  $x$ , the graph fourier transform is defined as  $\hat{x} = U^T x$ , and the inverse transform is  $x = U^T \hat{x}$ . The eigenvalue matrix  $\Lambda = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$  with  $0 \leq \lambda_1 \leq \lambda_2, \dots, \lambda_N \leq 2$  and  $\lambda_i$  is frequency [41]. Here, smaller  $\lambda_i$  corresponds to low-frequency signals (smooth information), while larger  $\lambda_i$  corresponds to high-frequency signals (non-smooth information) [42]. According to spectral graph theory, graph convolution can be expressed as:

$$g * x = U g(\Lambda) U^T = \sum_{i=0}^N g(\lambda_i) u_i u_i^T x,$$

where  $g(\Lambda) = \text{Diag}(g(\lambda_1), \dots, g(\lambda_N))$  is the spectral graph filter to re-weight frequencies.

In practice, performing eigen-decomposition on large-scale graphs is infeasible in terms of time complexity. Hence, it is common to utilize the polynomial approximation with regard to  $\tilde{L}$  as an approximate filter  $g(\tilde{L})$  [43]:

$$g * x = U g(\Lambda) U^T \approx g(\tilde{L}) \cdot x.$$

Such method is typically considered for smooth signals, while non-smooth signals are more important under graph heterophily. Therefore, various graph filters are designed to capture complex patterns on graphs. The common low-pass filter  $F_{LP}$  is built based on the affinity matrix  $F_{LP} = \tilde{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ , while the corresponding high-pass filter is  $F_{HP} = \tilde{L} = I - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$  [41].

### 2.5 Learning paradigms

Currently, mainstream graph learning paradigms include supervised

learning and self-supervised learning. With the development of supervised learning and self-supervised pre-training techniques, prompt learning has recently emerged which integrates the advantages of both supervised and self-supervised paradigms.

- Supervised learning. Based on supervision signals, some effective end-to-end GNNs have been designed. Given a GNN model, node classification is semi-supervised as it uses unlabeled nodes from the test set [15], while graph classification is supervised since the test set is ignored [44]. In this paper, we simplify by treating both as supervised learning for convenience.
- Self-supervised learning. Despite the remarkable success of supervised learning, the heavy reliance on supervisions brings high annotation cost, weak model robustness, and the overfitting problem. To this end, self-supervised learning [45,46] on graphs designs a series of pretext tasks, leverages the input itself as supervision to learn informative representations from unlabeled data.
- Prompt learning. Built on self-supervised pre-training techniques [47], prompt learning is primarily applied to the downstream tuning of pre-trained models in NLP [48]. The target is to bridge the gap between pretexts and downstream tasks through a unified task template, thereby fully leveraging the pre-trained model. Free from the costly fine-tuning, prompt learning can achieve great results by efficient tuning, even when the supervision of specific task is limited. By fully leveraging the self-supervised pre-training and requiring minimal supervision signals, prompt learning can be seen as integrating the advantages of both supervised and self-supervised learning.

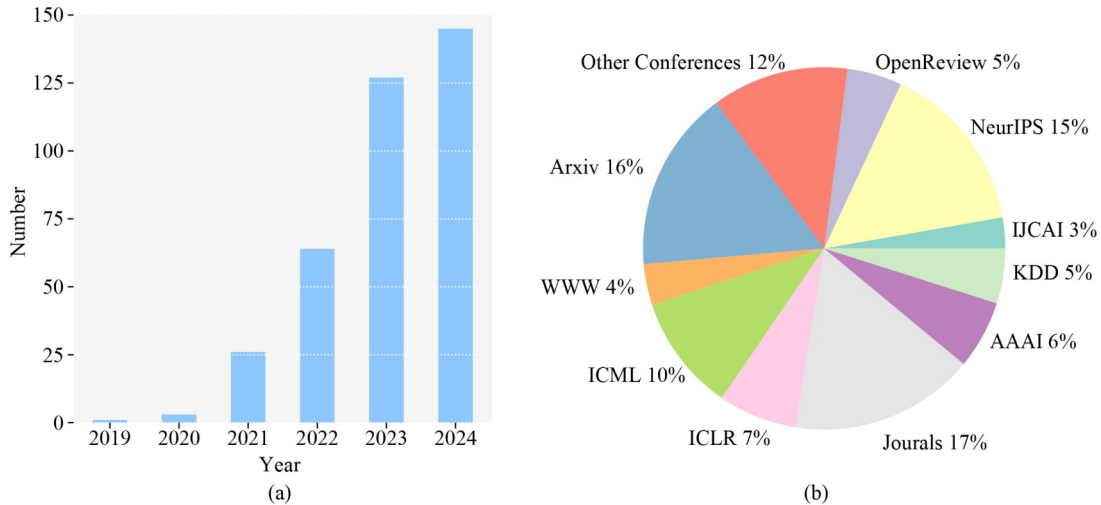
## 3 Overview

This section provides a brief literature overview, followed by the organizational structure of this survey.

### 3.1 Literature overview

In this survey, we collect over 500 papers, of which more than 300 are directly related to graph heterophily, with particular emphasis on those published in top journals or famous conferences. Here, journals include TPAMI, TKDE, TMLR, TNNLS, and Neural Networks, while conferences involve ICML, ICLR, NeurIPS, KDD, AAAI, among others. We also include papers that feature novel topics or have garnered widespread attention on OpenReview and Arxiv. All the resources related to this survey are presented in our GitHub repository<sup>1)</sup>. In Fig. 2, we present the statistics of collected papers. First, we compile the annual publication statistics of papers over the past six years. As can be seen in Fig. 2(a), the number of papers released per year shows a significant growth trend, indicating the enormous potential of this topic. Moreover, the source distribution is given in Fig. 2(b). It is worth noting that more than half of collected papers have been published in top journals or conferences, indicating

<sup>1)</sup> <https://github.com/gongchenghua/Papers-Graphs-with-Heterophily>.



**Fig. 2** The statistics of collected papers related to learning from graphs with heterophily. (a) Number of papers published per year; (b) the distribution of sources for the collected papers

the reliability. In Fig. 3, we analyze the general content and present the most frequently words in titles. Notably, the keywords are closely related to the topic, which center around learning from graphs with heterophily.

### 3.2 Organizational structure

We present the organizational structure of this survey in Fig. 4. In Section 2, we provide notions and preliminaries, and present an overview in Section 3. In Section 4, we introduce metrics of graph heterophily and provide a detailed description of benchmarks. In Section 5, we summarize representative GNN models and beyond, and provide a detailed grouping of them. Apart from supervised learning, we also introduce other popular learning paradigms on heterophilic graphs in Section 6, including self-supervised learning and prompt learning. Apart from remarkable GNN models and

learning paradigms, some extensible topics are also mentioned in Section 7. In Section 8, we discuss the practical applications, and provide unique insights for future explorations in Section 9.

## 4 Measures and benchmarks

The success of graph learning depends on high-quality data. To evaluate proposed models, various heterophilic benchmarks have been released [49–54]. Meanwhile, some metrics for graph heterophily have been proposed to characterize datasets. This section discusses the heterophily measures, and then provides a comprehensive review on benchmarks.

### 4.1 Measuring heterophily

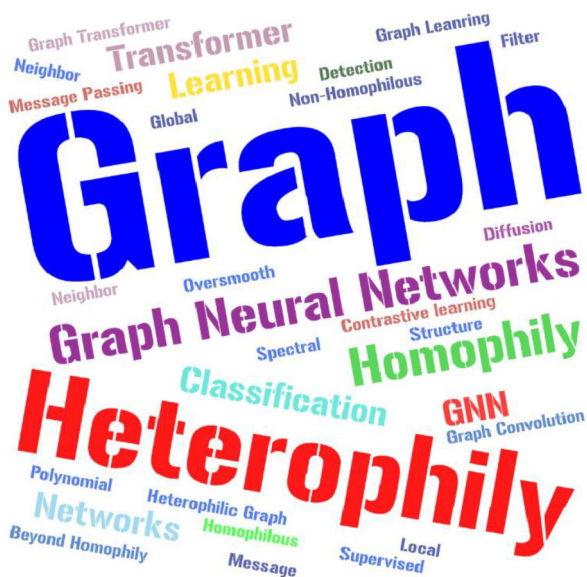
Graph heterophily refers to the phenomenon that connected nodes tend to share different features or labels. Understanding this concept and establishing relevant metrics is crucial for further research. Here, we introduce representative metrics for graph heterophily, typically presented in terms of homophily level. For example, node homophily [50] and edge homophily [20] are most frequently used. Specifically, node homophily measures homophily at the node level, where homophily degree for each node is computed as the proportion of neighbors sharing the same label. Then node homophily  $\mathcal{H}_{\text{node}}$  is defined as the average homophily degree of all nodes:

$$\mathcal{H}_{\text{node}} = \frac{1}{|\mathcal{V}|} \sum_{v \in \mathcal{V}} \frac{|\{u \in \mathcal{N}(v) : y_v = y_u\}|}{|\mathcal{N}(v)|},$$

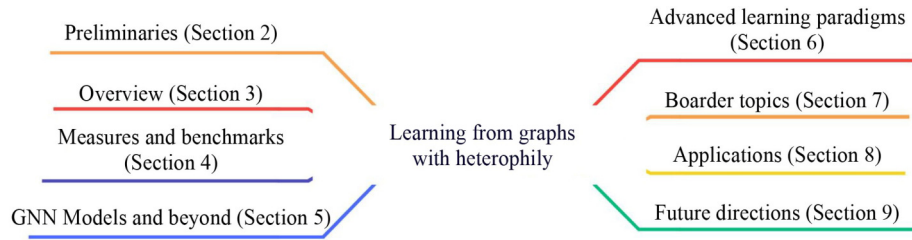
where  $y_v, y_u$  denote the labels of nodes  $v$  and  $u$ . It is worth noting that  $\mathcal{H}_{\text{node}}$  only reflects homophily within 1-hop neighbors. Recent works [55,56] extend this definition to the  $k$ -hop to measure high-order homophily:

$$\mathcal{H}_{\text{high-order}} = \frac{|\{u \in \mathcal{N}_k(v) : y_v = y_u\}|}{|\mathcal{N}_k(v)|},$$

where  $\mathcal{N}_k(v)$  denotes the  $k$ -hop neighbor set of  $v$ . Edge homophily  $\mathcal{H}_{\text{edge}}$  measures homophily at the edge level, and is defined as the fraction of edges connecting nodes with the same label:



**Fig. 3** The word cloud composed of high-frequency words from the titles of collected articles in this survey



**Fig. 4** The overall organizational structure of this survey.

$$\mathcal{H}_{\text{edge}} = \frac{|\{(v, u) \in \mathcal{E} : y_v = y_u\}|}{|\mathcal{E}|}$$

Both node and edge homophily range from  $[0, 1]$ , and high homophily indicates low heterophily, and vice versa. While widely used, these two simple metrics are highly sensitive to the number of class, leading to limited utility [51]. To mitigate the class imbalance issue, another metric named class homophily [51] is defined as:

$$\mathcal{H}_{\text{class}} = \frac{1}{C-1} \sum_{k=1}^C \left[ \mathcal{H}_k - \frac{|C_k|}{|\mathcal{V}|} \right]_+$$

where  $[x]_+ = \max\{x, 0\}$ ,  $C$  is the number of class,  $C_k$  is the set of nodes in class  $k$ , and  $\mathcal{H}_k$  is the class-wise homophily metric:

$$\mathcal{H}_k = \frac{\sum_{v: y_v=c} |\{u \in \mathcal{N}(v) : y_u = y_v\}|}{\sum_{v: y_v=c} |\mathcal{N}(v)|}$$

Although  $\mathcal{H}_{\text{class}}$  measures heterophily more fairly, some issues still exist [57]. For example,  $\mathcal{H}_{\text{class}}$  neglects the variation of node degrees when correcting the fraction of intra-class edges by its expected value. To this end, combined with assortativity coefficient [58], adjusted homophily is defined as:

$$\mathcal{H}_{\text{adj}} = \frac{\mathcal{H}_{\text{edge}} - \sum_{k=1}^C \mathcal{D}_k^2 / (2|\mathcal{E}|)^2}{1 - \sum_{k=1}^C \mathcal{D}_k^2 / (2|\mathcal{E}|)^2}$$

where  $\mathcal{D}_k$  is the sum of degrees of all nodes with label  $k$ .

**Discussion.** Overall, node homophily  $\mathcal{H}_{\text{node}}$  and edge homophily  $\mathcal{H}_{\text{edge}}$  are widely used in measuring heterophily due to their simplicity and ease of implementation. However, they overlook the impact of class imbalance and node degree issues on heterophily, leading to the introduction of class homophily  $\mathcal{H}_{\text{class}}$  and adjusted homophily  $\mathcal{H}_{\text{adj}}$ . Recent study [57] summarizes the desirable properties of heterophily measures and the aboved metrics besides adjusted homophily have critical drawbacks in comparison of heterophily across datasets. Moreover, various statistic-based metrics for graph heterophily are continuously emerging [59–64]. Recently, some studies have used classifier-based [65] and unsupervised-based [66] methods. Zheng et al. [67] attempted to disentangle heterophily from label, structural, and feature aspects, and provide a comprehensive review of existing metrics. Although the proposed metrics help characterize heterophily, recent studies show that GNN performance does not always align with heterophily [61,65,68]. Therefore, how graph heterophily affects GNNs remains unclear and is worth exploring.

## 4.2 Benchmark datasets

To boost learning from graphs with heterophily, there is an urgent need for trustworthy and high-quality benchmarks. We further categorize existing benchmarks into three categories: basic, large-scale, and advanced. Note that all datasets focus on the node classification task, and detailed statistics are presented in Table 1.

### 4.2.1 Basic benchmark

Inspired by the studies in complex networks [49,71,72], Pei et al. [50] summarized the first benchmark for heterophilic graphs. To date, most studies have evaluated their models on it, which we refer to as basic benchmark. Basic benchmark includes six datasets: **Cornell**, **Texas**, **Wisconsin**, **Chameleon**, **Squirrel** and **Actor**. Next, we categorize them based on their sources and provide detailed descriptions.

–**WebKB.** WebKB is a webpage dataset collected from computer science departments by Carnegie Mellon University. **Cornell**, **Texas**, **Wisconsin** are three sub-datasets of WebKB [72], where nodes represent web pages, and edges represent hyperlinks between web pages.

–**Wikipedia.** **Chameleon** and **Squirrel** are two page-page networks on specific topics collected from Wikipedia [49], which is a free, online encyclopedia that anyone can edit. In these datasets, nodes represent web pages and edges are mutual links between web pages.

–**Actor.** **Actor**, also called **Film**, is the actor-only induced subgraph of a “film-director-actor-writer” network [71]. Each node represents an actor, and the edge between two nodes denotes their co-occurrence on the same Wikipedia page.

Despite the widespread use, basic benchmark suffers from limitations due to the small scale and narrow domain, which is inadequate for evaluation.

### 4.2.2 Large-scale benchmark

Aside from the overfitting risks caused by mall-scale datasets [73], evaluation on basic benchmark is plagued by high variance across different splits [20]. To this end, a series of large-scale datasets from diverse domains are collected and released in [51,52], forming the large-scale benchmark: **ArXiv–Year**, **Snap–Patents**, **Wiki**, **Penn94**, **Pokec**, **Genius**, **Twitch–Gamers** and **Deezer–Europe**. Additionally, there are also some representative large-scale datasets, such as **Wiki–Cooc**, **BlogCatalog** and **Flickr**. Since these datasets show the significant heterophily, we also include them in the large-scale benchmark.

**Table 1** The statistics of current benchmark datasets for heterophilic graphs

Dataset	# Nodes	# Edges	# Features	# Classes	$\mathcal{H}_{\text{node}}$	$\mathcal{H}_{\text{edge}}$	Source	Literature
Cornell	183	295	1,703	5	0.20	0.30	WebKB	[50]
Texas	183	309	1,703	5	0.06	0.06	WebKB	[50]
Wisconsin	251	499	1,703	5	0.10	0.17	WebKB	[50]
Chameleon	2,277	36,101	2,325	5	0.25	0.23	Wikipedia	[50]
Squirrel	5,201	217,073	2,089	5	0.22	0.22	Wikipedia	[50]
Actor	7,600	33,544	931	5	0.20	0.22	Actor	[50]
Penn94	41,554	1,362,229	5	2	0.48	0.47	Social Media	[52]
Pokec	1,632,803	30,622,564	65	2	0.39	0.44	Social Media	[52]
Genius	421,961	984,979	12	2	0.10	0.62	Social Media	[52]
Twitch-Gamers	168,114	6,797,557	7	2	0.56	0.55	Social Media	[52]
Deezer-Europe	28,281	92,752	31,241	2	0.53	0.53	Social Media	[51]
Blog-Catalog	5,196	171,743	8,189	6	0.39	0.40	Social Media	[69]
Flickr	7,575	239,738	12,047	9	0.24	0.24	Social Media	[69]
ArXiv-Year	169,343	1,166,243	128	5	0.28	0.22	Citation	[52]
Snap-Patents	2,923,922	139,975,788	269	5	0.19	0.07	Citation	[52]
Wiki	1,925,342	303,434,860	600	5	–	0.39	Wikipedia	[52]
Wiki-Cooc	10,000	2,243,042	100	5	0.18	0.34	Wikipedia	[70]
Roman-Empire	22,662	32,927	300	18	0.05	0.05	Wikipedia	[53]
Amazon-Ratings	24,492	93,050	300	5	0.38	0.38	E-commerce	[53]
Minesweeper	10,000	39,402	7	2	0.68	0.68	Digital Game	[53]
Tolokers	11,758	519,000	10	2	0.63	0.59	Crowdsourcing	[53]
Questions	48,921	153,540	301	2	0.90	0.84	Q&A Platform	[53]

–Social Media. **Penn94** [74] is a social network from the Facebook of university students, where nodes represent students and edges represent the friendships. Similarly, **Pokec** [75] is the friendship network of a Slovak online social network, where nodes represent users and edges represent directed friendship relations. **Genius** [76] is a subset of the social network on a web site for crowdsourced annotations of song lyrics. The nodes are users, and edges connect users that follow each other on the website. **Twitch-Gamers** [77] is a network of relationships between accounts on the livestreaming platform Twitch. Each node represents a Twitch account, and edges exist between accounts sharing mutual followers. **Deezer-Europe** [78] is a social network of users on Deezer from European countries, where edges represent mutual follower relationships. **BlogCatalog** [69] is a also social network created from an online community where bloggers can follow each other. **Flickr** [69] is a dataset created from an online website sharing images and videos where users can follow each other, forming a social network.

–Citation. **ArXiv-Year** [79] is the Ogbn-ArXiv citation network

labeled by the posted year, instead of subject areas. The nodes are ArXiv papers, and directed edges connect a paper to other papers that it cites. **Snap-Patents** [75,80] is a dataset of utility patents granted between 1963 to 1999 in the US, where each node is a patent, and edges connect patents that cite each other.

–Wikipedia. **Wiki** [52] is a dataset of Wikipedia articles, where nodes represent pages and edges represent links between them. **Wiki-Cooc** [70] is a dataset based on the English Wikipedia, where nodes denote unique words and edges connect frequently co-occurring words.

Motivated by the large-scale benchmark, research papers on learning from graph heterophily start to flourish. To better evaluate models and verify whether they capture heterophily patterns, we urgently need more advanced, high-quality datasets.

#### 4.2.3 Advanced benchmark

Recently, a critical study has suggested existing benchmarks for heterophily-specific GNNs have serious drawbacks, rendering evaluation based on them unreliable [53]. The fatal drawback is the presence of duplicate nodes in **Chameleon** and **Squirrel**,

causing data leakage. Experiments have shown that removing duplicate nodes strongly affects many heterophily-specific GNNs. In addition, **Cornell**, **Texas**, and **Wisconsin** are not innocent. These datasets have very imbalanced classes. For example, **Texas** has a class that consists of only one node, making this class for training and evaluation meaningless. Aware of these issues, several new datasets with different nature and with diverse structural properties are released, collectively forming the advanced benchmark [53]. The advanced benchmark includes five datasets for evaluation: **Roman – Empire**, **Amazon – Ratings**, **Minesweeper**, **Tolokers**, and **Questions**.

–Wikipedia. **Roman – Empire** is based on the one of the longest articles on Wikipedia, Roman Empire. Each node corresponds to one (non-unique) word in the text and two nodes are connected with an edge if subjected to the proposed conditions.

–E-commerce. **Amazon – Ratings** is based on the Amazon product co-purchasing network metadata dataset from SNAP [75]. Nodes are products (books, CDs, DVDs, video tapes), and edges connect products that are frequently bought together.

–Digital Game. **Minesweeper** is a synthetic network based on the Minesweeper game. It is a  $100 \times 100$  grid where node is connected to eight neighbors, except for nodes at the edge of the grid. 20% of the nodes are randomly selected as mines and the task is to predict which nodes are mines.

–Crowdsourcing. **Tolokers** is based on data from the Toloka crowdsourcing platform [81]. The nodes represent tolokors (workers) in at least one of 13 selected projects. An edge connects two tolokors if they are in the same task.

–Q&A Platform. **Questions** is based on the data from a question-answering website Yandex Q. The nodes are users, and an edge connects two nodes if one user answered the other user’s question during a one-year time interval.

Evidently, we can find that the advanced benchmark not only has unique graph heterophily characteristics, but also covers a broader scope, which is more closely related to our daily lives.

#### 4.2.4 Discussion

As can be seen in Table 2, we summarize the macroscopic characteristics across different heterophilic graph benchmarks. Basic benchmark focuses on ease of use; however, due to data leakage, this issue must be processed before utilization. Large-scale benchmark, on the other hand, emphasizes evaluating the scalability of models. Advanced benchmark, owing to its reliability and high quality, has

become the mainstream evaluation benchmark for assessment. In addition, we find that basic benchmark, due to their limited scope, exhibit relatively narrow heterophily patterns. In contrast, large-scale and advanced benchmark cover a broader range of domains, thus exhibiting richer heterophily patterns, which are more beneficial for comprehensive evaluation. The common heterophily metrics and literature sources for each dataset in these benchmarks can be found in Table 1. For further advancement of this field, we advocate for the release of high-quality open-source datasets from a broader spectrum of fields. We also recommend further cross-pollination of heterophilic datasets with NLP, CV, and other fields, enabling models to truly address real-world application challenges.

#### 4.3 Model reassessment

Given flaws identified in current benchmarks, several studies have suggested a comprehensive reassessment of existing models. Platonov et al. [53] first uncovered fatal problems within popular datasets and revealed that the previous evaluation is unreliable. They further proposed the advanced benchmark for reassessment and showed that standard GNNs generally outperform heterophily-specific methods. Through extensive empirical analysis, Luo et al. [82] investigated the influence of many GNN configurations such as dropout [83,84], normalization [85], residual connections [86,87], network depth [21,87,88], and jumping knowledge [89] on node classification. Experiments show that classic GNNs (GCN, GAT, and GraphSAGE) with these configurations can match or even outperform recent Graph Transformers on most heterophilic and homophilic datasets with slight hyper-parameter tuning. Beyond classic GNNs, Liao et al. [43] extensively benchmarked spectral GNNs from the frequency perspective. They proposed that most spectral models can be divided into three categories: fixed filter, variable filter and filter bank. All three types can achieve satisfactory performance on homophilic graphs, while variable filters and filter bank models excel under graph heterophily. To identify the real challenging subsets of heterophilic datasets, Luan et al. [5,54] benchmarked the commonly used heterophilic datasets and classify them into benign, malignant and ambiguous datasets. They proposed that good models for heterophily should match classic methods on homophilic graphs, and perform better especially on malignant and ambiguous datasets.

We advocate for the comprehensive benchmarks and fair evaluation to compare proposed models. This approach not only lays a solid foundation for future research but also further promotes advancements in learning from graphs with heterophily.

### ■ 5 GNN models and beyond

Graph heterophily has led to the emergence of heterophilic GNN

**Table 2** The comparison of macroscopic characteristics across different heterophilic graph benchmarks

Benchmark	# Datasets	Dataset scale	Dataset scope	Remark
Basic benchmark	6	Small	Knowledge graph & social networks	Data leakage
Large-scale benchmark	11	Large	Social & citation networks	Focus on model scalability
Advanced benchmark	5	Medium	Broader applications	Reliable & high-quality

models. In this section, we categorize representative methods according to the model architecture and its underlying mechanism into six types, as can be seen in Fig. 5. The first four type methods extend the classic GNN architecture, making improvements from four perspectives: spectral graph filters, high-order neighbor expansion, global homophily modeling, and advanced message passing, which aim to adapt GNN models for heterophilic graphs. Drawing on the powerful modeling capabilities of Transformers [24] and Neural ODEs [189], we also introduce studies based on Graph Transformers and Neural Diffusion Processes. These advanced architectures extend beyond conventional GNNs, offering novel insights to handle graph heterophily.

### 5.1 Spectral graph filters

Based on graph signal processing [190], spectral filters have been integrated into GNNs to enhance the expressive power. Most GNNs smooth representations of connected nodes, which is equivalent to low-pass filtering [191]. For example, GCN [15] utilizes the self-loop operation to enhance the low-pass filter, and its approximate filter  $g(\tilde{\mathbf{L}})$  can be expressed as:

$$g(\tilde{\mathbf{L}}) = \mathbf{I} + \mathbf{F}_{\text{LP}} = \mathbf{I} + \tilde{\mathbf{A}} = 2\mathbf{I} - \tilde{\mathbf{L}}.$$

However, high-frequency signals that capture the dissimilarity between nodes are often neglected, restricting the expressive power of heterophily. To this end, many studies design advanced spectral filters to capture heterophily patterns. Here, we follow Liao et al. [43] and categorize existing spectral methods into fixed filter, variable filter, and filter bank.

#### 5.1.1 Fixed filter

For the first type, the basis and parameters of graph filters are both constant, resulting in fixed filters  $g(\tilde{\mathbf{L}})$ . As an extension of the homophilic scenario [192–194], GCNII [90] enhance GCN under heterophily with simple yet effective techniques: Initial residual and Identity mapping, to preserve ego or local information. The spectral interpretation of the polynomial approximation is:

$$g(\tilde{\mathbf{L}}) = \sum_{k=0}^K \alpha(1-\alpha)^k (\mathbf{I} - \tilde{\mathbf{L}})^k,$$

where  $K$  is the spectral filter order,  $\alpha \in [0, 1]$  is the coefficient for balancing neighbor propagation and ego preservation. Due to the preset parameters, methods based on fixed filters still face the problem of insufficient expressive power.

#### 5.1.2 Variable filter

Compared to fixed filters, variable filters extend the approximate filter to a variable version  $g(\tilde{\mathbf{L}}, \theta)$ . Inspired by generalized PageRank [194, 195], GPR-GNN [91] formulates the spectral filter as:

$$g(\tilde{\mathbf{L}}, \theta) = \sum_{k=0}^K \theta^k (\mathbf{I} - \tilde{\mathbf{L}})^k,$$

where  $\theta$  is the learnable weight to fit heterophily patterns. ASGC [92] adopts the same principle and provides a more simplified version. ChebNet [196] is also a typical example, which utilizes the

Chebyshev polynomial as basis:

$$g(\tilde{\mathbf{L}}, \theta) = \sum_{k=0}^K \theta^k T^{(k)}(\tilde{\mathbf{L}}).$$

Each term  $T^{(k)}(\tilde{\mathbf{L}})$  can be expressed recursively:

$$\begin{aligned} T^{(k)}(\tilde{\mathbf{L}}) &= 2\tilde{\mathbf{L}}T^{(k-1)}(\tilde{\mathbf{L}}) - T^{(k-2)}(\tilde{\mathbf{L}}), \\ T^{(1)}(\tilde{\mathbf{L}}) &= \tilde{\mathbf{L}}, T^{(0)}(\tilde{\mathbf{L}}) = \mathbf{I}. \end{aligned}$$

Based on ChebNet, ChebNetII [93] employs the Chebyshev interpolation [197] to enhance the filter:

$$g(\tilde{\mathbf{L}}, \theta) = \frac{2}{K+1} \sum_{k=0}^K \sum_{\kappa=0}^K \theta_{\kappa} T^{(k)}(x_{\kappa}) T^{(k)}(\tilde{\mathbf{L}}),$$

where  $T^{(k)}(x_{\kappa})$ ,  $T^{(k)}(\tilde{\mathbf{L}})$  follow the Chebyshev basis, and  $x_{\kappa} = \cos \frac{(\kappa+1/2)\pi}{K+1}$  are Chebyshev nodes of  $T^{K+1}$  [93]. ClenshawGCN [94] applies Clenshaw algorithm to incorporate explicit residual connections. The form of its filter is similar to chebNet, but with some differences in the iteration:

$$\begin{aligned} T^{(k)}(\tilde{\mathbf{L}}) &= 2\tilde{\mathbf{L}}T^{(k-1)}(\tilde{\mathbf{L}}) - T^{(k-2)}(\tilde{\mathbf{L}}), \\ T^{(1)}(\tilde{\mathbf{L}}) &= 2\tilde{\mathbf{L}}, T^{(0)}(\tilde{\mathbf{L}}) = \mathbf{I}. \end{aligned}$$

To avoid oversimplified learnt weights, BernNet [95] substitutes the generalized PageRank with Bernstein polynomial:

$$g(\tilde{\mathbf{L}}, \theta) = \sum_{k=0}^K \frac{\theta_k}{2^K} T^{(k)}(\tilde{\mathbf{L}}), T^{(k)} = \binom{K}{k} (2\mathbf{I} - \tilde{\mathbf{L}})^{K-k} (\tilde{\mathbf{L}})^k.$$

In addition to the aforementioned models, other models, such as LegendreNet [96], JacobiConv [97], FavardGNN [98], also focus on designing advanced variable filters to enhance the expressive power for heterophily. For more details, please refer to [43].

#### 5.1.3 Filter bank

GNNs with filter bank integrate multiple fixed or variable filters to enhance the expressive power. Given the number of filters  $Q$ , the integration can be formulated as:

$$\hat{g}(\tilde{\mathbf{L}}, \gamma) = \bigoplus_{q=1}^Q \gamma^q \cdot g_q(\tilde{\mathbf{L}}),$$

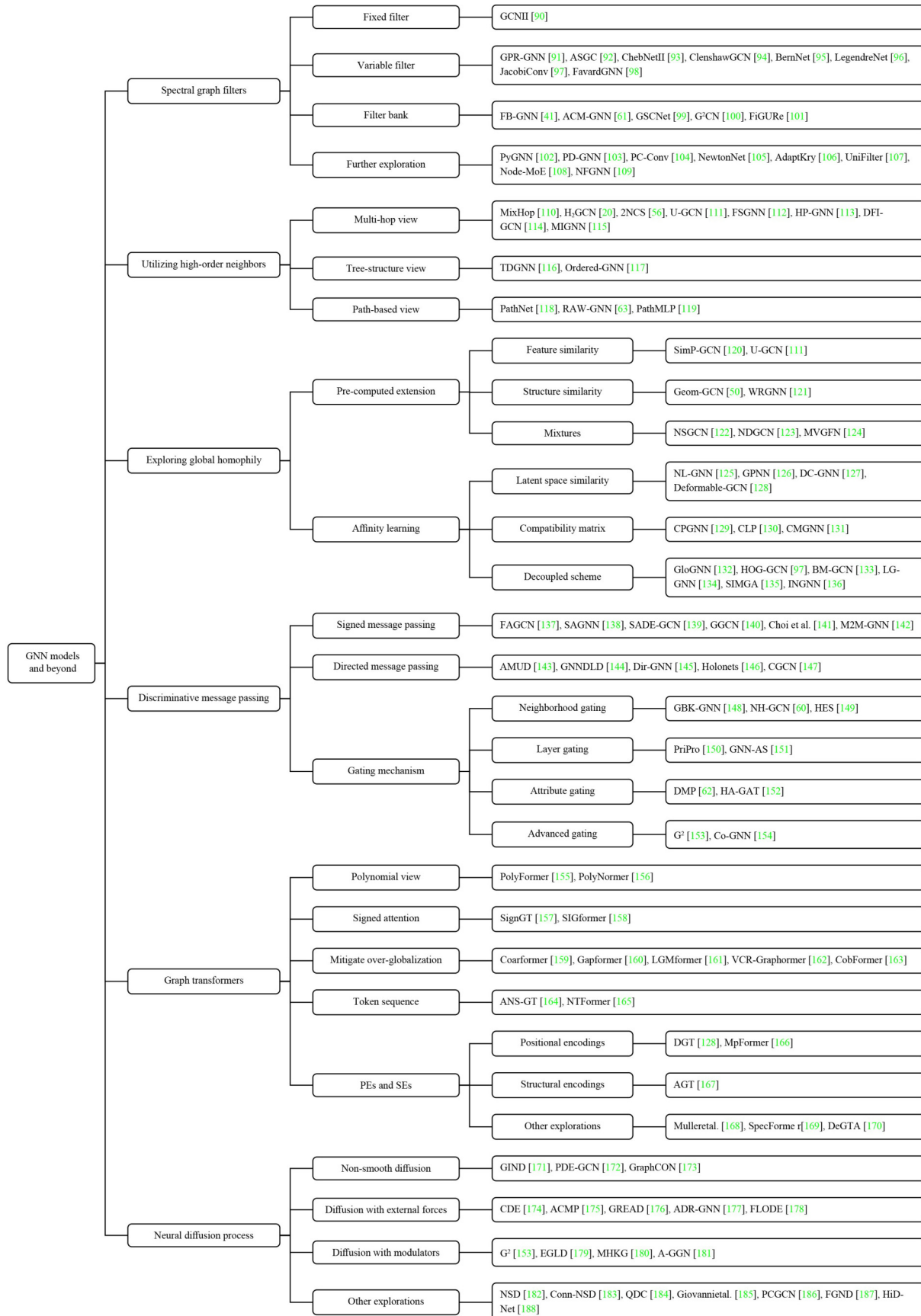
where  $\bigoplus$  denotes an arbitrary combination such as sum or concatenation, and each filter  $g_q(\tilde{\mathbf{L}})$  is assigned with a weight  $\gamma^q$ . FB-GNN [41] first introduces the concept of filter bank under the heterophily setting. It designs a dual-channel scheme with low-pass and high-pass filters to learn the smooth and non-smooth components:

$$\hat{g}(\tilde{\mathbf{L}}; \gamma) = \gamma_1 \mathbf{F}_{\text{LP}} + \gamma_2 \mathbf{F}_{\text{HP}} = \gamma_1 \tilde{\mathbf{L}} + \gamma_2 (\mathbf{I} - \tilde{\mathbf{L}}),$$

where  $\gamma_1, \gamma_2 \in [0, 1]$  is the learnable scalar parameters. ACM-GNN [61] extends FBGNN to three filters, including an all-pass filter to maintain node identity:

$$\hat{g}(\tilde{\mathbf{L}}; \gamma) = \gamma_1 \tilde{\mathbf{L}} + \gamma_2 (\mathbf{I} - \tilde{\mathbf{L}}) + \gamma_3 \mathbf{I}.$$

Similarly, FAGCN [137] integrates two spectral filters with bias to capture both low- and high-frequency signals:



**Fig. 5** Taxonomy of heterophilic GNNs and beyond with representative examples, the categorization criterion is based on the model architecture and underlying mechanisms

$$\hat{g}(\tilde{\mathbf{L}}; \gamma) = \gamma_1((\beta + 1)\mathbf{I} - \tilde{\mathbf{L}}) + \gamma_2((\beta - 1)\mathbf{I} + \tilde{\mathbf{L}}),$$

where  $\beta \in [0, 1]$  is the scaling coefficient. GSCNet [99] proposes a simple basis that decouples the positive and negative activation:

$$\hat{g}(\tilde{\mathbf{L}}; \gamma) = \sum_{i=0}^{K_1} \gamma_i(2\mathbf{I} - \tilde{\mathbf{L}})^i + \sum_{j=0}^{K_2} \gamma_j(\tilde{\mathbf{L}})^j,$$

where activation ratios can be adjusted by hyper-parameters  $K_1$  and  $K_2$ . G<sup>2</sup>CN [100] proposes the Gaussian filters with sufficient flexibility:

$$\begin{aligned} \hat{g}(\tilde{\mathbf{L}}; \gamma) &= \gamma_1 \sum_{k=1}^{\lfloor K/2 \rfloor} \theta_{1,k} T_1^{(k)} + \gamma_2 \sum_{k=1}^{\lfloor K/2 \rfloor} \theta_{2,k} T_2^{(k)}, \\ T_1^{(k)} &= ((1 + \beta_1)\mathbf{I} - \tilde{\mathbf{L}}), \theta_{1,k} = \frac{\alpha_1^k}{k!}, \\ T_2^{(k)} &= ((1 - \beta_2)\mathbf{I} - \tilde{\mathbf{L}}), \theta_{2,k} = \frac{\alpha_2^k}{k!}, \end{aligned}$$

where  $\alpha, \beta \in [0, 1]$  is the decay coefficient and scaling coefficient. Typically, FiGURe [101] utilizes the filter-level parameters to control each component:

$$\hat{g}(\tilde{\mathbf{L}}; \gamma) = \sum_{q=1}^Q \gamma^q \cdot g_q(\tilde{\mathbf{L}}).$$

Various filters, including Chebyshev, Bernstein and others, can be used to compose the filter bank. It is worth noting that variable filters and filter bank models outperform fixed filter under heterophily [43]. Therefore, we need to consider the trade-off between model capacity and computational cost.

#### 5.1.4 Further explorations

In addition to filter design, there are also further explorations of spectral graph filter. For example, PyGNN [102] enhances spectral filters with sampling techniques, while PD-GNN [103] introduces a new Laplacian matrices, offering more flexibility. Inspired by heat equation [198], PC-Conv [104] introduces the Poisson-Charlier polynomial [199] to make an exact numerical approximation. NewtonNet [105] establishes the connections between spectral frequencies and graph heterophily, and further integrates filter with Newton Interpolation [200]. Huang et al. [106] optimized the polynomial filters in Krylov subspace [201], and further reveal how universal polynomial bases enhance spectral GNNs [107]. Motivated by Mixture of Experts [202], Node-MoE [108] proposes that node-wise filtering can achieve linear separability, and further designs a Mixture of Experts GNN framework. Besides Node-MoE, both DSF [98] and NFGNN [109] are dedicated to designing adaptive node-specific filtering.

#### 5.1.5 Discussion

Based on spectral graph theory, spectral GNNs focus on frequency signals of whole graphs and are more theoretically grounded. Despite remarkable success, existing spectral GNNs still face two major issues: the polynomial limitation and the transductive limitation [203]. The flexibility of polynomial filters, the co-existence of homophilic and heterophilic patterns, as well as scalability on large

graphs, remain challenges that spectral GNNs urgently need to address.

## 5.2 Utilizing high-order neighbors

Under heterophilic setting, simply weighted averaging representations of neighbors [15,17] in message passing will inevitably introduce noise, potentially resulting in low-quality representations [20].

### 5.2.1 Multi-hop view

To mitigate graph heterophily, a natural idea is to utilize informative nodes within multi-hops. Formally, the  $k$ -hop neighbor set can be defined as:

$$\mathcal{N}_k(v) = \{u : d(u, v) = k\},$$

where  $d(u, v)$  measures the shortest distance between  $u$  and  $v$ . MixHop [110] demonstrates that utilizing multi-hop neighbors provides a wider class of representations. The extension of message passing to multi-hops can be formulated as:

$$\begin{aligned} r_{i,k}^{(l)} &= \text{AGG}(\{h_j^{(l-1)} : v_j \in \bigcup_{q=0}^k \mathcal{N}_q(v)\}), (k = 0, 1, 2), \\ h_i^{(l)} &= \text{CONCAT}(r_{i,0}^{(l)}, r_{i,1}^{(l)}, r_{i,2}^{(l)}), \end{aligned}$$

where CONCAT denotes the column-wise combination. The arbitrary combinations can be learned to model more complex patterns, such as graph heterophily. Similar to Mixhop, H<sub>2</sub>GCN [20] considers 2-hop message passing, and utilizes the combination of intermediate representations:

$$\begin{aligned} r_{i,k}^{(l)} &= \text{AGG}(\{h_j^{(l-1)} : v_j \in \mathcal{N}_k(v)\}), (k = 0, 1, 2), \\ h_i^{(l)} &= \text{CONCAT}(r_{i,0}^{(l)}, r_{i,1}^{(l)}, r_{i,2}^{(l)}), \\ h_i &= \text{CONCAT}(h_i^{(0)}, h_i^{(1)}, h_i^{(2)}). \end{aligned}$$

Both Mixhop and H<sub>2</sub>GCN utilize CONCAT operation to separate the ego and neighbor representations, which has been proven to be indeed helpful under heterophily [53]. Particularly, H<sub>2</sub>GCN emphasizes the 2-hop neighbors, and demonstrates that 2-hop neighborhood for a node is always homophily-dominant. Cavallo et al. [56] also find that 2-hop Neighbor Class Similarity (2NCS) correlates with GNN performance more strongly. U-GCN [111] discovers that 1-hop,  $k$ NN and 2-hop neighbors are more suitable as neighborhoods in networks with complete homophily, randomness and complete heterophily, respectively. To extract information from different neighborhoods, U-GCN performs a multi-type convolution based on attention mechanism. Moreover, FSGNN [112] employs the softmax operation to regularize messages from different hops, and HP-GNN [113] designs a memory unit to retain multi-hop information. DFI-GCN [114] extracts multi-hop interactions between different features via Newton's identities, and MIGNN [115] utilizes the mutual information to model the dependence between nodes within  $k$ -hop neighbors.

### 5.2.2 Tree-structure view

Another line of utilizing high-order neighbors is to model the neighborhood as a tree structure. For example, TDGNN [116] views

higher-order neighborhoods as tree structures to provide insights into addressing heterophily. Ordered-GNN [117] emphasizes the orders of tree structure, integrates inductive bias from rooted-tree hierarchy, and encodes neighbors at some orders. This method prevents the mixing of node features within hops, leading to superior performance on both heterophilic and homophilic graphs.

### 5.2.3 Path-based view

High-order neighborhoods can similarly be transformed into a multitude of paths. PathNet [118] first sheds light on path-level patterns, which explicitly reflect rich semantic and structural information on graphs. This work proposes a novel path aggregation strategy with path sampling, and designs a structure-aware cell for path-level aggregation. PathMLP [119] designs a similarity-based path sampling, encodes path-level messages through simple transformation and concatenation, and performs adaptive path aggregation. Similarly, RAW-GNN [63] employs Breadth-First Search (BFS) and Depth-First Search (DFS) views to model graph homophily and heterophily patterns. After obtaining BFS and DFS paths, RAW-GNN utilizes RNN to encode each path and learns the path-level attention for aggregation.

### 5.2.4 Discussion

Obviously, in a larger neighborhood with nodes containing rich information, fully utilizing higher-order neighbors can enhance model performance. However, selecting an appropriate high-order neighborhood size remains challenging: a larger neighborhood increases computational costs, while a smaller one may not adequately address the heterophily issue. While empirical results suggest that a 2-hop neighborhood might be suitable [20,56,111], this does not guarantee applicability in complex real-world scenarios.

## 5.3 Exploring global homophily

Since local heterophily has a negative impact on GNNs, another approach is to explore homophily globally. Generally, we can introduce a potential neighbor set  $\mathcal{N}_p$ :

$$\mathcal{N}_p = \{u : \text{sim}(u, v) > \rho\},$$

where  $\rho$  is a threshold, and  $\text{sim}(\cdot)$  is a similarity function that can be implemented based on feature, structure or other metrics. In this way, we extend the neighbor set in Eq. (1) from a global perspective to mitigate the heterophily issue.

### 5.3.1 Pre-computed extension

A straightforward approach is to extend the neighbor set through pre-computation.

–Feature similarity. Globally exploring homophily based on feature is the most widely adopted approach. Here,  $\text{sim}(\cdot)$  is implemented by dot product or cosine similarity, and the similarity matrix can be pre-computed as:

$$S_{ij} = \cos(x_i, x_j) = \frac{x_i^\top x_j}{\|x_i\| \|x_j\|}.$$

We can choose  $k$  Nearest Neighbors ( $k$ NN) for each node as the potential neighbor set. For example, SimP-GCN [120] and U-GCN

[111] adaptively combine information from the original graph and  $k$ NN feature graph, and find this simple method achieves good performance under heterophilic setting.

–Structure Similarity. Structure-based methods place emphasis on using graph structure to discover global neighbors. Typically, Geom-GCN [50] uses three structural metrics: Isomap [204], Poincare [205], and Struc2Vec [206], to implement  $\text{sim}(\cdot)$ , and maps nodes to latent spaces for geometric relation mining. Neighbors that conform to the geometric relationships also participate in the message passing. WRGNN [121] takes the degree sequence of neighbors as the metric to construct a multi-relational graph. Message passing on this computation graph breaks the limit of local assortative and facilitates global integration.

–Mixtures. It is also feasible to simultaneously utilize feature and structure for neighbor extension. For example, NSGCN [122] computes the common neighbors distribution and feature similarity for extension, while NDGCN [123] further considers the higher-order neighborhood distribution. MVGFN [124] defines the semantic  $k$ NN neighbors based on feature similarity and the structural  $k$ NN neighbors based on embedding algorithms, such as diffusion process [194] and node2vec [207]. Combined with 1-hop and 2-hop neighbors, MVGFN aggregates hybrid information from these neighbor sets and introduces a multi-view fusion framework.

### 5.3.2 Affinity learning

While pre-computed extension is available, exploring global homophily through learning offers more flexibility. Generally, this type of methods learn an affinity matrix to model global relationships and guide message passing.

–Latent space similarity. Considering the computation cost, an intuitive idea is to map nodes to the latent space to explore global homophily. For example, NL-GNN [125] directly utilizes the GNNs to embed nodes, and GPNN [126] leverages the Pointer Networks [208] to obtain node embeddings. After that, both of them compute the affinity matrix based on embeddings through attention, and then perform non-local aggregation. Similar approach are also applied in DC-GNN [127] and Deformable-GCN [128] to combat the graph heterophily issue.

–Compatibility matrix. Some studies introduce the compatibility matrix [129] to boost homophily mining. Compatibility matrix models the connection probability of nodes between classes, which can be transformed into an affinity matrix to guide message passing. CPGNN [129] first incorporates the compatibility matrix into GNNs, and designs the compatibility-guided propagation. CLP [130] learns the class compatibility matrix and generalizes label propagation to accommodate the heterophily assumption. CMGNN [131] revisits the message passing under heterophily, reformulates them into a unified framework, and reveals that enhancing the class compatibility matrix is the key to heterophily-specific GNNs. Moreover, LRGNN [209] extends the compatibility matrix to a signed global relationship matrix, and reformulates the matrix prediction into a low-rank approximation problem.

–Decoupled scheme. Another way to model the affinity matrix is by decoupling the structure and feature. Typically, GloGNN [132] uses MLPs to map the node feature and adjacency matrix into feature and topology views separately, and further fuses them with a term weight. After decoupling, GloGNN aggregates information from global nodes and designs an acceleration strategy to avoid the quadratic time complexity. HOG-GCN [97], BM-GCN [133], and LG-GNN [134] follow GloGNN and perform the decoupled scheme. Specifically, HOG-GCN applies label propagation to enhance topology view, BM-GCN introduces block modeling-guided propagation and LG-GNN integrates SimRank [210] and ListNet loss [211] to enhance two views. Further, SIMGA [135] and INGNN [136] theoretically proves that decoupled strategy is effective in discovering global relationship and grouping similar nodes under graph heterophily.

### 5.3.3 Discussion

We acknowledge that global computation significantly boosts model performance on heterophilic graphs but incurs higher computational costs. It is worth investigating whether global neighbors are necessary and if exploring global homophily could have negative effects. For fully connected graphs, Transformer-based models seem more suitable, raising questions about whether GNNs are the optimal choice for exploring global homophily.

## 5.4 Discriminative message passing

Besides the above methods, heterophilic information in local neighborhood is also worth consideration. We can adopt the discriminative mechanism to retain useful information and filter noisy messages.

### 5.4.1 Signed message passing

Aggregation weights in message passing are typically set positive, limiting the capture of heterophily patterns. To this end, FAGCN [137] first allows the aggregation weights in message passing to be negative to accommodate graph heterophily. SAGNN [138] and SADE-GCN [139] regularize the weights and restrict the range to  $[-1, 1]$ , while GGCN [140] approximates the sign function with cosine similarity. Choi et al. [141] pointed that signed messages escalate the inconsistency between neighbors and increase the uncertainty in predictions. Therefore, they proposed to adopt calibration for signed GNNs to reduce uncertainty. Further, Liang et al. [142] proposed that signed message passing has limitations: undesirable representation update for multi-hop neighbors and vulnerability against over-smoothing. To address these issues, they proposed the M2M-GNN and the core idea is to ensure that heterophilic node representations are not intertwined.

### 5.4.2 Directed message passing

Edge directionality is often overlooked in graph learning, yet recent studies have shown that directed message passing can alleviate heterophily issues. AMUD [143] investigates the impact of directed topology in graph heterophily, and offers modeling guidance for digraphs from a statistical perspective. GNNDDL [144] integrates edge directionality into heterophily-specific models, and highlights

that this simple concept can significantly boost performance. Koke et al. [146] integrated spectral filters with digraphs and propose directed spectral convolution with detailed analysis from the frequency perspective. CGCN [147] introduces a new Laplacian formulation for digraphs, and leverages the path asymmetry to address graph heterophily. Rossi et al. [145] observed that edge directionality substantially increases its effective homophily under heterophily, while posing negligible impact on homophilic graphs. Inspired by this, they introduced Dir-GNN to separate aggregations for directed edges, and proved that Dir-GNN is more expressive than vanilla GNNs.

### 5.4.3 Gating mechanism

Designing gating mechanisms for heterophily in message passing is highly flexible, and can be applied within node neighborhoods, between model layers, and even at the attribute level.

–Neighborhood gating. GBK-GNN [148] first proposes to use bi-kernel gated mechanism to capture homophily and heterophily, respectively. Through gate selection for neighbor aggregation, GBK-GNN can theoretically enhance the discriminative ability for heterophily. NH-GCN [60] introduces Neighborhood Homophily (NH) and designs a NH-based gating mechanism to separate neighbors into dual channels with channel-specific weights. HES [149] proposes the Snowflake Hypothesis [212] underpinning the concept of “one node, one receptive field”, and establishes the gates for nodes to filter out local heterophilic message.

–Layer gating. Apart from filtering heterophilic neighbors, gating mechanism can also be used to customize personalized aggregation between GNN layers. For example, PriPro [150] designs the gating mechanism between GNN layers to adaptively integrate or discard inter-layer information for each node, implicitly leveraging higher-order relations to address the heterophily issue. Following PriPro, the similar idea of customizing personalized receptive field has also been applied in GNN-AS [151].

–Attribute gating. At a finer granularity, DMP [62] sets dimension-level gates on the node representations and calibrates the propagation of each attribute, thereby enhancing the discriminative ability for heterophily. Moreover, HA-GAT [152] designs the edge preference matrix to enhance the heterophily-aware aggregation at the attribute level.

–Advanced gating. Meanwhile, we also observe that some studies have utilized more advanced gating mechanisms. Considering that graph learning may benefit from different propagation rates, Rusch et al. [153] proposed a multi-rate message gating scheme called  $G^2$  that leverages graph gradients to ameliorate the heterophily issue. Inspired by social interactions, Co-GNN [154] proposes a cooperative framework where each node is viewed as a dynamic player. Co-GNN endows nodes with four states: Standard, Listen, Broadcast and Isolate, allowing nodes to choose whether to receive or broadcast messages. This advanced gating mechanism can better capture heterophilic patterns, and allows the exploration of typical topology while learning.

#### 5.4.4 Discussion

Overall, signed propagation, directed propagation, and various gating mechanisms can be seen as enhancements to the message passing mechanism. They still adhere to the message passing mechanism, and inevitably encounter issues such as over-smoothing, over-squashing, and the graph heterophily issue. Future research should focus on whether to improve model architectures to adapt to various scenarios or to develop a sufficiently powerful model that can handle diverse and complex data, such as graph heterophily.

#### 5.5 Graph transformers

Due to the attention mechanism, this powerful architecture, Graph Transformer (GT), is considered to have inherent advantages in addressing heterophily issue.

1) Polynomial view. Spectral GNNs utilize polynomial bases to approximate graph convolution and demonstrate powerful capabilities on heterophilic graphs. Inspired by that, PolyFormer [155] defines the polynomial token to perform node-wise filtering, and then designs the global attention mechanism for polynomial tokens to balance node-specific and global patterns. PolyNormer [156] introduces the first polynomial-expressive GT, where graph topology and node features are integrated into polynomial coefficients separately. With a linear local-to-global attention scheme, PolyNormer can learn high-degree equivariant polynomials and perform well under both homophily and heterophily settings.

2) Signed attention. Due to the inherent positive attention, the self-attention mechanism of GTs fails to capture high-frequency signals on graphs. To this end, SignGT [157] extends self-attention to a signed version and utilizes multi-hop topology to maintain local structural bias. The same concept has also been extended to the field of recommender systems, improved as SIGformer [158].

3) Mitigate over-globalization. While global attention partially addresses heterophily, massive distant nodes inevitably divert significant attention, regardless of actual relevance. To this end, Coarformer [159] downsamples the graph by grouping nodes into a less number of super-nodes, and then explore interactions of super-nodes and capture coarse but long-range dependencies. Similarly, Gapformer [160] introduces graph pooling to mitigate the influx of irrelevance, LGMformer [161] adopts the  $K$ -Mean method, and VCR-Graphormer [162] specifically encodes the heterophilic information and rewires graphs using virtual connections and super-nodes. Xing et al. [163] explored whether global attention benefits GTs, revealing the over-globalizing problem. They proposed CobFormer to mitigate this issue while retaining the ability to extract valuable information from distant interactions.

4) Token sequence. This category of GTs samples a token sequence for each node as input, implicitly leveraging multi-hop neighbor information. For example, ANS-GT [164] samples the token sequence from 1-hop, 2-hop and  $k$ NN neighbors or based on PageRank, and then formulates the sampling optimization as an adversary bandit problem. NTFormer [165] samples node-wise and

neighborhood-wise token sequences from attribute and topology perspectives, and then uses a Transformer-based backbone with an adaptive fusion module to learn final node representations.

5) PEs and SEs. Due to the inherent global mechanism, GTs treat inputs as fully connected graphs, making PEs and SEs crucial for supplementing structure information, especially under heterophilic settings.

–Positional encodings. Inspired by Katz index [213], DGT [128] designs the learnable PE named Katz PE to improve the expressive power of GTs by incorporating structural and semantic similarity. MpFormer [166] introduces a novel PE called HeterPos, which uses shortest path distances to define relative positions and captures feature distinctions between ego-nodes and neighbors, facilitating the incorporation of heterophilic information into GTs.

–Structural encodings. Apart from PEs, AGT [167] emphasizes the importance of SEs, introducing learnable centrality encoding and kernelized local structure encoding from node centrality and subgraph perspectives. This framework addresses the gap in SEs for learning from heterophilic graphs in GTs.

–Other explorations. To assess the effectiveness of PEs and SEs in addressing heterophily, Muller et al. [168] benchmarked multiple model variants with PEs and SEs on heterophilic datasets. They concluded that PEs and SEs lead to significant performance gains, while global attention only offers only minor improvements. Noting the rich information conveyed by graph Laplacian, SpecFormer [169] proposes the Laplacian eigenvalue encoding based on spectral theory. DeGTA [170] identifies two primary challenges in existing GTs: multi-view chaos and local-global chaos. It proposes a decoupled model that clearly defines positional, structural, and attribute information, as well as local and global interactions, integrating them in a rational manner.

Compared to the message-passing mechanism, the global modeling capability of the Transformer architecture offers a promising approach to addressing heterophily. Although Transformers have achieved significant success in NLP, questions remain about their effective transfer to graph-structured data. Therefore, investigating the relationship between GNNs and Transformers remains an interesting topic [214–216]. Additionally, GTs require substantial computational resources, and despite techniques to reduce complexity, they still face challenges with large-scale graph. Increasing parameters and computational load does not significantly enhance modeling capabilities for both homophilic and heterophilic patterns, which is a topic worth discussing.

#### 5.6 Neural diffusion process

Intuitively, heat diffusion naturally corresponds to the message passing in GNNs. Recent studies have revealed connections between diffusion dynamics and message passing [217], leading to the emergence of numerous neural diffusion GNNs [218]. Here, we focus on neural diffusion methods to address graph heterophily.

### 5.6.1 Non-smooth diffusion

Previous works [217,219,220] have extended the isotropic diffusion in GCN [15] to anisotropic diffusion to enhance expressive power. However, the heat diffusion mechanism remains susceptible to graph heterophily issues. To this end, GIND [171] extends linear isotropic diffusion to an expressive nonlinear version, thereby avoiding the aggregation of noisy information from dissimilar neighbors. PDE-GCN [172] employs an oscillation-based, non-smooth dynamical system to enhance neural diffusion. Similarly, GraphCON [173] considers a more general dynamics which combines a damped oscillation with a coupling function. Non-smooth dynamics approaches can, to some extent, alleviate over-smoothing and demonstrate superior performance in addressing heterophily.

### 5.6.2 Diffusion with external forces

However, single dynamics like diffusion or oscillation may still be limited in handling heterophily. In fact, we can enhance these dynamics by explicitly introducing external forces, such as convection, advection, and reaction. For example, CDE [174] incorporates heterophily principles by modeling information flow on nodes using convection-diffusion equation. Based on this, the homophilic “diffusion” and heterophilic “convection” are effectively combined to capture complex graph patterns. ACMP [175] further incorporates Allen-Cahn reaction term [221] into neural diffusion, forming a reaction-diffusion process. The reaction process in ACMP is implemented by repulsive forces between nodes, which can be interpreted as a negative diffusion coefficient. GREAD [176] proposes a general reaction-diffusion framework which integrates various dynamical processes, with reaction terms selected from diverse domains. For instance, Fisher reaction [222] is used to describe the spreading of biological populations; Zeldovich reaction [223] is to describe the phenomena in combustion theory; Allen-Cahn reaction [221] is also included in GREAD. ADR-GNN [177] further incorporates an advection term into reaction-diffusion process, forming advection-diffusion-reaction process. Moreover, FLODE [178] proposes the fractional heat equation for modeling anomalous diffusion. This framework, incorporating the fractional Laplacian to capture non-local interactions and establish long-range dependencies, is well-suited for heterophilic graphs. In summary, these methods introduce external force terms into diffusion process, acting conceptually as a high-pass filter to sharpen effects and address heterophily.

### 5.6.3 Diffusion with modulators

Different from imposing external forces, modulating the diffusion process, such as using a gating mechanism, can alleviate heterophily issues to some extent. [218]. For example,  $G^2$  [153] proposes that controlling diffusion speed can counteract over-smoothing and address heterophily. Therefore, this framework explicitly introduces a gating function to modulate neural diffusion. Similarly, EGLD [179] introduces a dual-channel neural diffusion framework with low-pass and high-pass filtering, and incorporates a dimension-level gate to coordinate representations. In contrast to dimension-level gating, MHKG [180] introduces a filter-level gating strategy based

on the reverse heat kernel. Specifically, MHKG integrates high-pass and low-pass filters into heat diffusion, controlling signal smoothing and sharpening via filter gate weights. Moreover, A-GGN [181] captures long-range dependencies between nodes by imposing stability and conservation constraints via anti-symmetric weight matrices in diffusion.

### 5.6.4 Other explorations

Here, we briefly outline relevant exploration works below. From the perspective of higher-order geometries, some studies [182,183] address heterophily through sheaf Laplacian [224] diffusion. Inspired by quantum diffusion, QDC [184] proposes a graph convolution kernel and improves message passing with quantum mechanics. Giovanni et al. [185] attempted to understand graph convolution via energies, and Zhang et al. [186] explored steering GNNs with pinning control [225]. Park et al. [226] found that the reverse of diffusion process produces more distinguishable representations for heterophilic graphs. FGND [187] combines latent class representation learning with graph topology to reconstruct the diffusion matrix. From the macro perspective, Li et al. [188] proposed a general diffusion framework that formally establishes the relationship between the diffusion process and many classic GNNs.

### 5.6.5 Discussion

Viewing graph message passing as a dynamical system provides a unique modeling perspective, supported by physical theory, which has propelled the development of GNNs. Currently, GNNs based on neural diffusion begin to flourish, with more studies taking graph heterophily into account. However, pitfalls of model stability and robustness, gradient vanishing and explosion are still obstacles to overcome in this field [218].

## 6 Advanced learning paradigms

With advancements in deep learning, real-world demands are driving the emergence of paradigms beyond supervised learning. This section explores self-supervised and prompt learning, which are hot topics across various research fields.

### 6.1 Self-supervised learning

Supervised learning from heterophilic graphs has achieved remarkable success as listed above. However, it requires a amount of labeled data for training, resulting in high annotation costs. Recently, Self-Supervised Learning (SSL) [45,227–229] has emerged as a novel paradigm for learning from heterophilic graphs. In general, graph SSL follow the encoder-decoder framework [45]. Given the GNN encoder  $f_\theta$  parameterized by  $\theta$  and pretext decoder  $p_\phi$  parameterized by  $\phi$ , the graph SSL loss can be formulated as:

$$\theta^*, \phi^* = \arg \min_{\theta, \phi} \mathcal{L}_{\text{SSL}}(f_\theta, p_\phi, \mathcal{G}), \quad (2)$$

where  $\mathcal{L}_{\text{SSL}}$  regularizes the output of pretext without supervision. For specific downstream, we can introduce a downstream decoder  $q_\psi$  parameterized by  $\psi$ , and then formulate the downstream task as a supervised learning task:

$$\theta^{**}, \psi^* = \arg \min_{\theta^*, \psi} \mathcal{L}_{\text{sup}}(f_{\theta^*}, q_\psi, \mathcal{G}, Y),$$

where  $\mathbf{Y}$  denotes downstream task labels and  $\mathcal{L}_{\text{sup}}$  is the supervised loss. Typically, SSL for heterophilic graphs primarily includes two forms: Contrastive Learning and Graph Auto-Encoders.

### 6.1.1 Contrastive Learning

Graph Contrastive Learning (GCL) is one of the most extensively studied SSL methods [230,231]. Its core idea is to bring positive samples closer and push negative samples further apart in the latent space. For GCL, Eq. (32) can be reformulated as:

$$\theta^*, \phi^* = \arg \min_{\theta, \phi} \mathcal{L}_{\text{SSL}}(p_{\phi}(f_{\theta}(\mathcal{G}^1), f_{\theta}(\mathcal{G}^2))),$$

where  $\mathcal{G}^1$  and  $\mathcal{G}^2$  are different views of  $\mathcal{G}$ , the pretext decoder  $p_{\phi}$  indicate the discriminator that estimates the agreement between two instances (e.g., the bilinear function or the dot product). However, most GCL methods follow the homophily assumption [232], limiting the applicability to heterophilic scenarios. Therefore, some GCL works have focused on learning from graphs with heterophily.

–Spectral filters. Given the success of spectral filters in supervised learning, HLCL [233] extends this powerful tool to GCL. This framework identifies homophilic and heterophilic views based on feature similarity, and customizes a corresponding filter for each view. After dual-view contrast, HLCL then uses the homophilic branch as the final output. GREET [234] employs an edge discriminator to distinguish heterophilic and homophilous edges, and performs dual-view contrast through low-pass and high-pass filters. To jointly learn edge distinction and node representations, it introduces an alternating training strategy for iterative optimization. Moreover, PolyGCL [235] explores the expressive power of various polynomial spectral filters in GCL, and integrates low-frequency and high-frequency information through linear combination.

–Data augmentation. Data augmentation has proven effective in enhancing GCL performance [232,236–238]. However, simple random augmentation on heterophilic graphs does not effectively enhance GCL, necessitating more advanced augmentation strategies. SimGCL [239] computes feature similarity and local feature assortativity [62] to perform pre-computed augmentation, while HGRL [240] introduces structural learning augmentation based on feature similarity. HeterGCL [241] proposes that random structure augmentation can lead to topology destruction. It introduces an adaptive aggregation strategy to connect high-order neighbors and explores structural information using an local-to-global contrastive loss. GASSER [242] injects perturbations into specific spectral frequencies, with edge selection guided by spectral hints. This augmentation technique is adaptive, controllable, and heuristically fits homophily ratios and spectrum. However, some methods question the necessity of data augmentation. For example, SP-GCL [243], AF-GCL [244], and GraphACL [55] propose augmentation-free GCL architectures, questioning the necessity of data augmentation. Whether data augmentation is necessary and how to conduct it under heterophily remain topics worth exploring.

–Multi-view contrast. Due to graph heterophily, constructing multiple views in GCL to capture different hierarchical aspects of

graphs is one solution. Khan et al. [245] employ diffusion wavelets [246] to create augmented-view graphs, and utilize a multi-view contrast for learning invariant representations. Diffusion wavelet filters capture the band-pass response of graph signals, explicitly highlighting higher-order information. Realizing the importance of node attributes in heterophilic graphs, MUSE [247] constructs semantic and contextual views to capture ego node and neighborhood information. Rather than simply combining views, MUSE fuses dual-view representations using a fusion controller, significantly enhancing performance by emphasizing semantic information.

–Capture monophily. Monophily is a common phenomenon in real-world graphs, for example, the attributes of a node’s friends tend to be similar to the attributes of that node’s other friends [248]. Intuitively, monophily describes the two-hop neighbor similarity. Inspired by this, GraphACL [55] presents a asymmetric contrastive framework, and proves that the asymmetric design can capture one-hop neighborhood context and monophily patterns. Based on GraphACL, a efficient version named GraphECL [249] has been proposed for fast inference. S3GCL [250] introduces a parameterized Chebyshev filters to enhance GCL, and establishes a cross-pass GCL objective between full-pass and biased-pass filters. Similar to GraphACL, S3GCL treats neighboring nodes as positive pairs, eliminates random augmentation, and captures monophily patterns beyond the homophily assumption.

–Homophily enhancement. Given that heterophily can limit GCL performance, we consider enhancing homophily to address this issue. NeCo [251] demonstrates that intra-class edges impact GCL performance. To address this, it integrates the positive neighbor sampling and homophily discrimination into a unified framework. By removing inter-class edges and enhancing homophily during training, GCL performance can be improved on both heterophilic and homophilic graphs. HomoGCL [252] introduces soft clustering to discover potential positive and negative samples from neighbors. HEATS [253] and ROSEN [254] both learn an affinity matrix in an unsupervised manner to capture global homophily beyond local affinity. Therefore, how to measure graph heterophily without supervision will be an interesting topic.

–Alleviate ambiguity. Research indicates that GNNs can produce ambiguous node representations due to neighborhood aggregation, particularly in heterophilic graphs [61,140]. DisamGCL [255] first combines the heterophily issue with the ambiguity of GNNs, and proposes that disambiguation can enhance GNNs in heterophilic scenarios. It introduces a memory cell to identify ambiguous nodes and disambiguates them using contrastive learning.

### 6.1.2 Graph auto-encoders

One objective of Graph Auto-Encoder (GAE) is to obtain low-dimensional embeddings via graph reconstruction for subsequent tasks. In such a case, Eq. (32) can be further derived as:

$$\theta^*, \phi^* = \arg \min_{\theta, \phi} \mathcal{L}_{\text{SSL}}(p_{\phi}(f_{\theta}(\mathcal{G})), \mathcal{G}),$$

where the reconstruction target can be node features, graph structure,

or both simultaneously. Existing GAEs are primarily designed to reconstruct direct links [256,257], implicitly following the homophily assumption and performing poorly on heterophilic graphs. Given that link reconstruction is less effective under heterophily, SELENE [258], MVGE [259], and AGCN [260] reconstruct node attributes and network structure simultaneously to adapt GAEs for heterophilic graphs. Moreover, PairE [261] introduces aggregated feature and assortativity versus reconstruction as extra pretexts to retain high-frequency signals. NWR-GAE [262] addresses the information loss from oversimplified link reconstruction, which can degrade downstream task performance. It introduces a graph decoder to reconstruct both proximity and structure through Neighborhood Wasserstein Reconstruction (NWR). By considering proximity, structure, and feature information, NWR-GAE excels in heterophilic settings. Recently, many studies attempt to combine GAE with GCL for more powerful SSL models [263–266]. However, these efforts are mostly limited to homophilic scenarios, overlooking heterophily. Integrating learning from heterophilic graphs into SSL approaches is a direction worth exploring.

### 6.1.3 Others

Xiao et al. [267] proposed that the semantic structure of graph can be decoupled into latent variables that capture different aspects, including attribute, label, and link patterns. They introduce a Decoupled Self-Supervised Learning (DSSL) framework, and models the generative process of nodes and links through latent variable, decoupling diverse semantics in neighborhoods into SSL process. By decoupling local neighborhood contexts, DSSL operates without relying on graph augmentations or downstream labels, effectively handling graph heterophily.

## 6.2 Prompt learning

Originating from NLP, the “pre-training, prompt-tuning” paradigm [48] reformulates various downstream tasks into a unified template, and designs specific prompts for downstream adaptation. Given a pre-trained GNN encoder  $f_\theta$  and a prompted adapter  $q_\psi$ , prompt learning for downstream tasks can be formulate as:

$$\psi^* = \arg \min_{\psi} \mathcal{L}_{\text{prompt}}(f_\theta, q_\psi, \mathcal{G}, \mathbf{Y}),$$

where the pre-trained  $f_\theta$  is frozen, and  $q_\psi$  is composed of light-weight parameters. Free from fine-tuning, prompt learning fully unleashes the potential of pre-trained models, where adjusting only a few parameters can achieve excellent results, even the supervisions of specific tasks is limited. Inspired by the success of prompt learning in NLP [26,268] and CV [269,270], the graph domain is beginning to shift the focus towards the “pre-training, prompting” paradigm [271,272].

Since prompt learning in the graph domain is still in its early stages, we begin by surveying existing advancements.

–Unified frameworks. The core of prompt learning is the unified task framework, which prevents “negative transfer” [273] between pre-training pretexts and downstream tasks. GPPT [274] first introduces prompt learning into graph learning, and presents a unified prompt

framework based on link prediction. Following GPPT, GraphPrompt [275,276] and Prodigy [277] adopt the unified template based on subgraph similarity calculation. SGL-PT [278] packages various graph tasks as node generation, while ProG [279,280] reformulates tasks at different levels to graph level and introduces meta-learning to boost multi-task learning. Additionally, OFA [281] aims to provide a general solution for building and training a foundation GNN model with in-context learning ability across domains.

–Prompt strategies. In addition to unified frameworks, some studies focus on designing prompt adapters specifically tailored to graph data. For example, GPF [282] injects learnable perturbations into feature space to adapt the pre-trained model to certain downstream tasks. VNT [283] inserts a set of virtual nodes into the input as prompt, while ProG [279,280] and SUPT [284] introduce virtual prompt graphs. GraphPrompt [275,276] utilizes prompt tokens to efficiently adjust model outputs, and MultiGPrompt [285] designs multiple pretext tokens to avoid negative influence of multi-tasks. Other works, including GSPF [286], IGAP [287], ULTRA-DP [288], and TGPT [289] aim to design prompts from diverse perspectives to better match downstream tasks.

–Various extensions. Recognizing the powerful capabilities of graph prompt learning, related extensions have emerged in both the graph domain and other fields. Considering heterogeneous relations in graphs, HetGPT [290] and HGPrompt [291] focus on harnessing prompt tuning in pre-trained heterogeneous GNNs. To model dynamic relations in graphs, DyGPrompt [292] introduces a prompting learning framework for dynamic graphs. Krait [293] reveals that backdoor disguise can benignly affect graph prompts, and CrossBA [294] investigates backdoor attacks in cross-context graph prompts. Not only limited to the graph domain, graph prompting techniques are making significant progress in fields such as drug prediction [295], text-attribute graphs [281,296–299], urban computing [300] and recommendation systems [301].

Despite the flourishing development of graph prompt learning, little work has focused on the impact of graph heterophily. Self-Pro [302] first pays attention to the heterophily issue in graph prompt. To accommodate heterophilic graphs, Self-Pro introduces asymmetric graph contrastive learning [55] for pre-training, and unifies the pretext and downstream tasks to avoid negative knowledge transfer [273]. Through self-adapter and semantic prompt injection, Self-Pro performs well in few-shot settings without additional parameters. ProNoG [303] revisits existing pre-training methods on heterophilic graphs and introduces some non-homophily pre-training methods. For downstream adaptation, it proposes the condition-net [304] to generate a series of prompts conditioned on heterophilic patterns. From structural perspective, PSP [305] introduces virtual nodes in the prompting phase, enabling downstream tasks to benefit from topological patterns. By decoupling node attributes from structure [51] during the pre-training phase, PSP can to some extent address the heterophily issue.

## ■ 7 Broader topics

Alongside GNN models and learning paradigms, extensible topics

related to heterophilic graphs warrant attention. In this section, we introduce broader topics to expand the research perspective.

## 7.1 Diversified learning tasks

Graph tasks can generally be categorized into three classes: node-level, edge-level, and graph-level. Most heterophilic graph studies focus on the node-level tasks, particularly node classification. With growing attention to heterophily, researchers are exploring the potential of other tasks.

### 7.1.1 Node clustering

Node clustering involves grouping similar nodes into the same category and dissimilar nodes into different categories in an unsupervised manner [306]. Deep graph learning has facilitated the development of numerous advanced graph clustering methods [256,307–309]. Similar to node classification, graph heterophily in node clustering has also drawn attention.

–Spectral filters. Spectral graph filters have demonstrated the strong capabilities in heterophilic scenarios. To capture heterophily patterns, CGC [310] automatically learns suitable filters for node clustering in heterophilic settings, extracting comprehensive information beyond low-frequency components. Moreover, AHGFC [311] designs a hybrid filter based on joint aggregation with node features and adjacency relationships to make low and high-frequency signals on graphs more distinguishable.

–Advanced reconstruction. Under unsupervised settings, most methods use GAEs to obtain node embeddings for clustering. Given the complex link patterns, reconstructing only structural links is insufficient. Therefore, many methods attempt to employ advanced reconstruction techniques. For example, DGCN [312] uses dual encoders to separately map node feature and structure into low-dimensional spaces, and then fuses them for feature reconstruction. SELENE [258] restructures both node attributes and graph structure, and utilizes a dual-view contrast to enhance the discrimination of inter-class nodes. PFGC [313] notes that potential homophily may exist in multi-hop neighborhoods, suggesting the reconstruction of high-order topology. PLCSR [314] integrates curriculum and contrastive learning to enhance GAEs for node clustering in heterophilic graphs.

–Homophily enhancement. From a data-centric perspective, HoLe [315] observes that enhancing graph homophily can significantly improve node clustering. Therefore, it proposes a structure homophily-enhanced method that removes inter-class edges or adds intra-class edges, allowing structure learning and node clustering to mutually reinforce each other.

### 7.1.2 Link prediction

The target of link prediction is to infer missing links or predict potential links based on the input graphs [316]. GNN-based link prediction methods have achieved state-of-the-art performance [256,257,317–319], but most assume graph homophily, overlooking potentially complex patterns. Zhou et al. [320] first introduced link prediction to heterophilic graphs, proposing that connected nodes

with low feature similarity may share similarities in latent factors. They proposed the DisenLink framework to model heterophilic patterns by disentangled views, and learn representations by edge factor discovery and factor-aware message passing. Sharing the same topic, GRAFF-LP [321] introduces GRAFF [185], a physics-inspired GNN, to enhance link prediction under heterophily with physics biases in message passing. Zhu et al. [322] analyzed how different link prediction encoders and decoders adapt to varying levels of heterophily, and highlights the importance of adopting learnable decoders with ego and neighbor representation separation for link prediction beyond homophily.

### 7.1.3 Graph classification

In addition to node-level tasks, such as graph classification, are closely related to heterophily. Unlike node-level tasks, graph classification requires adaptation to graphs with varying homophily ratios. Therefore, uniform aggregation and simple readout functions, such as sum, ignore graph heterophily, leading to performance degradation [323]. Inspired by H2GCN [20], IHGNN [323] separates the ego and neighbor representation and designs an adaptive aggregation strategy for different layers. For graphs with varying homophily ratios, IHGNN designs a graph-level readout function to reorganize nodes within graphs and align them across graphs. To verify the impact of heterophily on graph classification, Ding et al. [324] applied spectral GNNs to molecular and protein datasets, and found that FAGCN [137] and ChebNet [196] outperform vanilla GCN, indicating the importance of high-frequency signals.

## 7.2 Model scalability

When addressing heterophily, large-scale graphs inevitably pose challenges, making it urgent to enhance model scalability. We categorize existing methods into two types based on architecture: Message Passing Framework and Graph Transformer.

### 7.2.1 Message passing framework

The early model LINKX [52] separately embeds the adjacency and feature matrices using MLPs, and combines them through concatenation. This simple method enables mini-batch training and inference, demonstrating good performance on large-scale heterophilic datasets. Based on LINX, GLINX [325] adds PEs and monophilous propagation for further improvement. LD<sup>2</sup> [326] generates node embeddings from the adjacency and feature matrices through pre-computation, and then applies multi-hop discriminative propagation. Theoretical and empirical results show that LD<sup>2</sup> has a time complexity  $O(N)$ , linear to the number of nodes. HopGNN [327] pre-computes neighborhood information and models interactions between multi-hops, fully utilizing high-order neighbors to generalize under heterophily. AGS-GNN [328] introduces an attribute-guided sampling strategy to scale. Specifically, it selects neighbor subsets based on feature similarity and diversity with pre-computed sampling distribution. Through neighbor pruning and group aggregation, AGS-GNN performs well on both homophilic and heterophilic graphs with desired scalability.

### 7.2.2 Graph transformer

One challenge for GTs is the scalability issue due to quadratic

complexity, which is prohibitive for large graphs. NodeFormer [329] introduces the kernelized Gumbel-Softmax to reduce the complexity of all-pair message passing to linear. Thanks to efficient propagation between arbitrary node pairs, NodeFormer demonstrates its promising potential for tackling heterophily, long-range dependencies and large-scale graphs. Inspired by neural diffusion, DifFormer [330] elucidates the relationship between energy-driven diffusion and GTs. This diffusion-based Transformer also computes pairwise diffusivity with quadratic complexity. To address this, an acceleration strategy via state updating is proposed, reducing the complexity to linear. SGFormer [331] modifies the multi-head attention in GTs to a single-layer and single-head version. Combined with a simple local propagation, SGFormer retains the necessary expressiveness without PEs or SEs, data pre-processing, or extra loss functions. GOAT [332] introduces dimension reduction based on EMA  $K$ -Means, and proves that this approximate has bounded error compared to global attention. Moreover, SpikeGraphormer [333] integrates Spiking Neural Networks (SNNs) [334,335] into GTs, enabling all-pair node interactions on large-scale graphs with limited GPU memory.

### 7.3 Adversarial attack and robustness

Studying adversarial attacks [336] and model robustness [337,338] aids in understanding model principles, enhancing model performance and credibility. Currently, the community of graph learning is focusing on these topics, with heterophily as a primary concern.

- Adversarial attacks. Recent studies [339,340] show that GNNs are sensitive to adversarial attacks, where minor, intentionally introduced changes in graph structure can lead to significant performance degradation. Zhu et al. [341] first investigated the relationship between graph heterophily and GNN robustness against structural attacks. They found that on homophilic graphs, effective structural attacks lead to increased heterophily, whereas on heterophilic graphs, attacks alter the homophily level contingent on node degrees. Moreover, they proposed that separating ego and neighbor representation can improve the robustness of GNNs against adversarial attacks. Combining with spectral graph theory, Huang et al. [342] designed a mid-pass filtering GCN model named Mid-GCN, leveraging the robustness of middle-frequency signals against adversarial attacks. NSPGNN [343] employs low-pass and high-pass filters on  $k$ NN graphs to enhance model robustness. Lei et al. [344] pointed out that ignoring odd-hop neighbors improves the robustness of GNNs and present EvenNet, a simple yet effective spectral GNN based on even-polynomial filter. Qiu et al. [345] further uncovered that the predominant vulnerability on heterophilic graphs is caused by structural out-of-distribution. Therefore, they presented LHS, a framework that strengthens GNNs against attacks by refining latent homophilic structures under heterophily. Inspired by knowledge distillation [346], Deng et al. [347] introduced the MLP-to-GNN distillation

framework against structure attacks. It indicates that Prospect-MLP corrects the wrong knowledge of Prospect-GNN regardless of homophily ratios, endowing the adversarial robustness.

- Resisting label noise. In addition to adversarial attacks, data noise can also significantly degrade the generalization of GNNs. Cheng et al. [348] studied the impact of label noise in the context of arbitrary heterophily, and found that a high homophily rate can mitigate the effect of label noise on GNNs. Therefore, they proposed the R<sup>2</sup>LP framework, and iteratively performed graph reconstruction with homophily, label propagation for noisy label refinement, and high-confidence sampling over multi-rounds.
- Model privacy. The homophily assumption and message passing of GNNs, combined with publicly available user information, can provide adversaries with opportunities to infer private attributes, leading to privacy breaches [349–351]. Therefore, developing privacy-preserving GNN models to resist inference attacks is of significant importance. Yuan et al. [352] investigated Graph Privacy Leakage via Structure (GPS) and introduced heterophily metrics to quantify privacy breach risks. To counter privacy attacks, they proposed a graph data publishing method with learnable graph sampling, making sampled graphs suitable for publication. Moreover, other recent privacy-preserving methods [353,354] also focus on addressing graph heterophily.
- Model fairness. Unfortunately, the homophily assumption fails to account for local deviations where unfairness may impact certain groups, potentially amplifying unfairness [355,356]. Loveland et al. [357] first considered graph heterophily to improve model fairness, and demonstrate that heterophily-specific GNNs can address disassortative group labels and promote fairness in graphs.

### 7.4 Graph structure learning

Recent studies have sparked efforts in Graph Structure Learning (GSL), aiming to jointly learn an optimized graph structure and corresponding node representations for downstream tasks [358]. Heterophily introduces complex link patterns in graphs, posing significant challenges to GSL and attracting considerable attention.

- Edge discriminator. Identifying heterophilic or noisy edges in GSL has been shown to enhance model performance [359]. A natural approach is to train a discriminator to distinguish heterophilic edges. For example, DC-GNN [360] introduces a learnable edge classifier to transform the original heterophilic graph into its corresponding homophilic counterpart. GNN-SATA [361] designs an edge discriminator to dynamically remove or add edges, enhancing the GNN's performance on heterophilic graphs. Moreover, ECG-GNN [362] builds an edge discriminator based on the pre-trained representations, and selects top- $k$  neighbors to form a complementary graph. Parallel message passing on both original and new structures benefits downstream tasks and structural optimization.

- **Dual views.** A dual-view model, another GSL approach to address heterophily, decouples the input graph into homophily and heterophily views for further analysis. For example, GOAL [363] proposes to reconstruct graphs into dual views to complement each other. It first groups intra-class nodes together, ranks them based on feature similarity, and introduces a complemented aggregation strategy. Similarly, ATL [364] decomposes the graph into two components and extracts complementary graph signals. This dual-view framework can adaptively filter and modulate complex graph signals, which is critical to address heterophilic patterns.
- **Neighborhood similarity.** To better characterize heterophily in GSL, it is necessary to comprehensively consider neighborhood patterns for graph rewiring. DHGR [365] introduces two metrics: Neighborhood Feature Distribution and Neighborhood Label Distribution, to identify edge polarity and further guide graph rewiring. Choi et al. [366] proposed to measure the node similarity with local subgraphs based on optimal transport, better adapting to heterophilic graphs.
- **Spectral clustering.** Since spectral clustering [367] can capture long-range dependencies in graphs, some GSL methods use it for graph rewiring. For example, GCN-SL [368] proposes an efficient spectral clustering method to encode nodes, and constructs an affinity matrix based on it. By combining the affinity matrix with feature similarity, GCN-SL learns an optimized structure that enhances downstream prediction tasks. Moreover, Li et al. [59] constructed the adjacency matrix based on the result of adaptive spectral clustering, with the aim of maximizing the proposed homophilic scores.
- **Probabilistic modeling.** Based on Bayesian Inference [369], Wang et al. [370] introduced a Graph structure Estimation Networks (GEN). In addition to observed links and node features, GEN incorporates high-order neighborhood information to reduce bias, and presents a model that jointly treats this multi-view information as observations of the optimal graph. Moreover, L2A [371] uses a variational inference framework to perform maximum likelihood estimation for GNNs and optimal graph structure learning, improving applicability to heterophilic graphs.
- **Self-supervised manner.** Requiring no extensive supervision, self-supervised learning methods have become a popular paradigm in GSL. GPS [372] estimates edge likelihood based on self-supervised link prediction and rewires edges based on reconstruction uncertainty. SUBLIME [373] adopts contrastive learning, while HES-GSL [374] introduces the denoising auto-encoder [375] to obtain node representations. Both methods then conduct structural learning with homophily-enhanced self-supervision. Moreover, GSSC [376] applies structural sparsification to remove potentially uninformative or heterophilic edges, and then performs structural self-contrasting in the sparsified neighborhood.

**Benchmarks.** Recently, many GSL benchmarks including

OpenGSL [70] and GSLB [377] have been released, attracting widespread attention. These benchmarks include heterophilic graphs and evaluate the performance of existing GSL methods, highlighting graph heterophily as a hot topic in GSL.

## ■ 8 Applications

Graph heterophily is prevalent in real-world graph structures and is receiving increasing attention in application-level research. In this section, we will explore practical applications and provide detailed introductions.

### 8.1 Cyberspace security

Graph heterophily poses challenges for social network analysis in social cyberspace but also offers significant potential for cyberspace security, providing new insights and innovative approaches.

- **Anomaly detection.** The rich relations between normal and abnormal objects can be modeled as graphs, giving rise to the Graph-based Anomaly Detection (GAD). GAD faces challenges from graph heterophily, where anomalies are often overshadowed by numerous normal neighbors. Traditional GNNs, which smooth neighboring nodes, can undermine the discriminative features of anomalies.

–**Edge discrimination.** Existing GAD methods suffer from heterophily induced by hidden anomalies connected to numerous benign nodes. Therefore, SparseGAD [378] introduces a framework that sparsifies the graph structure to reduce noise and collaboratively learns node representations. It retains strong homophilic and heterophilic edges while removing irrelevant ones, then performs heterophily-aware aggregation using GPR-GNN [91]. To capture the discriminative information of anomalies, TA-Detector [379] introduces a trust classifier to distinguish between trust and distrust connections using label supervision. Meanwhile, GHRN [380] proposes a label-aware edge indicator to compute the post-aggregation similarity for pruning heterophilic edges. Moreover, TAM [381] introduces an anomaly scoring measure, local node affinity, to guide edge discrimination and iteratively removes heterophilic edges. In addition to edge pruning, HedGe [382] introduces a metric called class homophily variance, and emphasizes that generating potential homophilic edges based on this metric can enhance GAD performance.

–**Spectral view.** Tang et al. [383] rethoughted GAD from a spectral view, observing that anomalies cause a “right-shift” phenomenon, where spectral energy distribution concentrates less on low frequencies and more on high frequencies. To this end, they proposed the Beta Wavelet Graph Neural Network (BWGNN), which employs spectral and spatial localized band-pass filters to address the “right-shift” phenomenon. Moreover, AMNET [384] directly integrates high-pass and low-pass filters to establish a multi-frequency filter bank and utilizes the attention mechanism to combine them adaptively.

–**Self-supervised manner.** Self-supervised learning methods, such as GAE, can also be applied to GAD to address heterophily. For example, MELON [385] explicitly models anomaly patterns and

incorporates prior knowledge of various anomalies into an enhanced data augmentation strategy. It then proposes a dual-channel graph encoder with an edge discriminator and performs multi-view contrastive learning. Roy et al. [386] explored the feasibility of GAEs to GAD under heterophily, and find that existing GAEs excel at detecting cluster-type structural anomalies but struggle with non-cluster anomalies. Therefore, they proposed GAD-NR, an extension of NWR-GAE [262], designed specifically for heterophily, to enhance anomaly detection by leveraging its robust modeling capabilities.

–Distribution shift. Gao et al. [387] observed that heterophily and homophily distribution may shift between training and test data due to time factors and annotation preferences, a phenomenon known as structural distribution shift (SDS). Ignoring SDS can lead to poor generalization in anomaly detection. To address this, they propose the Graph Decomposition Network (GDN) with homophily guidance.

- Fraud detection. Fraud detection [388], widely used in financial, e-commerce, and insurance industries, aims to detect users exhibiting suspicious behaviors in the communication networks. Under the Graph-based Fraud Detection (GFD) setting, fraud nodes tend to interact with normal users, displaying heterophilic characteristics. While GFD and GAD overlap to some extent, this survey distinguishes between the two concepts and provides separate introductions.

–Edge discrimination. Considering the heterophilic patterns in fraud graphs, H<sup>2</sup>-FDetector [389] identifies homophilic and heterophilic connections based on label supervision, and customizes a propagation strategy for heterophilic connections. Moreover, DRAG [390] models inherent heterophily in graphs through different relation types, and then performs a relation-attentive aggregation strategy at the edge level.

–Group aggregation. Given that fraud graphs exhibit both homophily and heterophily, exploring advanced aggregation techniques is advisable to address this complexity. For example, GAGA [391] introduces a Transformer-based method for fraud detection in multi-relation graphs. To capture high-order information from distant neighbors, GAGA segregates neighboring nodes into fraudulent, benign, and unlabeled groups and performs group aggregation over multi-hops. DGA-GNN [392] employs decision tree binning encoding for feature transformation, and designs a dynamic grouping strategy to classify nodes into distinct groups for hierarchical aggregation. Moreover, PMP [391] emphasizes that the key to GFD is distinguishing inter-class neighbors rather than excluding them. It utilizes label information to discriminate neighbors and customizes distinct group aggregation.

–Spectral view. Another approach for GFD addresses graph heterophily from graph spectral perspective. SplitGNN [393] analyzes the spectral distribution under varying degrees of heterophily and observes that fraud nodes cause spectral energy to shift from low-frequency to high-frequency. Therefore, it employs an

edge classifier to split the edges, enhancing signal expressions across different frequency bands, and uses flexible band-pass spectral filters to learn node representations. Moreover, SEC-GFD [394] decomposes the graph spectrum and performs complicated message passing based on frequency bands to improve GFD performance.

- Bot detection. Social bots are automated programs that simulate human activities on social media, often used to spread false information [10], manipulate elections [395–397], leading to cybersecurity risks. These bots hide within social networks, establishing heterophilic connections with normal users, which challenges Graph-based Bot Detection (GBD).

–Edge discrimination. Interactions with real accounts lead to social networks containing massive camouflaged, heterophilic, and unreliable edges. To this end, HOVER [398] proposes that the key to GBD is identifying and reducing these edges for alleviating graph heterophily. HOVER prunes inter-class edges using heuristic criteria and further proposes an oversampling strategy for GBD. SIRAN [399] combines relation discrimination with initial residual connections to reduce neighbor noise, and enhance the capability to distinguish different types of nodes in human-bot graphs. BothH [400] constructs a combination graph integrating the original graph and feature similarity graph, and uses an edge classifier to distinguish heterophilic connections for message passing. Instead of removing heterophilic edges, increasing graph homophily is also a feasible approach. HOFA [401] introduces homophily-oriented edge augmentation, adding homophilic edges based on representation similarity to mitigate the impact of heterophily.

–Spectral view. Rethinking GBD from the spectral view, existing methods tend to focus on the low-frequency information while neglecting the high-frequency information. To address this, MSGS [402] proposes a multi-scale architecture using adaptive graph filters to intelligently exploit the low-frequency and high-frequency graph signals.

–Contrastive learning. Due to the label scarcity, graph contrastive learning, a self-supervised learning paradigm has been extended to the GBD scenario. BotSCL [403] employs data augmentation to generate diverse graph views, and designs a channel-wise, attention-free encoder to address heterophily. This framework leverages valuable label supervision to guide the encoder in aggregating class-specific information for GBD.

- Rumor detection. Rumors on social media pose an increasingly critical cybersecurity issue, potentially threatening societies [404]. Thanks to deep graph learning, graph-based rumor detection has recently garnered significant attention [404–406]. Real-world social networks exhibit low homophily, making heterophily in rumor graphs more challenging due to the involvement of diverse modalities like users, posts, links, and hashtags. To handle multi-modal heterophily, Nguyen et al. [407] proposed a Portable Heterophilic Graph Aggregation for Rumor detection On Social media (PHAROS). This framework generalizes direct relations into

multi-hop and modality-aware aggregations, using a Graph Transformer to explore global homophily. It encodes rumor graph from three perspective: features, labels, and graph topology, and reduces training workload through multi-head self-attention.

- Crime forecasting. Since nearby regions typically exhibit similar socioeconomic characteristics and crime patterns, recent solutions construct distance-based region graphs and utilize GNNs for crime forecasting. However, distance-based graphs cannot fully capture crime correlations between distant but similar regions. Motivated by this, HAGEN [408] introduces a heterophily-aware constraint to regularize the original graph. The learned graph structure in HAGEN reveals dependencies between regions in crime occurrences and captures temporal patterns from historical crime records.

## 8.2 Recommender system

Recently, graph-based learning methods have become a popular paradigm to enhance recommender systems [409–413]. Social networks are integrated into recommender systems based on the social homophily assumption [1], wherein users tend to form connections with individuals who share similar interests. Connections among like-minded users are leveraged to compensate for information scarcity in the interaction graph, thereby enhancing personalized recommendations. TGIN [414] first observes that user interests and click behaviors may exhibit heterophily in networks. SHaRe [415] calculates the preference-aware homophily ratios across real-world datasets and observes that user connections in social networks can be heterophilic. To fully leverage social connections, SHaRe adopts graph rewiring to add highly homophilic relations and remove heterophilic ones. This ensures the retention of critical social relations while introducing beneficial potential relations for recommendations. Considering the fairness of recommendation, HetroFair [416] designs a fairness-aware attention mechanism to generate fair embeddings for users and items, and assigns distinct weights to different heterophilic features during the aggregation.

## 8.3 Geographic information

The advancement of GNNs provides novel insights into geographic information research, enhancing the analysis of geographic data and geographic science. However, geographic networks exhibit more complex patterns, including heterophily.

- Urban computing. Urban graphs, widely applied in urban computing [13], where nodes represent urban objects (e.g., regions or points of interest), and edges denote urban dependencies (e.g., human mobility or road connections). Heterophily is prevalent in urban graphs, reflecting that dissimilar urban objects can be interconnected urban system. Therefore, SHGNN [14] proposes a metric named the Spatial Diversity Score to uncover the spatial heterophily. This framework employs a rotation-scaling module to cluster spatially close neighbors, and processes each group with less internal diversity. Subsequently, it introduces a heterophily-sensitive spatial interaction module to adaptively capture complex patterns within different spatial groups.

- Remote sensing. In the multimodal remote sensing context, different modality or their combinations results in distinct node types and heterophilic interactions. On remote sensing graphs [417], Label Propagation (LP) is essential to improve labeled data sparsity, enhance learning effectiveness, and support decision-making. However, traditional LP algorithms are based on the assumption of graph homophily and heterogeneity. To address this, Taelman et al. [418] design a novel LP method inspired by ZooBP [419] for multimodal remote sensing. This method performs well on fully heterogeneous graphs and incorporates both homophilic and heterophilic interactions.

## 8.4 Computer vision

The applications of graph learning in computer vision are continuously expanding, offering new insights for understanding complex interactions between vision objects in scenes [420–423]. However, the relationships between objects in the visual domain are not necessarily homophilic.

- Scene generation. The objective of Scene Graph Generation (SGG) is to detect objects and predict pairwise relations in an image [424,425]. Current SGG methods employ GNNs to model the relations, assuming homophily in scene graphs while ignoring heterophily. Inspired by learning from heterophilic graphs, HLNet [426] presents a Heterophily Learning Network to explore homophily and heterophily between objects in scene graphs. HLNet introduces an adaptive reweighting transformer, equivalent to general polynomial graph filtering [427], to handle both high-frequency and low-frequency contexts. Using a heterophily-aware message passing strategy, it fully explores the interactions between objects in complex visual scenes, accounting for both heterophily and homophily. KWGNN [428] rethinks the SGG from spectral view and demonstrates that the spectral energy shifts towards the high-frequency part as heterophily in scene graph increases. Therefore, KWGNN adaptively generates band-pass filters inspired by kumaraswamy wavelet transform [429] and integrates the filtering results to better accommodate varying levels of smoothness in scene graphs.
- Point cloud segmentation. Point cloud segmentation [430] is a crucial tasks in 3D computer vision, aiming to divide point clouds into different regions based on their attributes or functions. Recently, performing point cloud segmentation based on GNNs has become the mainstream trend [431,432]. In point cloud, some regions inevitably contain nodes from multiple categories, indicating the graph heterophily. Traditional GNN-based methods overlook crucial heterophilic information, leading to blurred segmentation boundaries. To address this, Du et al. [99] modeled the point cloud network as a homophilic-heterophilic graph and propose a graph regulation network to produce finer segmentation boundaries. They first evaluated the extent of homophily between nodes

and apply different weight strategies for homophilic and heterophilic relationships. After adaptively propagation, they designed a prototype feature extraction module to mine high-order homophily from the global prototype space. This framework theoretically constrains node representation similarity based on the degree of heterophily.

- 3D object detection. The target of 3D Object Detection (3DOD) is to accurately locate 3D objects in point clouds [433]. Chen et al. [434] noted that the relational knowledge in 3DOD should encompass both homophily and heterophily. They proposed a Joint Homophily and Heterophily Relational Knowledge Distillation (H2RKD) framework for lidar 3DOD, which models both homophilic and heterophilic relations to enhance the intra-object similarity and inter-object discrimination.

### 8.5 Biochemical research

Graph learning has already achieved significant breakthroughs in biochemistry [435–438], with modeling biochemical molecules as graph structures becoming a mainstream learning paradigm. However, recent research highlights graph heterophily as a pressing issue that needs addressing in this field.

- Drug discovery. Combination therapy [439], involving multiple drugs to improve clinical outcomes, has demonstrated advantages over monotherapy. To avoid costly high-throughput testing, researchers have established drug-drug networks to explore potential combinations and accelerate drug discovery. Chen et al. [440] found that drug pairs with complementary exposure to the disease tend to be effective combinations, which is consistent with the principle of non-overlapping pharmacology [441]. In other words, heterophily is widely present in drug-drug networks used for combination therapy. Chen et al. [442] confirmed that drug-drug networks exhibit heterophily and sparseness, which limits the effectiveness of homophily-based GNNs. Therefore, they introduced DCMGCN, a framework that simultaneously optimizes drug representations and predictions. Specifically, DCMGCN expands the local neighborhood of drug nodes, searching globally for distant but related nodes to enhance learning process. Moreover, Liu et al. [443] found that the heterophily issue is also widespread in drug repositioning [444–446]. They further proposed a Structure-enhanced Line Graph Convolutional Network (SLGCN) for learning from drug-disease pairs. It utilizes the transformation of line graphs to capture graph topology, and assigns appropriate weights for homophilic and heterophilic structures in message passing with a gating mechanism.
- Molecule generation. Conditional molecule generation is crucial for materials discovery and drug design, and its integration with deep learning is becoming increasingly seamless [447–449]. Existing methods often assume strong homophily in molecular structures while overlooking heterophily between dissimilar atoms. To address this,

HTFlows [450] proposes a flow-based method for conditional molecule generation. By leveraging multiple interactive flows, it effectively captures both homophilic and heterophilic patterns, providing a more versatile representation of the balance between molecular affinities and repulsions.

- Neuroscience. Modeling brain networks is crucial for the early diagnosis of neurodegenerative diseases, and deep learning methods in this field are rapidly advancing [451–455]. With dissimilar regions of interest physically connecting, brain networks exhibit heterophily, making modeling challenging due to the interplay between homophily and heterophily. To address this, AGT [12] introduces spectral node-wise filters based on wavelet transform [456,457] to adaptively capture localized homophily and heterophily. Considering sequential variations in progressive degeneration, AGT uses temporal regularization to control distances between diagnostic groups in the latent space, ensuring effective capture of temporal dynamics.

### 8.6 Software engineering

As GNNs are widely applied in the intelligent process of software engineering, the issue of graph heterophily also emerges.

- Program management. Modeling software programs as graphs and using graph learning for management have been widely adopted in software engineering [458]. To handle heterophilic interactions between programs, HAGCN [459] introduces a subtraction operation into GNNs to separate dissimilar nodes in the representation space. It also separately encodes each edge type and employs a global relation-aware attention mechanism to aggregate messages from different edge types, enhancing homophily-oriented interactions.
- Defect localization. In the current booming open-source ecosystem, a myriad of new developers register on GitHub, and millions of new code repositories are established. Frequent code changes impose higher demands on Just-In-Time (JIT) defect localization [460–462]. Zhang et al. [463] observed that real-world code graphs exhibit very low homophily and proposed to use heterophilic methods to model this pattern. Therefore, they used FAGCN [137], designed for heterophilic graphs, to extract both homophilic and heterophilic patterns from code graphs, and enhance defective file prediction through contrastive learning.

## 9 Future directions

After reviewing recent advances of learning from heterophilic graphs, several challenges and promising directions for further exploration remain. In this section, we analyze these directions to provide insights for future research.

### 9.1 More complex scenarios

Currently, most studies on graph heterophily focus on static, homogeneous graphs, neglecting the complex relationships in real-world scenarios, such as graph heterogeneity, dynamic changes, and higher-order connections. Addressing graph heterophily presents

greater challenges in these contexts but also offers insights into understanding complex patterns under real-world settings.

- **Heterogeneous graphs.** Heterogeneous graphs, also known as heterogeneous information networks, are network structures that comprise multiple types of nodes and edges [464]. The multitude of node and edge types poses great challenges to graph learning and has spurred the development of a series of Heterogeneous Graph Neural Networks [465–471]. Recent studies [472,473] note the heterophily in heterogeneous graphs and attempt to measure this property. Meanwhile, self-supervised learning on heterogeneous graphs has also taken the heterophily issue into account [474]. Recently, the release of new benchmark [475] marks that this field will garner more widespread attention.
- **Temporal graphs.** Temporal graphs formalize evolving graphs with dynamically changing nodes, edges, and features, and GNNs are powerful tools for analyzing temporal graphs [476,477]. For example, Greto [478] empirically and theoretically elucidates the topology-task discordance and explains why homophily-based GNNs fail on dynamic graphs. Additionally, the analysis of heterophily in temporal graphs remains limited, offering ample room for future exploration.
- **Hypergraphs.** Hypergraphs generalize graphs by allowing edges to connect any number of vertices, enabling high-order interactions [479]. Heterophily is more common in hypergraphs than in simple graphs [480]. Wang et al. [481] first explored this issue and propose a hypergraph-based framework, shows significant superiority in processing heterophily patterns. Further follow-ups have been made in [482,483], and we look forward to deeper research on the heterophily issue in hypergraphs.

## 9.2 Deeper theoretical insights

Deeper theoretical insights can significantly inspire learning from graphs with heterophily. Thus, we organize related theories and highlight promising perspectives for theoretical understanding.

- **Model performance.** Intuitively, heterophily can degrade GNN performance, while strong homophily tends to improve it. However, Ma et al. [68] first revealed that strong homophily is not always necessary, and vanilla GCN can perform well on heterophilic graphs under certain conditions. Luan et al. [61,65] demonstrated that heterophily is not always detrimental to model, and mid-level homophily is the main culprit of bad performance, a phenomenon termed the “mid-homophily pitfall”. In addition to heterophily distribution [484], some studies provide crucial insights into GNN performance related to conditional shift [485], structural disparity [486,487], and other factors [64,488,489]. We expect more theoretical research on the relationship between model performance and heterophily, laying a solid theoretical foundation for the development of this field.
- **Over-smoothing & over-squashing.** Two notorious issues regarding GNNs are over-smoothing and over-squashing.

Over-smoothing refers to the phenomenon where node representations become gradually similar as the depth of GNNs gets deeper, leading to a loss of discriminative power [21,490–492]. Yan et al. [140] first proposed a unified perspective on over-smoothing and heterophily, suggesting that addressing graph heterophily can also mitigate the over-smoothing issue [182,226,493]. Another issue of GNNs is over-squashing, where information from distant neighbors is compressed into fixed-length vectors, leading to a loss of messages from high-order neighbors [494]. Rubin et al. [495] first established a theoretical framework to understand the combined effect of heterophily and over-squashing. Notably, Huang et al. [107] introduced UniFilter, a polynomial filter-based GNN that addresses heterophily from a spectral perspective, effectively preventing over-smoothing and mitigating over-squashing. As three major pitfalls of GNNs, we anticipate more research exploring the connections among heterophily, over-smoothing, and over-squashing, and developing comprehensive solutions [496].

- **Other discoveries.** In addition to the above, other theoretical analyses regarding heterophily are also worth attention. Yang et al. [497] study the training dynamics of GNNs in the function space, and establish a correlation between generalization and homophily. They derived a data-dependent generalization bound that highly depends on heterophily. MGNN [498] provides a comprehensive analysis of the universality of spatial GNNs from geometric and physical perspectives. Inspired by the Distance Geometry Problem, the proposed framework effectively handles both homophilic and heterophilic graphs. Inspired by the statistical physics and random matrix theory, Shi et al. [499] explored the double descent phenomenon in GNNs, and further explained its relationship with heterophily.

## 9.3 Broader learning scopes

We need not only deeper theoretical insights but also encourage researchers to explore a broader range of areas, including diverse learning tasks, settings, paradigms, real-world applications, and more.

- **More learning tasks.** Most existing studies on heterophilic graphs focus primarily on node classification, with research on edge-level and graph-level tasks still in early stages. Beyond these, we encourage exploring more diverse tasks that consider heterophilic scenarios, such as graph generation [500,501], graph condensation [502–504], critical node identification [505], and influence maximization [506–508].
- **More task settings.** Node classification is the most studied topic in heterophilic graphs, but existing works primarily focus on semi-supervised settings, neglecting complex real-world scenarios. Meanwhile, existing models suffer significant performance degradation when labels are extremely limited. To bridge this gap, further exploration of models designed for few-shot settings [509,510] or settings

with very limited labels [511] is needed. Moreover, imbalanced class settings [512,513] and long-tail settings [514], which affect homophilic graphs, should also be considered under heterophily. Additionally, we can rethink heterophilic graph learning from the perspective of weakly-supervised settings [515,516], including missing or noisy structures, features, and labels.

- More learning paradigms. In the field of learning from heterophilic graphs, supervised learning dominates, while self-supervised learning is rapidly advancing and prompt learning remains in its early stages. Beyond these paradigms, graph heterophily is also expanding into areas such as reinforcement learning [517], knowledge distillation [518,519], graph self-training [520], and neural architecture search [521–524]. We are also curious about whether graph heterophily can spark synergies with meta learning [525], multi-task learning [526], positive unlabeled learning [527], and other paradigms.
- More applications. As outlined above, current applications of heterophilic graphs mainly focus on social networks. In fields like biology, chemistry, and geography, the understanding of graph heterophily is still superficial, limited to structural inconsistencies. Therefore, there is an urgent need to deepen the study of heterophily in these fields to better address specific application requirements. Moreover, we are eager to see further integration of graph heterophily with fields such as Agents [528,529], AI4Finance [530], and AI4Science [531]. Domains like task planning [532,533], portfolio management [534], climate change [535], tectonic movements [536,537], and epidemic modeling [538] exhibit typical heterophily in data structure, making them prime candidates for such integration. Compared to the outstanding performance of Large Language Models (LLMs) in text and dialogue generation, the “killer application” for learning from graphs with heterophily has yet to be discovered.

#### 9.4 Advanced learning architectures

The design of backbone architectures is crucial for learning from heterophilic graphs. In the preceding text, we detail two such architectures: the Message Passing Framework and Graph Transformer. Here, we explore prospects for advanced architectures in this domain.

- Advanced backbones. Inspired by State Space Models (SSMs) [539], a novel architecture named Mamba [540] has emerged as a strong competitor to Transformer, offering effective and efficient modeling of long-range dependencies in sequential data. Inspired by this, GMN [541] and Graph-Mamba [170] pioneer the adaptation of SSMs to graph-structured data, achieving performance comparable to GTs. The state selection mechanism, comparable to global attention, certainly captures long-range dependencies and address graph heterophily issues. However, naturally adapting SSMs to graph-structured data requires further exploration.

Additionally, we hope to see other advanced architectures, such as Kolmogorov Arnold Network [542–544], explored in the context of graph heterophily.

- LLM for heterophily. With great advancements in LLMs, enhancing graph learning using the extensive knowledge within LLMs is a promising direction. Existing methods have demonstrated the capability of LLMs to empower learning for Text-Attributed Graphs [298,545–550]. Recent advances have primarily focused on homophilic graphs, with graph heterophily remaining unexplored. Notably, LLM4HeG [551] pioneers the integration of LLMs into learning from heterophilic graphs, offering new insights for future development. Key questions that have become current research hotspots include: Can LLMs effectively identify graph heterophily? How can we reduce costs and efficiently use LLMs to enhance heterophilic graph learning? Should we follow paradigms like LLM4GNN and GNN4LLM, using LLMs to replace or integrate with GNNs in a unified architecture under heterophilic settings [216,552,553]?
- Graph foundation models. The goal of Graph Foundation Models (GFMs) [553] is to develop graph models trained on vast, diverse datasets to enhance applicability across various tasks and domains, emerging as a significant focus in the graph domain. Current works [554–557] indicate that this field is still in its early stages. Remarkably, AnyGraph [558] employs a Mixture-of-Experts [202] architecture to effectively manage cross-domain distribution shifts. This is the first attempt to enable graph models to exhibit scaling law behavior [559], where model performance improves with more data and parameters. Despite this progress, graph heterophily remains a challenge for GFMs with strong generalization capabilities. We call for greater attention to graph heterophily in GFM studies.

#### ■ 10 Conclusion

In this paper, we presented a comprehensive survey of the benchmark datasets, GNN models, learning paradigms, real-world applications, and future directions for heterophilic graphs. Through a detailed overview and an in-depth analysis of recent advances, we aim to provide inspiration and insights for this field, thereby promoting further development of learning from graphs with heterophily.

#### ■ Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 62202172).

#### ■ Competing interests

The authors declare that they have no competing interests or financial conflicts to disclose.

#### ■ Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as

you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## ■ References

- [1] McPherson M, Smith-Lovin L, Cook J M. Birds of a feather: homophily in social networks. *Annual Review of Sociology*, 2001, 27: 415–444
- [2] Kimura D, Hayakawa Y. Coevolutionary networks with homophily and heterophily. *Physical Review E*, 2008, 78(1): 016103
- [3] Zheng X, Wang Y, Liu Y, Li M, Zhang M, Jin D, Yu P S, Pan S. Graph neural networks for graphs with heterophily: a survey. 2022, arXiv preprint arXiv: 2202.07082
- [4] Zhu J, Yan Y, Heimann M, Zhao L, Akoglu L, Koutra D. Heterophily and graph neural networks: past, present and future. *IEEE Data Engineering Bulletin*, 2023, 46(2): 12–34
- [5] Luan S, Hua C, Lu Q, Ma L, Wu L, Wang X, Xu M, Chang X W, Precup D, Ying R, Li S Z, Tang J, Wolf G, Jegelka S. The heterophilic graph learning handbook: benchmarks, models, theoretical analysis, applications and challenges. 2024, arXiv preprint arXiv: 2407.09618
- [6] Brachten F, Stieglitz S, Hofeditz L, Kloppenborg K, Reimann A. Strategies and influence of social bots in a 2017 German state election - a case study on twitter. 2017, arXiv preprint arXiv: 1710.07562
- [7] Shao C, Ciampaglia G L, Varol O, Flammini A, Menczer F. The spread of fake news by social bots. 2017, arXiv preprint arXiv: 1707.07592
- [8] Shao C, Ciampaglia G L, Varol O, Yang K C, Flammini A, Menczer F. The spread of low-credibility content by social bots. *Nature Communications*, 2018, 9(1): 4787
- [9] Doshi J, Novacic I, Fletcher C, Borges M, Zhong E, Marino M C, Gan J, Mager S, Sprague D, Xia M. Sleeper social bots: a new generation of AI disinformation bots are already a political threat. 2024, arXiv preprint arXiv: 2408.12603
- [10] Wan H, Luo M, Ma Z, Dai G, Zhao X. How do social bots participate in misinformation spread? A comprehensive dataset and analysis. 2024, arXiv preprint arXiv: 2408.09613
- [11] Lynn C W, Bassett D S. The physics of brain network structure, function and control. *Nature Reviews Physics*, 2019, 1(5): 318–332
- [12] Cho H, Sim J, Wu G, Kim W H. Neurodegenerative brain network classification via adaptive diffusion with temporal regularization. In: *Proceedings of the 41st International Conference on Machine Learning*. 2024
- [13] Zheng Y, Capra L, Wolfson O, Yang H. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2014, 5(3): 38
- [14] Xiao C, Zhou J, Huang J, Xu T, Xiong H. Spatial heterophily aware graph neural networks. In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023, 2752–2763
- [15] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. In: *Proceedings of the 5th International Conference on Learning Representations*. 2017
- [16] Hamilton W L, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, 1025–1035
- [17] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. 2017, arXiv preprint arXiv: 1710.10903
- [18] Gilmer J, Schoenholz S S, Riley P F, Vinyals O, Dahl G E. Neural message passing for quantum chemistry. In: *Proceedings of the 34th International Conference on Machine Learning*. 2017, 1263–1272
- [19] Gilmer J, Schoenholz S S, Riley P F, Vinyals O, Dahl G E. Message passing neural networks. In: Schütt K T, Chmiela S, von Lilienfeld O A, Tkatchenko A, Tsuda K, Müller K R, eds. *Machine Learning Meets Quantum Physics*. Cham: Springer, 2020, 199–214
- [20] Zhu J, Yan Y, Zhao L, Heimann M, Akoglu L, Koutra D. Beyond homophily in graph neural networks: current limitations and effective designs. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 2020, 653
- [21] Li Q, Han Z, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning. In: *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. 2018, 433
- [22] Alon U, Yahav E. On the bottleneck of graph neural networks and its practical implications. In: *Proceedings of the 9th International Conference on Learning Representations*. 2021
- [23] Corso G, Cavalleri L, Beaini D, Liò P, Veličković P. Principal neighbourhood aggregation for graph nets. In: *Proceedings of the 34th International Conference on Neural Information Processing System*. 2020, 1112
- [24] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, Polosukhin I. Attention is all you need. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, 6000–6010
- [25] Devlin J, Chang M W, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2019, 4171–4186
- [26] Brown T, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D M, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D. Language models are few-shot learners. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 2020, 159
- [27] Patwardhan N, Marrone S, Sansone C. Transformers in the real world: a survey on NLP applications. *Information*, 2023, 14(4): 242
- [28] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Hounsby N. An image is worth 16x16 words: transformers for image

- recognition at scale. In: Proceedings of the 9th International Conference on Learning Representations. 2021
- [29] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021, 9992–10002
- [30] Han K, Wang Y, Chen H, Chen X, Guo J, Liu Z, Tang Y, Xiao A, Xu C, Xu Y, Yang Z, Zhang Y, Tao D. A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(1): 87–110
- [31] Khan S, Naseer M, Hayat M, Zamir S W, Khan F S, Shah M. Transformers in vision: a survey. *ACM Computing Surveys (CSUR)*, 2022, 54(10s): 200
- [32] Min E, Chen R, Bian Y, Xu T, Zhao K, Huang W, Zhao P, Huang J, Ananiadou S, Rong Y. Transformer for graphs: an overview from architecture perspective. 2022, arXiv preprint arXiv: 2202.08455
- [33] Shehzad A, Xia F, Abid S, Peng C, Yu S, Zhang D, Verspoor K. Graph transformers: a survey. 2024, arXiv preprint arXiv: 2407.09777
- [34] Ying C, Cai T, Luo S, Zheng S, Ke G, He D, Shen Y, Liu T Y. Do transformers really perform bad for graph representation? In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 2212
- [35] Dwivedi V P, Luu A T, Laurent T, Bengio Y, Bresson X. Graph neural networks with learnable structural and positional representations. In: Proceedings of the 10th International Conference on Learning Representations. 2022
- [36] Bouritsas G, Frasca F, Zafeiriou S, Bronstein M M. Improving graph neural network expressivity via subgraph isomorphism counting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(1): 657–668
- [37] Kreuzer D, Beaini D, Hamilton W, Létourneau V, Tossou P. Rethinking graph transformers with spectral attention. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 1654
- [38] Ma L, Lin C, Lim D, Romero-Soriano A, Dokania P K, Coates M, Torr P H S, Lim S N. Graph inductive biases in transformers without message passing. In: Proceedings of the 40th International Conference on Machine Learning. 2023, 971
- [39] Zhang K, Zhu Y, Wang J, Zhang J. Adaptive structural fingerprints for graph attention networks. In: Proceedings of the 8th International Conference on Learning Representations. 2020
- [40] Chung F R K. *Spectral Graph Theory*. Providence: American Mathematical Society, 1997
- [41] Luan S, Zhao M, Hua C, Chang X W, Precup D. Complete the missing half: augmenting aggregation filtering with diversification for graph convolutional networks. 2020, arXiv preprint arXiv: 2008.08844
- [42] Daković M, Stanković L, Sejdović E. Local smoothness of graph signals. *Mathematical Problems in Engineering*, 2019, 2019(1): 3208569
- [43] Liao N, Liu H, Zhu Z, Luo S, Lakshmanan L V S. Benchmarking spectral graph neural networks: a comprehensive study on effectiveness and efficiency. 2024, arXiv preprint arXiv: 2406.09675
- [44] Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks? In: Proceedings of the 7th International Conference on Learning Representations. 2019
- [45] Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, Tang J. Self-supervised learning: generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(1): 857–876
- [46] Liu Y, Jin M, Pan S, Zhou C, Zheng Y, Xia F, Yu P S. Graph self-supervised learning: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(6): 5879–5900
- [47] Hu W, Liu B, Gomes J, Zitnik M, Liang P, Pande V S, Leskovec J. Strategies for pre-training graph neural networks. In: Proceedings of the 8th International Conference on Learning Representations. 2020
- [48] Liu P, Yuan W, Fu J, Jiang Z, Hayashi H, Neubig G. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 2023, 55(9): 195
- [49] Rožemberczki B, Allen C, Sarkar R. Multi-scale attributed node embedding. *Journal of Complex Networks*, 2021, 9(2): cnab014
- [50] Pei H, Wei B, Chang K C C, Lei Y, Yang B. Geom-GCN: geometric graph convolutional networks. In: Proceedings of the 8th International Conference on Learning Representations. 2020
- [51] Lim D, Li X, Hohne F, Lim S N. New benchmarks for learning on non-homophilous graphs. 2021, arXiv preprint arXiv: 2104.01404
- [52] Lim D, Hohne F, Li X, Huang S L, Gupta V, Bhalerao O, Lim S N. Large scale learning on non-homophilous graphs: new benchmarks and strong simple methods. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 1598
- [53] Platonov O, Kuznedelev D, Diskin M, Babenko A, Prokhorenkova L. A critical look at the evaluation of GNNs under heterophily: are we really making progress? In: Proceedings of the 11th International Conference on Learning Representations. 2023
- [54] Luan S, Lu Q, Hua C, Wang X, Zhu J, Chang X W, Wolf G, Tang J. Are heterophily-specific GNNs and homophily metrics really effective? Evaluation pitfalls and new benchmarks. 2024, arXiv preprint arXiv: 2409.05755
- [55] Xiao T, Zhu H, Chen Z, Wang S. Simple and asymmetric graph contrastive learning without augmentations. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 710
- [56] Cavallo A, Grohnfeldt C, Russo M, Lovisotto G, Vassio L. 2-hop neighbor class similarity (2NCS): a graph structural metric indicative of graph neural network performance. 2022, arXiv preprint arXiv: 2212.13202
- [57] Platonov O, Kuznedelev D, Babenko A, Prokhorenkova L. Characterizing graph datasets for node classification: homophily-heterophily dichotomy and beyond. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 25
- [58] Newman M E J. Mixing patterns in networks. *Physical Review E*, 2003, 67(2): 026126
- [59] Li S, Kim D, Wang Q. Restructuring graph for higher homophily via adaptive spectral clustering. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence. 2023, 8622–8630
- [60] Gong S, Zhou J, Xie C, Xuan Q. Neighborhood homophily-based graph convolutional network. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023, 3908–3912
- [61] Luan S, Hua C, Lu Q, Zhu J, Zhao M, Zhang S, Chang X W, Precup D. Revisiting heterophily for graph neural networks. In:

- Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 100
- [62] Yang L, Li M, Liu L, Niu B, Wang C, Cao X, Guo Y. Diverse message passing for attribute with heterophily. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 363
- [63] Jin D, Wang R, Ge M, He D, Li X, Lin W, Zhang W. RAW-GNN: RANdom walk aggregation based graph neural network. In: Proceedings of the 31st International Joint Conference on Artificial Intelligence. 2022, 2108–2114
- [64] Lee S Y, Kim S, Bu F, Yoo J, Tang J, Shin K. Feature distribution on graph topology mediates the effect of graph convolution: homophily perspective. In: Proceedings of the 41st International Conference on Machine Learning. 2024
- [65] Luan S, Hua C, Xu M, Lu Q, Zhu J, Chang X W, Fu J, Leskovec J, Precup D. When do graph neural networks help with node classification? Investigating the homophily principle on node distinguishability. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 1251
- [66] Ojha I, Bose K, Das S. Affinity-based homophily: can we measure homophily of a graph without using node labels? In: Proceedings of the 12th International Conference on Learning Representations. 2024
- [67] Zheng Y, Luan S, Chen L. What is missing in homophily? Disentangling graph homophily for graph neural networks. 2024, arXiv preprint arXiv: 2406.18854
- [68] Ma Y, Liu X, Shah N, Tang J. Is homophily a necessity for graph neural networks? In: Proceedings of the 10th International Conference on Learning Representations. 2022
- [69] Huang X, Li J, Hu X. Label informed attributed network embedding. In: Proceedings of the 10th ACM International Conference on Web Search and Data Mining. 2017, 731–739
- [70] Zhou Z Y, Zhou S, Mao B, Zhou X, Chen J, Tan Q, Zha D, Feng Y, Chen C, Wang C. OpenGSL: a comprehensive benchmark for graph structure learning. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 787
- [71] Tang J, Sun J, Wang C, Yang Z. Social influence analysis in large-scale networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2009, 807–816
- [72] García-Plaza A P, Fresno V, Unanue R M, Zubiaga A. Using fuzzy logic to leverage HTML markup for web page representation. IEEE Transactions on Fuzzy Systems, 2017, 25(4): 919–933
- [73] Dwivedi V P, Joshi C K, Luu A T, Laurent T, Bengio Y, Bresson X. Benchmarking graph neural networks. The Journal of Machine Learning Research, 2023, 24(1): 43
- [74] Traud A L, Mucha P J, Porter M A. Social structure of facebook networks. Physica A: Statistical Mechanics and its Applications, 2012, 391(16): 4165–4180
- [75] Jure L, Andrej J. Stanford large network dataset collection. See Snap.stanford.edu/data website, 2014
- [76] Lim D, Benson A R. Expertise and dynamics within crowdsourced musical knowledge curation: a case study of the genius platform. In: Proceedings of the 15th International AAAI Conference on Web and Social Media. 2021, 373–384
- [77] Rozemberczki B, Sarkar R. Twitch gamers: a dataset for evaluating proximity preserving and structural role-based node embeddings. 2021, arXiv preprint arXiv: 2101.03091
- [78] Rozemberczki B, Sarkar R. Characteristic functions on graphs: birds of a feather, from statistical descriptors to parametric models. In: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020, 1325–1334
- [79] Hu W, Fey M, Zitnik M, Dong Y, Ren H, Liu B, Catasta M, Leskovec J. Open graph benchmark: datasets for machine learning on graphs. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020, 1855
- [80] Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. 2005, 177–187
- [81] Lhoest Q, del Moral A V, Jernite Y, Thakur A, von Platen P, Patil S, Chaumond J, Drame M, Plu J, Tunstall L, Davison J, Šaško M, Chhablani M, Malik B, Brandeis S, Le Scao T, Sanh V, Xu C, Patry N, McMillan-Major A, Schmid P, Gugger S, Delangue C, Matussière T, Debut L, Bekman S, Cistac P, Goehringer T, Mustar V, Lagunas F, Rush A, Wolf T. Datasets: a community library for natural language processing. In: Proceedings of 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2021, 175–184
- [82] Luo Y, Shi L, Wu X M. Classic GNNs are strong baselines: reassessing GNNs for node classification. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. 2024
- [83] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 2014, 15(1): 1929–1958
- [84] Shu J, Xi B, Li Y, Wu F, Kamhoua C, Ma J. Understanding dropout for graph neural networks. In: Proceedings of Companion Proceedings of the Web Conference 2022. 2022, 1128–1138
- [85] Cai T, Luo S, Xu K, He D, Liu T Y, Wang L. GraphNorm: a principled approach to accelerating graph neural network training. In: Proceedings of the 38th International Conference on Machine Learning. 2021, 1204–1215
- [86] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, 770–778
- [87] Li G, Müller M, Thabet A, Ghanem B. DeepGCNs: can GCNs go as deep as CNNs? In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019, 9266–9275
- [88] Li G, Xiong C, Thabet A, Ghanem B. DeeperGCN: all you need to train deeper GCNs. 2020, arXiv preprint arXiv: 2006.07739
- [89] Xu K, Li C, Tian Y, Sonobe T, Kawarabayashi K I, Jegelka S. Representation learning on graphs with jumping knowledge networks. In: Proceedings of the 35th International Conference on Machine Learning. 2018, 5449–5458
- [90] Chen M, Wei Z, Huang Z, Ding B, Li Y. Simple and deep graph convolutional networks. In: Proceedings of the 37th International Conference on Machine Learning. 2020, 1725–1735
- [91] Chien E, Peng J, Li P, Milenkovic O. Adaptive universal generalized PageRank graph neural network. In: Proceedings of the 9th International Conference on Learning Representations. 2021

- [92] Chanpuriya S, Musco C. Simplified graph convolution with heterophily. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 1971
- [93] He M, Wei Z, Wen J R. Convolutional neural networks on graphs with chebyshev approximation, revisited. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 527
- [94] Guo Y, Wei Z. Clenshaw graph neural networks. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023, 614–625
- [95] He M, Wei Z, Huang Z, Xu H. BernNet: learning arbitrary graph spectral filters via bernstein approximation. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 1091
- [96] Chen J, Xu L. Improved modeling and generalization capabilities of graph neural networks with Legendre polynomials. IEEE Access, 2023, 11: 63442–63450
- [97] Wang T, Jin D, Wang R, He D, Huang Y. Powerful graph convolutional networks with adaptive propagation mechanism for homophily and heterophily. In: Proceedings of the 36th AAAI Conference on Artificial Intelligence. 2022, 4210–4218
- [98] Guo Y, Wei Z. Graph neural networks with learnable and optimal polynomial bases. In: Proceedings of the 40th International Conference on Machine Learning. 2023, 12077–12097
- [99] Du Z, Liang J, Liang J, Yao K, Cao F. Graph regulation network for point cloud segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 7940–7955
- [100] Li M, Guo X, Wang Y, Wang Y, Lin Z. G<sup>2</sup>CN: graph gaussian convolution networks with concentrated graph filters. In: Proceedings of the 39th International Conference on Machine Learning. 2022, 12782–12796
- [101] Ekbote C, Deshpande A P, Iyer A, Bairi R, Sellamanickam S. FiGURe: simple and efficient unsupervised node representations with filter augmentations. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 1538
- [102] Geng H, Chen C, He Y, Zeng G, Han Z, Chai H, Yan J. Pyramid graph neural network: a graph sampling and filtering approach for multi-scale disentangled representations. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023, 518–530
- [103] Lu Q, Zhu J, Luan S, Chang X W. Flexible diffusion scopes with parameterized Laplacian for heterophilic graph learning. 2024, arXiv preprint arXiv: 2409.09888
- [104] Li B, Pan E, Kang Z. PC-Conv: unifying homophily and heterophily with two-fold filtering. In: Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence. 2024, 13437–13445
- [105] Xu J, Dai E, Luo D, Zhang X, Wang S. Shape-aware graph spectral learning. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024, 2692–2701
- [106] Huang K, Cao W, Ta H, Xiao X, Liò P. Optimizing polynomial graph filters: a novel adaptive Krylov subspace approach. In: Proceedings of the ACM Web Conference 2024. 2024, 1057–1068
- [107] Huang K, Wang Y G, Li M, Liò P. How universal polynomial bases enhance spectral graph neural networks: heterophily, over-smoothing, and over-squashing. In: Proceedings of the 41st International Conference on Machine Learning. 2024
- [108] Han H, Li J, Huang W, Tang X, Lu H, Luo C, Liu H, Tang J. Node-wise filtering in graph neural networks: a mixture of experts approach. 2024, arXiv preprint arXiv: 2406.03464
- [109] Zheng S, Zhu Z, Liu Z, Li Y, Zhao Y. Node-oriented spectral filtering for graph neural networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(1): 388–402
- [110] Abu-El-Haija S, Perozzi B, Kapoor A, Alipourfard N, Lerman K, Harutyunyan H, Ver Steeg G, Galstyan A. MixHop: higher-order graph convolutional architectures via sparsified neighborhood mixing. In: Proceedings of the 36th International Conference on Machine Learning. 2019, 21–29
- [111] Jin D, Yu Z, Huo C, Wang R, Wang X, He D, Han J. Universal graph convolutional networks. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 815
- [112] Maurya S K, Liu X, Murata T. Simplifying approach to node classification in graph neural networks. Journal of Computational Science, 2022, 62: 101695
- [113] Xu J, Dai E, Zhang X, Wang S. HP-GMN: graph memory networks for heterophilous graphs. In: Proceedings of 2022 IEEE International Conference on Data Mining (ICDM). 2022, 1263–1268
- [114] Zhao Z, Yang Z, Li C, Zeng Q, Guan W, Zhou M. Dual feature interaction-based graph convolutional network. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(9): 9019–9030
- [115] Choi S, Kim G, Yun S Y. Node mutual information: enhancing graph neural networks for heterophily. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023
- [116] Wang Y, Derr T. Tree decomposed graph neural network. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021, 2040–2049
- [117] Song Y, Zhou C, Wang X, Lin Z. Ordered GNN: ordering message passing to deal with heterophily and over-smoothing. In: Proceedings of the 11th International Conference on Learning Representations. 2023
- [118] Sun Y, Deng H, Yang Y, Wang C, Xu J, Huang R, Cao L, Wang Y, Chen L. Beyond homophily: structure-aware path aggregation graph neural network. In: Proceedings of the 31st International Joint Conference on Artificial Intelligence. 2022, 2233–2240
- [119] Zhou J, Xie C, Gong S, Qian J, Yu S, Xuan Q, Yang X. PathMLP: smooth path towards high-order homophily. Neural Networks, 2024, 180: 106650
- [120] Jin W, Derr T, Wang Y, Ma Y, Liu Z, Tang J. Node similarity preserving graph convolutional networks. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 2021, 148–156
- [121] Suresh S, Budde V, Neville J, Li P, Ma J. Breaking the limit of graph neural networks by improving the assortativity of graphs with local mixing patterns. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021, 1541–1551
- [122] Ai G, Gao Y, Wang H, Li X, Wang J, Yan H. Neighbors selective graph convolutional network for homophily and heterophily. Pattern

- Recognition Letters, 2024, 184: 44–51
- [123] Liu S, He D, Yu Z, Jin D, Feng Z. Beyond homophily: neighborhood distribution-guided graph convolutional networks. *Expert Systems with Applications*, 2025, 259: 125274
- [124] Wang Y, Xiang S, Pan C. Improving the homophily of heterophilic graphs for semi-supervised node classification. In: *Proceedings of 2023 IEEE International Conference on Multimedia and Expo (ICME)*. 2023, 1865–1870
- [125] Liu M, Wang Z, Ji S. Non-local graph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(12): 10270–10276
- [126] Yang T, Wang Y, Yue Z, Yang Y, Tong Y, Bai J. Graph pointer neural networks. In: *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. 2022, 8832–8839
- [127] Dong Y, Dupty M H, Deng L, Liu Z, Goh Y L, Lee W S. Differentiable cluster graph neural network. 2024, arXiv preprint arXiv: 2405.16185
- [128] Park J, Yun S, Park H, Kang J, Jeong J, Kim K M, Ha J W, Kim H J. Deformable graph transformer. 2022, arXiv preprint arXiv: 2206.14337
- [129] Zhu J, Rossi R A, Rao A, Mai T, Lipka N, Ahmed N K, Koutra D. Graph neural networks with heterophily. In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. 2021, 11168–11176
- [130] Zhong Z, Ivanov S, Pang J. Simplifying node classification on heterophilous graphs with compatible label propagation. 2022, arXiv preprint arXiv: 2205.09389
- [131] Zheng Z, Bei Y, Zhou S, Ma Y, Gu M, Xu H, Lai C, Chen J, Bu J. Revisiting the message passing in heterophilous graph neural networks. 2024, arXiv preprint arXiv: 2405.17768
- [132] Li X, Zhu R, Cheng Y, Shan C, Luo S, Li D, Qian W. Finding global homophily in graph neural networks when meeting heterophily. In: *Proceedings of the 39th International Conference on Machine Learning*. 2022, 13242–13256
- [133] He D, Liang C, Liu H, Wen M, Jiao P, Feng Z. Block modeling-guided graph convolutional neural networks. In: *Proceedings of the 36th AAAI Conference on Artificial Intelligence*. 2022, 4022–4029
- [134] Yu Z, Feng B, He D, Wang Z, Huang Y, Feng Z. LG-GNN: local-global adaptive graph neural network for modeling both homophily and heterophily. In: *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*. 2024, 2515–2523
- [135] Liu H, Liao N, Luo S. SIMGA: a simple and effective heterophilous graph neural network with efficient global aggregation. 2023, arXiv preprint arXiv: 2305.09958
- [136] Liu X, Zhang L, Guan H. Uplifting message passing neural network with graph original information. 2022, arXiv preprint arXiv: 2210.05382
- [137] Bo D, Wang X, Shi C, Shen H. Beyond low-frequency information in graph convolutional networks. In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. 2021, 3950–3957
- [138] Wu Y, Hu L, Wang Y. Signed attention based graph neural network for graphs with heterophily. *Neurocomputing*, 2023, 557: 126731
- [139] Lai Y, Zhang T, Fan R. Self-attention dual embedding for graphs with heterophily. 2023, arXiv preprint arXiv: 2305.18385
- [140] Yan Y, Hashemi M, Swersky K, Yang Y, Koutra D. Two sides of the same coin: heterophily and oversmoothing in graph convolutional neural networks. In: *Proceedings of 2022 IEEE International Conference on Data Mining (ICDM)*. 2022, 1287–1292
- [141] Choi Y, Choi J, Ko T, Kim C K. Is signed message essential for graph neural networks. 2023, arXiv preprint arXiv: 2301.08918
- [142] Liang L, Kim S, Shin K, Xu Z, Pan S, Qi Y. Sign is not a remedy: multiset-to-multiset message passing for learning on heterophilic graphs. In: *Proceedings of the 41st International Conference on Machine Learning*. 2024
- [143] Sun H, Li X, Wu Z, Su D, Li R H, Wang G. Breaking the entanglement of homophily and heterophily in semi-supervised node classification. In: *Proceedings of IEEE 40th International Conference on Data Engineering (ICDE)*. 2024, 2379–2392
- [144] Chaudhary C, Boran N K, Sangeeth N, Singh V. GNNDLD: graph neural network with directional label distribution. In: *Proceedings of the 16th International Conference on Agents and Artificial Intelligence*. 2024, 165–176
- [145] Rossi E, Charpentier B, Di Giovanni F, Frasca F, Günnemann S, Bronstein M M. Edge directionality improves learning on heterophilic graphs. In: *Proceedings of the 2nd Learning on Graphs Conference*. 2023, 25
- [146] Koke C, Cremers D. HoloNets: spectral convolutions do extend to directed graphs. In: *Proceedings of the 12th International Conference on Learning Representations*. 2024
- [147] Zhuo W, Tan G. Commute graph neural networks. 2024, arXiv preprint arXiv: 2407.01635
- [148] Du L, Shi X, Fu Q, Ma X, Liu H, Han S, Zhang D. GBK-GNN: gated bi-kernel graph neural networks for modeling both homophily and heterophily. In: *Proceedings of the ACM Web Conference 2022*. 2022, 1550–1558
- [149] Wang K, Zhang G, Zhang X, Fang J, Wu X, Li G, Pan S, Huang W, Liang Y. The heterophilic snowflake hypothesis: training and empowering GNNs for heterophilic graphs. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024, 3164–3175
- [150] Cheng Y, Chen M, Li X, Shan C, Gao M. Prioritized propagation in graph neural networks. 2023, arXiv preprint arXiv: 2311.02832
- [151] Deng G, Zhou H, Kannan R, Prasanna V. Learning personalized scoping for graph neural networks under heterophily. 2024, arXiv preprint arXiv: 2409.06998
- [152] Wang J, Guo Y, Yang L, Wang Y. Heterophily-aware graph attention network. *Pattern Recognition*, 2024, 156: 110738
- [153] Rusch T K, Chamberlain B P, Mahoney M W, Bronstein M M, Mishra S. Gradient gating for deep multi-rate learning on graphs. In: *Proceedings of the 11th International Conference on Learning Representations*. 2023
- [154] Finkelshtein B, Huang X, Bronstein M M, Ceylan I I. Cooperative graph neural networks. In: *Proceedings of the 41st International Conference on Machine Learning*. 2024
- [155] Ma J, He M, Wei Z. PolyFormer: scalable node-wise filters via polynomial graph transformer. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024, 2118–2129

- [156] Deng C, Yue Z, Zhang Z. Polynormer: polynomial-expressive graph transformer in linear time. In: Proceedings of the 12th International Conference on Learning Representations. 2024
- [157] Chen J, Li G, Hopcroft J E, He K. SignGT: signed attention-based graph transformer for graph representation learning. 2023, arXiv preprint arXiv: 2310.11025
- [158] Chen S, Chen J, Zhou S, Wang B, Han S, Su C, Yuan Y, Wang C. SIGformer: sign-aware graph transformer for recommendation. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2024, 1274–1284
- [159] Kuang W, Wang Z, Wei Z, Li Y, Ding B. When transformer meets large graphs: an expressive and efficient two-view architecture. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(10): 5440–5452
- [160] Liu C, Zhan Y, Ma X, Ding L, Tao D, Wu J, Hu W. Gapformer: graph transformer with graph pooling for node classification. In: Proceedings of the 32nd International Joint Conference on Artificial Intelligence. 2023, 2196–2205
- [161] Li W, Chen K, Liu S, Zheng T, Huang W, Song M. Learning a mini-batch graph transformer via two-stage interaction augmentation. In: Proceedings of the 27th European Conference on Artificial Intelligence. 2024, 3015–3022
- [162] Fu D, Hua Z, Xie Y, Fang J, Zhang S, Sancak K, Wu H, Malevich A, He J, Long B. VCR-Graphormer: a mini-batch graph transformer via virtual connections. In: Proceedings of the 12th International Conference on Learning Representations. 2024
- [163] Xing Y, Wang X, Li Y, Huang H, Shi C. Less is more: on the over-globalizing problem in graph transformers. In: Proceedings of the 41st International Conference on Machine Learning. 2024
- [164] Zhang Z, Liu Q, Hu Q, Lee C K. Hierarchical graph transformer with adaptive node sampling. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 1539
- [165] Chen J, Jiang S, He K. NTFormer: a composite node tokenized graph transformer for node classification. 2024, arXiv preprint arXiv: 2406.19249
- [166] Li D, Qi B, Gao J, Xiong H, Gu B, Chen X. MPformer: advancing graph modeling through heterophily relationship-based position encoding. In: Proceedings of the 12th International Conference on Learning Representations. 2024
- [167] Ma X, Chen Q, Wu Y, Song G, Wang L, Zheng B. Rethinking structural encodings: adaptive graph transformer for node classification task. In: Proceedings of the ACM Web Conference 2023. 2023, 533–544
- [168] Müller L, Galkin M, Morris C, Rampásek L. Attending to graph transformers. 2023, arXiv preprint arXiv: 2302.04181
- [169] Bo D, Shi C, Wang L, Liao R. Specformer: spectral graph neural networks meet transformers. In: Proceedings of the 11th International Conference on Learning Representations. 2023
- [170] Wang X, Zhu Y, Shi H, Liu Y, Hong C. Graph triple attention network: a decoupled perspective. 2024, arXiv preprint arXiv: 2408.07654
- [171] Chen Q, Wang Y, Wang Y, Yang J, Lin Z. Optimization-induced graph implicit nonlinear diffusion. In: Proceedings of the 39th International Conference on Machine Learning. 2022, 3648–3661
- [172] Eliasof M, Haber E, Treister E. PDE-GCN: novel architectures for graph neural networks motivated by partial differential equations. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 293
- [173] Rusch T K, Chamberlain B, Rowbottom J, Mishra S, Bronstein M M. Graph-coupled oscillator networks. In: Proceedings of the 39th International Conference on Machine Learning. 2022, 18888–18909
- [174] Zhao K, Kang Q, Song Y, She R, Wang S, Tay W P. Graph neural convection-diffusion with heterophily. In: Proceedings of the 32nd International Joint Conference on Artificial Intelligence. 2023, 4656–4664
- [175] Wang Y, Yi K, Liu X, Wang Y G, Jin S. ACMP: Allen-Cahn message passing with attractive and repulsive forces for graph neural networks. In: Proceedings of the 11th International Conference on Learning Representations. 2023
- [176] Choi J, Hong S, Park N, Cho S B. GREAD: graph neural reaction-diffusion networks. In: Proceedings of the 40th International Conference on Machine Learning. 2023, 5722–5747
- [177] Eliasof M, Haber E, Treister E. ADR-GNN: advection-diffusion-reaction graph neural networks. 2023, arXiv preprint arXiv: 2307.16092
- [178] Maskey S, Paolino R, Bacho A, Kutyniok G. A fractional graph Laplacian approach to oversmoothing. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 571
- [179] Zhang A, Li P. Unleashing the power of high-pass filtering in continuous graph neural networks. In: Proceedings of the 15th Asian Conference on Machine Learning. 2023, 1683–1698
- [180] Shao Z, Shi D, Han A, Guo Y, Zhao Q, Gao J. Unifying oversmoothing and over-squashing in graph neural networks: a physics informed approach and beyond. 2023, arXiv preprint arXiv: 2309.02769
- [181] Gravina A, Bacciu D, Gallicchio C. Anti-symmetric DGN: a stable architecture for deep graph networks. In: Proceedings of the 11th International Conference on Learning Representations. 2023
- [182] Bodnar C, Di Giovanni F, Chamberlain B P, Liò P, Bronstein M. Neural sheaf diffusion: a topological perspective on heterophily and oversmoothing in GNNs. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 1346
- [183] Barbero F, Bodnar C, de Ocariz Borde H S, Bronstein M, Veličković P, Liò P. Sheaf neural networks with connection Laplacians. In: Proceedings of Topological, Algebraic, and Geometric Learning Workshops 2022. 2022, 28–36
- [184] Markovich T. QDC: quantum diffusion convolution kernels on graphs. 2023, arXiv preprint arXiv: 2307.11234
- [185] Di Giovanni F, Rowbottom J, Chamberlain B P, Markovich T, Bronstein M M. Understanding convolution on graphs via energies. 2022, arXiv preprint arXiv: 2206.10991
- [186] Zhang A, Li P, Chen G. Steering graph neural networks with pinning control. 2023, arXiv preprint arXiv: 2303.01265
- [187] Wan L, Han H, Sun L, Zhang Z, Ning Z, Yan X, Xia F. Flexible graph neural diffusion with latent class representation learning. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 2936–2947
- [188] Li Y, Wang X, Liu H, Shi C. A generalized neural diffusion framework on graphs. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. 2024, 8707–8715

- [189] Chen R T Q, Rubanova Y, Bettencourt J, Duvenaud D. Neural ordinary differential equations. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018, 6572–6583
- [190] Ortega A, Frossard P, Kovačević J, Moura J M F, Vandergheynst P. Graph signal processing: overview, challenges, and applications. Proceedings of the IEEE, 2018, 106(5): 808–828
- [191] Nt H, Maehara T. Revisiting graph neural networks: all we have is low-pass filters. 2019, arXiv preprint arXiv: 1905.09550
- [192] Gasteiger J, Bojchevski A, Günnemann S. Predict then propagate: graph neural networks meet personalized PageRank. 2018, arXiv preprint arXiv: 1810.05997
- [193] Li Q, Wu X M, Liu H, Zhang X, Guan Z. Label efficient semi-supervised learning via graph filtering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, 9574–9583
- [194] Gasteiger J, Weißenberger S, Günnemann S. Diffusion improves graph learning. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019, 1197
- [195] Li P, Chien I, Milenkovic O. Optimizing generalized PageRank methods for seed-expansion community detection. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019, 1050
- [196] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. 2016, 3844–3852
- [197] Gil A, Segura J, Temme N M. Numerical Methods for Special Functions. Philadelphia: SIAM, 2007
- [198] Weber A. Analysis of the physical Laplacian and the heat flow on a locally finite graph. Journal of Mathematical Analysis and Applications, 2010, 370(1): 146–158
- [199] Kroecker J P. Wiener analysis of nonlinear systems using Poisson-Charlier crosscorrelation. Biological Cybernetics, 1977, 27(4): 221–227
- [200] Hildebrand F B. Introduction to Numerical Analysis. 2nd ed. St. Mineola: Dover Publications, Inc., 1987
- [201] Liesen J, Strakos Z. Krylov Subspace Methods: Principles and Analysis. Oxford: Oxford Academic, 2013
- [202] Cai W, Jiang J, Wang F, Tang J, Kim S, Huang J. A survey on mixture of experts. 2024, arXiv preprint arXiv: 2407.06204
- [203] Xu H, Yan Y, Wang D, Xu Z, Zeng Z, Abdelzaher T F, Han J, Tong H. SLOG: an inductive spectral graph neural network beyond polynomial filter. In: Proceedings of the 41st International Conference on Machine Learning. 2024
- [204] Tenenbaum J B, De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction. Science, 2000, 290(5500): 2319–2323
- [205] Nickel M, Kiela D. Poincaré embeddings for learning hierarchical representations. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017, 6341–6350
- [206] Ribeiro L F R, Saverese P H P, Figueiredo D R. *struc2vec*: learning node representations from structural identity. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017, 385–394
- [207] Grover A, Leskovec J. node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016, 855–864
- [208] Vinyals O, Fortunato M, Jaitly N. Pointer networks. In: Proceedings of the 29th International Conference on Neural Information Processing Systems. 2015, 2692–2700
- [209] Liang L, Hu X, Xu Z, Song Z, King I. Predicting global label relationship matrix for graph neural networks under heterophily. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 480
- [210] Jeh G, Widom J. SimRank: a measure of structural-context similarity. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2002, 538–543
- [211] Cao Z, Qin T, Liu T Y, Tsai M F, Li H. Learning to rank: from pairwise approach to listwise approach. In: Proceedings of the 24th International Conference on Machine Learning. 2007, 129–136
- [212] Wang K, Li G, Wang S, Zhang G, Wang K, You Y, Fang J, Peng X, Liang Y, Wang Y. The snowflake hypothesis: training deep GNN with one node one receptive field. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 3152–3163
- [213] Katz L. A new status index derived from sociometric analysis. Psychometrika, 1953, 18(1): 39–43
- [214] Zhang P, Yan Y, Li C, Wang S, Xie X, Kim S. Can transformer and GNN help each other? 2023, arXiv preprint arXiv: 2308.14355
- [215] Xu J, Wu Z, Lin M, Zhang X, Wang S. LLM and GNN are complementary: distilling LLM for multimodal graph learning. 2024, arXiv preprint arXiv: 2406.01032
- [216] Fan W, Wang S, Huang J, Chen Z, Song Y, Tang W, Mao H, Liu H, Liu X, Yin D, Li Q. Graph machine learning in the era of large language models (LLMs). 2024, arXiv preprint arXiv: 2404.14928
- [217] Chamberlain B, Rowbottom J, Gorinova M I, Bronstein M M, Webb S, Rossi E. Grand: graph neural diffusion. In: Proceedings of the 38th International Conference on Machine Learning. 2021, 1407–1418
- [218] Han A, Shi D, Lin L, Gao J. From continuous dynamics to graph neural networks: neural diffusion and beyond. 2023, arXiv preprint arXiv: 2310.10121
- [219] Chamberlain B, Rowbottom J, Eynard D, Di Giovanni F, Dong X, Bronstein M M. Beltrami flow and neural diffusion on graphs. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 123
- [220] Thorpe M, Nguyen T M, Xia H, Strohmmer T, Bertozzi A L, Osher S J, Wang B. GRAND++: graph neural diffusion with a source term. In: Proceedings of the 10th International Conference on Learning Representations. 2022
- [221] Allen S M, Cahn J W. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. Acta Metallurgica, 1979, 27(6): 1085–1095
- [222] Fisher R A. The wave of advance of advantageous genes. Annals of Eugenics, 1937, 7(4): 355–369
- [223] Gilding B H, Kersner R. Travelling Waves in Nonlinear Diffusion-Convection Reaction. Basel: Springer, 2004
- [224] Hansen J, Ghrist R. Toward a spectral theory of cellular sheaves. Journal of Applied and Computational Topology, 2019, 3(4): 315–358

- [225] Yu W, Chen G, Lü J, Kurths J. Synchronization via pinning control on general complex networks. *SIAM Journal on Control and Optimization*, 2013, 51(2): 1395–1416
- [226] Park M, Heo J, Kim D. Mitigating oversmoothing through reverse process of GNNs for heterophilic graphs. In: *Proceedings of the 41st International Conference on Machine Learning*. 2024, 1606
- [227] Jaiswal A, Babu A R, Zadeh M Z, Banerjee D, Makedon F. A survey on contrastive self-supervised learning. *Technologies*, 2020, 9(1): 2
- [228] Jing L, Tian Y. Self-supervised visual feature learning with deep neural networks: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(11): 4037–4058
- [229] Gui J, Chen T, Zhang J, Cao Q, Sun Z, Luo H, Tao D. A survey on self-supervised learning: algorithms, applications, and future trends. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(12): 9052–9071
- [230] Zhang Q, Zhao Z, Zhou H, Li X, Li C. Self-supervised contrastive learning on heterogeneous graphs with mutual constraints of structure and feature. *Information Sciences*, 2023, 640: 119026
- [231] Ju W, Wang Y, Qin Y, Mao Z, Xiao Z, Luo J, Yang J, Gu Y, Wang D, Long Q, Yi S, Luo X, Zhang M. Towards graph contrastive learning: a survey and beyond. 2024, arXiv preprint arXiv: 2405.11868
- [232] Guo X, Wang Y, Wei Z, Wang Y. Architecture matters: uncovering implicit mechanisms in graph contrastive learning. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. 2023, 1242
- [233] Yang W, Mirzasoleiman B. Graph contrastive learning under heterophily via graph filters. In: *Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence*. 2024, 184
- [234] Liu Y, Zheng Y, Zhang D, Lee V C S, Pan S. Beyond smoothing: unsupervised graph representation learning with edge heterophily discriminating. In: *Proceedings of the 37th AAAI Conference on Artificial Intelligence*. 2023, 4516–4524
- [235] Chen J, Lei R, Wei Z. PolyGCL: GRAPH CONTRASTIVE LEARNING via learnable spectral polynomial filters. In: *Proceedings of the 12th International Conference on Learning Representations*. 2024
- [236] Veličković P, Fedus W, Hamilton W L, Liò P, Bengio Y, Hjelm R D. Deep graph infomax. In: *Proceedings of the 7th International Conference on Learning Representations*. 2019
- [237] You Y, Chen T, Sui Y, Chen T, Wang Z, Shen Y. Graph contrastive learning with augmentations. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 2020, 488
- [238] Zhu Y, Xu Y, Yu F, Liu Q, Wu S, Wang L. Deep graph contrastive representation learning. 2020, arXiv preprint arXiv: 2006.04131
- [239] Liu C, Yu C, Gui N, Yu Z, Deng S. SimGCL: graph contrastive learning by finding homophily in heterophily. *Knowledge and Information Systems*, 2024, 66(3): 2089–2114
- [240] Chen J, Zhu G, Qi Y, Yuan C, Huang Y. Towards self-supervised learning on graphs with heterophily. In: *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 2022, 201–211
- [241] Wang C, Liu Y, Yang Y, Li W. HeterGCL: graph contrastive learning framework on heterophilic graph. In: *Proceedings of the 33rd International Joint Conference on Artificial Intelligence*. 2024, 2397–2405
- [242] Yang L, Hu W, Xu J, Shi R, He D, Wang C, Cao X, Wang Z, Niu B, Guo Y. GAUSS: GrAph-customized universal self-supervised learning. In: *Proceedings of the ACM Web Conference 2024*. 2024, 582–593
- [243] Wang H, Zhang J, Zhu Q, Huang W, Kawaguchi K, Xiao X. Single-pass contrastive learning can work for both homophilic and heterophilic graph. 2022, arXiv preprint arXiv: 2211.10890
- [244] Wang H, Zhang J, Zhu Q, Huang W. Augmentation-free graph contrastive learning with performance guarantee. 2022, arXiv preprint arXiv: 2204.04874
- [245] Khan A, Storkey A. Contrastive learning for non-local graphs with multi-resolution structural views. 2023, arXiv preprint arXiv: 2308.10077
- [246] Coifman R R, Maggioni M. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 2006, 21(1): 53–94
- [247] Yuan M, Chen M, Li X. MUSE: multi-view contrastive learning for heterophilic graphs. In: *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 2023, 3094–3103
- [248] Altenburger K M, Ugander J. Monophily in social networks introduces similarity among friends-of-friends. *Nature Human Behaviour*, 2018, 2(4): 284–290
- [249] Xiao T, Zhu H, Zhang Z, Guo Z, Aggarwal C C, Wang S, Honavar V G. Efficient contrastive learning for fast and accurate inference on graphs. In: *Proceedings of Forty-First International Conference on Machine Learning*. 2024
- [250] Wan G, Tian Y, Huang W, Chawla N V, Ye M. S3GCL: spectral, swift, spatial graph contrastive learning. In: *Proceedings of the 41st International Conference on Machine Learning*. 2024, 2044
- [251] He D, Zhao J, Guo R, Feng Z, Jin D, Huang Y, Wang Z, Zhang W. Contrastive learning meets homophily: two birds with one stone. In: *Proceedings of the 40th International Conference on Machine Learning*. 2023, 12775–12789
- [252] Li W Z, Wang C D, Xiong H, Lai J H. HomoGCL: rethinking homophily in graph contrastive learning. In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023, 1341–1352
- [253] Zhuo J, Qin F, Cui C, Fu K, Niu B, Wang M, Guo Y, Wang C, Wang Z, Cao X, Yang L. Improving graph contrastive learning via adaptive positive sampling. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024, 23179–23187
- [254] Zhuo J, Cui C, Fu K, Niu B, He D, Wang C, Guo Y, Wang Z, Cao X, Yang L. Graph contrastive learning reimagined: exploring universality. In: *Proceedings of the ACM Web Conference 2024*. 2024, 641–651
- [255] Zhao T, Zhang X, Wang S. Disambiguated node classification with graph neural networks. In: *Proceedings of the ACM Web Conference 2024*. 2024, 914–923
- [256] Kipf T N, Welling M. Variational graph auto-encoders. 2016, arXiv preprint arXiv: 1611.07308
- [257] Li J, Wu R, Sun W, Chen L, Tian S, Zhu L, Meng C, Zheng Z,

- Wang W. What's behind the mask: understanding masked graph modeling for graph autoencoders. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023, 1268–1279
- [258] Zhong Z, Gonzalez G, Grattarola D, Pang J. Unsupervised network embedding beyond homophily. 2022, arXiv preprint arXiv: 2203.10866
- [259] Lin B, Li Y, Gui N, Xu Z, Yu Z. Multi-view graph representation learning beyond homophily. *ACM Transactions on Knowledge Discovery from Data*, 2023, 17(8): 114
- [260] Li M, Zhang Y, Wang S, Hu Y, Yin B. Redundancy is not what you need: an embedding fusion graph auto-encoder for self-supervised graph representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(2): 3519–3533
- [261] Li Y, Lin B, Luo B, Gui N. Graph representation learning beyond node and homophily. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(5): 4880–4893
- [262] Tang M, Li P, Yang C. Graph auto-encoder via neighborhood Wasserstein reconstruction. In: Proceedings of the 10th International Conference on Learning Representations. 2022
- [263] Tian Y, Zhang C, Kou Z, Liu Z, Zhang X, Chawla N V. UGMAE: a unified framework for graph masked autoencoders. 2024, arXiv preprint arXiv: 2402.08023
- [264] Luo Y, Li S, Sui Y, Wu J, Wu J, Wang X. Masked graph modeling with multi-view contrast. In: Proceedings of the 40th IEEE International Conference on Data Engineering (ICDE). 2024, 2584–2597
- [265] Fang D, Zhu F, Xie D, Min W. Masked graph autoencoders with contrastive augmentation for spatially resolved transcriptomics data. In: Proceedings of 2024 IEEE International Conference on Bioinformatics and Biomedicine. 2024, 515–520
- [266] Yang W, Zhou L. CMGAE: enhancing graph masked autoencoders through the use of contrastive learning. In: Proceedings of the 2nd International Conference on Machine Learning, Control, and Robotics (MLCR). 2023, 42–47
- [267] Xiao T, Chen Z, Guo Z, Zhuang Z, Wang S. Decoupled self-supervised learning for graphs. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 45
- [268] Wei J, Bosma M, Zhao V Y, Guu K, Yu A W, Lester B, Du N, Dai A M, Le Q V. Finetuned language models are zero-shot learners. In: Proceedings of the 10th International Conference on Learning Representations. 2022
- [269] Jia C, Yang Y, Xia Y, Chen Y T, Parekh Z, Pham H, Le Q V, Sung Y H, Li Z, Duerig T. Scaling up visual and vision-language representation learning with noisy text supervision. In: Proceedings of the 38th International Conference on Machine Learning. 2021, 4904–4916
- [270] Jia M, Tang L, Chen B C, Cardie C, Belongie S, Hariharan B, Lim S N. Visual prompt tuning. In: Proceedings of the 17th European Conference on Computer Vision. 2022, 709–727
- [271] Sun X, Zhang J, Wu X, Cheng H, Xiong Y, Li J. Graph prompt learning: a comprehensive survey and beyond. 2023, arXiv preprint arXiv: 2311.16534
- [272] Long Q, Yan Y, Zhang P, Fang C, Cui W, Ning Z, Xiao M, Cao N, Luo X, Xu L, Jiang S, Fang Z, Chen C, Hua X S, Zhou Y. Towards graph prompt learning: a survey and beyond. 2024, arXiv preprint arXiv: 2408.14520
- [273] Wang L, Zhang M, Jia Z, Li Q, Ma K, Bao C, Zhu J, Zhong Y. AFEC: active forgetting of negative transfer in continual learning. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 1714
- [274] Sun M, Zhou K, He X, Wang Y, Wang X. GPPT: graph pre-training and prompt tuning to generalize graph neural networks. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022, 1717–1727
- [275] Liu Z, Yu X, Fang Y, Zhang X. GraphPrompt: unifying pre-training and downstream tasks for graph neural networks. In: Proceedings of the ACM Web Conference 2023. 2023, 417–428
- [276] Yu X, Liu Z, Fang Y, Liu Z, Chen S, Zhang X. Generalized graph prompt: toward a unification of pre-training and downstream tasks on graphs. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(11): 6237–6250
- [277] Huang Q, Ren H, Chen P, Kržmanc G, Zeng D, Liang P S, Leskovec J. PRODIGY: enabling in-context learning over graphs. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 718
- [278] Zhu Y, Guo J, Tang S. SGL-PT: a strong graph learner with graph prompt tuning. 2023, arXiv preprint arXiv: 2302.12449
- [279] Sun X, Cheng H, Li J, Liu B, Guan J. All in one: multi-task prompting for graph neural networks. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023, 2120–2131
- [280] Zi C, Zhao H, Sun X, Lin Y, Cheng H, Li J. ProG: a graph prompt learning benchmark. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. 2024
- [281] Liu H, Feng J, Kong L, Liang N, Tao D, Chen Y, Zhang M. One for all: towards training one graph model for all classification tasks. In: Proceedings of the 12th International Conference on Learning Representations. 2024
- [282] Fang T, Zhang Y, Yang Y, Wang C, Chen L. Universal prompt tuning for graph neural networks. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023
- [283] Tan Z, Guo R, Ding K, Liu H. Virtual node tuning for few-shot node classification. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023, 2177–2188
- [284] Lee J, Yang W, Kang J. Subgraph-level universal prompt tuning. 2024, arXiv preprint arXiv: 2402.10380
- [285] Yu X, Zhou C, Fang Y, Zhang X. MultiGPrompt for multi-task pre-training and prompting on graphs. In: Proceedings of the ACM Web Conference 2024. 2024, 515–526
- [286] Jiang B, Wu H, Zhang Z, Wang B, Tang J. A unified graph selective prompt learning for graph neural networks. 2024, arXiv preprint arXiv: 2406.10498
- [287] Yan Y, Zhang P, Fang Z, Long Q. Inductive graph alignment prompt: bridging the gap between graph pre-training and inductive fine-tuning from spectral perspective. In: Proceedings of the ACM Web Conference 2024. 2024, 4328–4339
- [288] Chen M, Liu Z, Liu C, Li J, Mao Q, Sun J. ULTRA-DP: unifying

- graph pre-training with multi-task graph dual prompt. 2023, arXiv preprint arXiv: 2310.14845
- [289] Wang J, Deng Z, Lin T, Li W, Ling S. A novel prompt tuning for graph transformers: tailoring prompts to graph topologies. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 3116–3127
- [290] Ma Y, Yan N, Li J, Mortazavi M, Chawla N V. HetGPT: harnessing the power of prompt tuning in pre-trained heterogeneous graph neural networks. In: Proceedings of the ACM Web Conference 2024. 2024, 1015–1023
- [291] Yu X, Fang Y, Liu Z, Zhang X. HGPrompt: bridging homogeneous and heterogeneous graphs for few-shot prompt learning. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. 2024, 1848
- [292] Yu X, Liu Z, Fang Y, Zhang X. DyGPrompt: learning feature and time prompts on dynamic graphs. 2024, arXiv preprint arXiv: 2405.13937
- [293] Song Y, Singh R, Palanisamy B. Krait: a backdoor attack against graph prompt tuning. 2024, arXiv preprint arXiv: 2407.13068
- [294] Lyu X, Han Y, Wang W, Qian H, Tsang I, Zhang X. Cross-context backdoor attacks against graph prompt learning. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 2094–2105
- [295] Wang Y, Xiong Y, Wu X, Sun X, Zhang J, Zheng G. DDIPrompt: drug-drug interaction event prediction based on graph prompt learning. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024, 2431–2441
- [296] Ye R, Zhang C, Wang R, Xu S, Zhang Y. Language is all a graph needs. In: Proceedings of the Association for Computational Linguistics. 2024, 1955–1973
- [297] Duan Y, Liu J, Chen S, Chen L, Wu J. G-Prompt: graphon-based prompt tuning for graph classification. *Information Processing & Management*, 2024, 61(3): 103639
- [298] Fang Y, Fan D, Zha D, Tan Q. GAUGLLM: improving graph contrastive learning for text-attributed graphs with large language models. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 747–758
- [299] Jiang W, Wu W, Zhang L, Yuan Z, Xiang J, Zhou J, Xiong H. Killing two birds with one stone: cross-modal reinforced prompting for graph and language tasks. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 1301–1312
- [300] Jin J, Song Y, Kan D, Zhu H, Sun X, Li Z, Sun X, Zhang J. Urban region pre-training and prompting: a graph-based approach. 2024, arXiv preprint arXiv: 2408.05920
- [301] Zhang P, Yan Y, Zhang X, Kang L, Li C, Huang F, Wang S, Kim S. GPT4Rec: graph prompt tuning for streaming recommendation. In: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2024, 1774–1784
- [302] Gong C, Li X, Yu J, Cheng Y, Tan J, Yu C. Self-pro: a self-prompt and tuning framework for graph neural networks. In: Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases. 2024, 197–215
- [303] Yu X, Zhang J, Fang Y, Jiang R. Non-homophilic graph pre-training and prompt learning. 2024, arXiv preprint arXiv: 2408.12594
- [304] Zhou K, Yang J, Loy C C, Liu Z. Conditional prompt learning for vision-language models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, 16795–16804
- [305] Ge Q, Zhao Z, Liu Y, Cheng A, Li X, Wang S, Yin D. Enhancing graph neural networks with structure-based prompt. 2023, arXiv preprint arXiv: 2310.17394
- [306] Wang S, Yang J, Yao J, Bai Y, Zhu W. An overview of advanced deep graph node clustering. *IEEE Transactions on Computational Social Systems*, 2024, 11(1): 1302–1314
- [307] Tian F, Gao B, Cui Q, Chen E, Liu T Y. Learning deep representations for graph clustering. In: Proceedings of the 28th AAAI Conference on Artificial Intelligence. 2014
- [308] Pan E, Kang Z. Multi-view contrastive graph clustering. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 165
- [309] Zhao H, Yang X, Wang Z, Yang E, Deng C. Graph debiased contrastive learning with joint representation clustering. In: Proceedings of the 30th International Joint Conference on Artificial Intelligence. 2021, 3434–3440
- [310] Xie X, Chen W, Kang Z, Peng C. Contrastive graph clustering with adaptive filter. *Expert Systems with Applications*, 2023, 219: 119645
- [311] Wen Z, Ling Y, Ren Y, Wu T, Chen J, Pu X, Hao Z, He L. Homophily-Related: adaptive hybrid graph filter for multi-view graph clustering. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. 2024, 15841–15849
- [312] Pan E, Kang Z. Beyond homophily: reconstructing structure for graph-agnostic clustering. In: Proceedings of the 40th International Conference on Machine Learning. 2023, 26868–26877
- [313] Xie X, Pan E, Kang Z, Chen W, Li B. Provable filter for real-world graph clustering. 2024, arXiv preprint arXiv: 2403.03666
- [314] Zhu P, Li J, Wang Y, Xiao B, Zhang J, Lin W, Hu Q. Boosting pseudo-labeling with curriculum self-reflection for attributed graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2024
- [315] Gu M, Yang G, Zhou S, Ma N, Chen J, Tan Q, Liu M, Bu J. Homophily-enhanced structure learning for graph clustering. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023, 577–586
- [316] Lü L, Zhou T. Link prediction in complex networks: a survey. *Physica A: Statistical Mechanics and Its Applications*, 2011, 390(6): 1150–1170
- [317] Zhang M, Chen Y. Link prediction based on graph neural networks. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018, 5171–5181
- [318] Li X, Ye T, Shan C, Li D, Gao M. SeeGera: self-supervised semi-implicit graph variational auto-encoders with masking. In: Proceedings of the ACM Web Conference 2023. 2023, 143–153
- [319] Tan Q, Liu N, Huang X, Choi S H, Li L, Chen R, Hu X. S2GAE: self-supervised graph autoencoders are generalizable learners with graph masking. In: Proceedings of the 16th ACM International Conference on Web Search and Data Mining. 2023, 787–795
- [320] Zhou S, Guo Z, Aggarwal C C, Zhang X, Wang S. Link

- prediction on heterophilic graphs via disentangled representation learning. 2022, arXiv preprint arXiv: 2208.01820
- [321] Di Francesco A G, Caso F, Bucarelli M S, Silvestri F. Link prediction under heterophily: a physics-inspired graph neural network approach. 2024, arXiv preprint arXiv: 2402.14802
- [322] Zhu J, Li G, Yang Y A, Zhu J, Cui X, Koutra D. On the impact of feature heterophily on link prediction with graph neural networks. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. 2024
- [323] Yang J, Medya S, Ye W. Incorporating heterophily into graph neural networks for graph classification. In: Proceedings of 2024 IEEE International Conference on Systems, Man, and Cybernetics. 2024, 1544–1551
- [324] Ding Y, Liu Z, Hao H. Self-supervised learning and graph classification under heterophily. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023, 3849–3853
- [325] Papachristou M, Goel R, Portman F, Miller M, Jin R. GLINKX: a scalable unified framework for homophilous and heterophilous graphs. 2022, arXiv preprint arXiv: 2211.00550
- [326] Liao N, Luo S, Li X, Shi J. LD<sup>2</sup>: scalable heterophilous graph neural network with decoupled embeddings. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 446
- [327] Chen J, Li Z, Zhu Y, Zhang J, Pu J. From node interaction to hop interaction: new effective and scalable graph learning paradigm. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023, 7876–7885
- [328] Das S S, Ferdous S, Halappanavar M M, Serra E, Pothan A. AGS-GNN: attribute-guided sampling for graph neural networks. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 538–549
- [329] Wu Q, Zhao W, Li Z, Wipf D, Yan J. NodeFormer: a scalable graph structure learning transformer for node classification. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 1986
- [330] Wu Q, Yang C, Zhao W, He Y, Wipf D, Yan J. DIFFormer: scalable (graph) transformers induced by energy constrained diffusion. In: Proceedings of the 11th International Conference on Learning Representations. 2023
- [331] Wu Q, Zhao W, Yang C, Zhang H, Nie F, Jiang H, Bian Y, Yan J. SGFormer: simplifying and empowering transformers for large-graph representations. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 2826
- [332] Kong K, Chen J, Kirchenbauer J, Ni R, Bruss C B, Goldstein T. GOAT: a global transformer on largescale graphs. In: Proceedings of the 40th International Conference on Machine Learning. 2023, 17375–17390
- [333] Sun Y, Zhu D, Wang Y, Tian Z, Cao N, O'Hared G. SpikeGraphormer: a high-performance graph transformer with spiking graph attention. 2024, arXiv preprint arXiv: 2403.15480
- [334] Ghosh-Dastidar S, Adeli H. Spiking neural networks. *International Journal of Neural Systems*, 2009, 19(4): 295–308
- [335] Tavanaei A, Ghodrati M, Kheradpisheh S R, Masquelier T, Maida A. Deep learning in spiking neural networks. *Neural Networks*, 2019, 111: 47–63
- [336] Chakraborty A, Alam M, Dey V, Chattopadhyay A, Mukhopadhyay D. Adversarial attacks and defences: a survey. 2018, arXiv preprint arXiv: 1810.00069
- [337] Silva S H, Najafirad P. Opportunities and challenges in deep learning adversarial robustness: a survey. 2020, arXiv preprint arXiv: 2007.00753
- [338] Xu J, Chen J, You S, Xiao Z, Yang Y, Lu J. Robustness of deep learning models on graphs: a survey. *AI Open*, 2021, 2: 69–78
- [339] Zügner D, Akbarnejad A, Günnemann S. Adversarial attacks on neural networks for graph data. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018, 2847–2856
- [340] Dai H, Li H, Tian T, Huang X, Wang L, Zhu J, Song L. Adversarial attack on graph structured data. In: Proceedings of the 35th International Conference on Machine Learning. 2018, 1123–1132
- [341] Zhu J, Jin J, Loveland D, Schaub M T, Koutra D. How does heterophily impact the robustness of graph neural networks? Theoretical connections and practical implications. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2022, 2637–2647
- [342] Huang J, Du L, Chen X, Fu Q, Han S, Zhang D. Robust mid-pass filtering graph convolutional networks. In: Proceedings of the ACM Web Conference 2023. 2023, 328–338
- [343] Zhu Y, Lai Y, Ai X, Zhou K. Universally robust graph neural networks by preserving neighbor similarity. 2024, arXiv preprint arXiv: 2401.09754
- [344] Lei R, Wang Z, Li Y, Ding B, Wei Z. EvenNet: ignoring odd-hop neighbors improves robustness of graph neural networks. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 339
- [345] Qiu C, Nan G, Xiong T, Deng W, Wang D, Teng Z, Sun L, Cui Q, Tao X. Refining latent homophilic structures over heterophilic graphs for robust graph convolution networks. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. 2024, 8930–8938
- [346] Zhang S, Liu Y, Sun Y, Shah N. Graph-less neural networks: teaching old MLPs new tricks via distillation. In: Proceedings of the Tenth International Conference on Learning Representations. 2022
- [347] Deng B, Chen J, Hu Y, Xu Z, Chen C, Zhang T. PROSPECT: learn MLPs on graphs robust against adversarial structure attacks. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024, 425–435
- [348] Cheng Y, Shan C, Shen Y, Li X, Luo S, Li D. Resurrecting label propagation for graphs with heterophily and label noise. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 433–444
- [349] He X, Wen R, Wu Y, Backes M, Shen Y, Zhang Y. Node-level membership inference attacks against graph neural networks. 2021, arXiv preprint arXiv: 2102.05429
- [350] Liao P, Zhao H, Xu K, Jaakkola T S, Gordon G J, Jegelka S, Salakhutdinov R. Information obfuscation of graph neural networks. In: Proceedings of the 38th International Conference on Machine Learning. 2021, 6600–6610
- [351] Hu H, Cheng L, Vap J P, Borowczak M. Learning privacy-

- preserving graph convolutional network with partially observed sensitive attributes. In: Proceedings of the ACM Web Conference 2022. 2022, 3552–3561
- [352] Yuan H, Xu J, Wang C, Yang Z, Wang C, Yin K, Yang Y. Unveiling privacy vulnerabilities: investigating the role of structure in graph data. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 4059–4070
- [353] Wu R, Fang G, Zhang M, Pan Q, Liu T, Wang W, Zhao W. On provable privacy vulnerabilities of graph representations. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. 2024
- [354] Mueller T T, Chevli M, Daigavane A, Rueckert D, Kaissis G. Privacy-utility trade-offs in neural networks for medical population graphs: insights from differential privacy and graph structure. 2023, arXiv preprint arXiv: 2307.06760
- [355] Dai E, Wang S. Say no to the discrimination: learning fair graph neural networks with limited sensitive attribute information. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 2021, 680–688
- [356] Li P, Wang Y, Zhao H, Hong P, Liu H. On dyadic fairness: exploring and mitigating bias in graph connections. In: Proceedings of the 9th International Conference on Learning Representations. 2021
- [357] Loveland D, Zhu J, Heimann M, Fish B, Schaub M T, Koutra D. On graph neural network fairness in the presence of heterophilous neighborhoods. 2022, arXiv preprint arXiv: 2207.04376
- [358] Zhu Y, Xu W, Zhang J, Liu Q, Wu S, Wang L. Deep graph structure learning for robust representations: a survey. 2021, arXiv preprint arXiv: 2103.03036
- [359] Ye Y, Ji S. Sparse graph attention networks. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(1): 905–916
- [360] Xue Y, Jin Z, Gao W. A data-centric graph neural network for node classification of heterophilic networks. *International Journal of Machine Learning and Cybernetics*, 2024, 15(8): 3413–3423
- [361] Yang Y, Sun Y, Wang S, Guo J, Gao J, Ju F, Yin B. Graph neural networks with soft association between topology and attribute. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. 2024, 1030
- [362] Deac A, Tang J. Evolving computation graphs. 2023, arXiv preprint arXiv: 2306.12943
- [363] Zheng Y, Zhang H, Lee V C S, Zheng Y, Wang X, Pan S. Finding the missing-half: graph complementary learning for homophily-prone and heterophily-prone graphs. In: Proceedings of the 40th International Conference on Machine Learning. 2023, 42492–42505
- [364] Xu Z, Chen Y, Zhou Q, Wu Y, Pan M, Yang H, Tong H. Node classification beyond homophily: towards a general solution. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023, 2862–2873
- [365] Bi W, Du L, Fu Q, Wang Y, Han S, Zhang D. Make heterophilic graphs better fit GNN: a graph rewiring approach. *IEEE Transactions on Knowledge and Data Engineering*, 2024, 36(12): 8744–8757
- [366] Choi Y, Choi J, Ko T, Byun H, Kim C K. Finding heterophilic neighbors via confidence-based subgraph matching for semi-supervised node classification. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022, 283–292
- [367] Nie F, Zeng Z, Tsang I W, Xu D, Zhang C. Spectral embedded clustering: a framework for in-sample and out-of-sample spectral clustering. *IEEE Transactions on Neural Networks*, 2011, 22(11): 1796–1808
- [368] Jiang M, Liu G, Su Y, Wu X. GCN-SL: graph convolutional networks with structure learning for graphs under heterophily. 2021, arXiv preprint arXiv: 2105.13795
- [369] Box G E P, Tiao G C. *Bayesian Inference in Statistical Analysis*. Hoboken: John Wiley & Sons, 2011
- [370] Wang R, Mou S, Wang X, Xiao W, Ju Q, Shi C, Xie X. Graph structure estimation neural networks. In: Proceedings of the Web Conference 2021. 2021, 342–353
- [371] Wu L, Tan C, Liu Z, Gao Z, Lin H, Li S Z. Learning to augment graph structure for both homophily and heterophily graphs. In: Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases. 2023, 3–18
- [372] Dong M, Kluger Y. Towards understanding and reducing graph structural noise for GNNs. In: Proceedings of the 40th International Conference on Machine Learning. 2023, 8202–8226
- [373] Liu Y, Zheng Y, Zhang D, Chen H, Peng H, Pan S. Towards unsupervised deep graph structure learning. In: Proceedings of the ACM Web Conference 2022. 2022, 1392–1403
- [374] Wu L, Lin H, Liu Z, Liu Z, Huang Y, Li S Z. Homophily-enhanced self-supervision for graph structure learning: insights and directions. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(9): 12358–12372
- [375] Vincent P, Larochelle H, Bengio Y, Manzagol P A. Extracting and composing robust features with denoising autoencoders. In: Proceedings of the 25th International Conference on Machine Learning. 2008, 1096–1103
- [376] Wu L, Lin H, Zhao G, Tan C, Li S Z. Learning to model graph structural information on MLPs via graph structure self-contrasting. *IEEE Transactions on Neural Networks and Learning Systems*, 2024
- [377] Li Z, Sun X, Luo Y, Zhu Y, Chen D, Luo Y, Zhou X, Liu Q, Wu S, Wang L, Yu J X. GSLB: the graph structure learning benchmark. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 1319
- [378] Gong Z, Wang G, Sun Y, Liu Q, Ning Y, Xiong H, Peng J. Beyond homophily: robust graph anomaly detection via neural sparsification. In: Proceedings of the 32nd International Joint Conference on Artificial Intelligence. 2023, 2104–2113
- [379] Wen J, Jiang N, Li L, Zhou J, Li Y, Zhan H, Kou G, Gu W, Zhao J. TA-Detector: a GNN-based anomaly detector via trust relationship. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2024
- [380] Gao Y, Wang X, He X, Liu Z, Feng H, Zhang Y. Addressing heterophily in graph anomaly detection: a perspective of graph spectrum. In: Proceedings of the ACM Web Conference 2023. 2023, 1528–1538
- [381] Qiao H, Pang G. Truncated affinity maximization: one-class homophily modeling for graph anomaly detection. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 2154
- [382] Zhang R, Cheng D, Liu X, Yang J, Ouyang Y, Wu X, Zheng Y. Generation is better than modification: combating high class homophily

- variance in graph anomaly detection. 2024, arXiv preprint arXiv: 2403.10339
- [383] Tang J, Li J, Gao Z, Li J. Rethinking graph neural networks for anomaly detection. In: Proceedings of the 39th International Conference on Machine Learning. 2022, 21076–21089
- [384] Chai Z, You S, Yang Y, Pu S, Xu J, Cai H, Jiang W. Can abnormality be detected by graph neural networks? In: Proceedings of the 31st International Joint Conference on Artificial Intelligence. 2022, 1945–1951
- [385] Jin W, Ma H, Zhang Y, Li Z, Chang L. Multi-view discriminative edge heterophily contrastive learning network for attributed graph anomaly detection. *Expert Systems with Applications*, 2024, 255: 124460
- [386] Roy A, Shu J, Li J, Yang C, Elshocht O, Smeets J, Li P. GAD-NR: graph anomaly detection via neighborhood reconstruction. In: Proceedings of the 17th ACM International Conference on Web Search and Data Mining. 2024, 576–585
- [387] Gao Y, Wang X, He X, Liu Z, Feng H, Zhang Y. Alleviating structural distribution shift in graph anomaly detection. In: Proceedings of the 16th ACM International Conference on Web Search and Data Mining. 2023, 357–365
- [388] Pourhabibi T, Ong K L, Kam B H, Boo Y L. Fraud detection: a systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, 2020, 133: 113303
- [389] Shi F, Cao Y, Shang Y, Zhou Y, Zhou C, Wu J. H2-FDetector: a GNN-based fraud detector with homophilic and heterophilic connections. In: Proceedings of the ACM Web Conference 2022. 2022, 1486–1494
- [390] Kim H, Choi J, Whang J J. Dynamic relation-attentive graph neural networks for fraud detection. In: Proceedings of 2023 IEEE International Conference on Data Mining Workshops (ICDMW). 2023, 1092–1096
- [391] Wang Y, Zhang J, Huang Z, Li W, Feng S, Ma Z, Sun Y, Yu D, Dong F, Jin J, Wang B, Luo J. Label information enhanced fraud detection against low homophily in graphs. In: Proceedings of the ACM Web Conference 2023. 2023, 406–416
- [392] Duan M, Zheng T, Gao Y, Wang G, Feng Z, Wang X. DGA-GNN: dynamic grouping aggregation GNN for fraud detection. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. 2024, 11820–11828
- [393] Wu B, Yao X, Zhang B, Chao K M, Li Y. SplitGNN: spectral graph neural network for fraud detection against heterophily. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023, 2737–2746
- [394] Xu F, Wang N, Wu H, Wen X, Zhao X, Wan H. Revisiting graph-based fraud detection in sight of heterophily and spectrum. In: Proceedings of the 38th AAAI Conference on Artificial Intelligence. 2024, 9214–9222
- [395] Bergstrom C T, Bak-Coleman J B. Information gerrymandering in social networks skews collective decision-making. *Nature*, 2019, 573(7772): 40–41
- [396] Deb A, Luceri L, Badaway A, Ferrara E. Perils and challenges of social media and election manipulation analysis: the 2018 US midterms. In: Proceedings of Companion Proceedings of the 2019 World Wide Web Conference. 2019, 237–247
- [397] Ferrara E. Disinformation and social bot operations in the run up to the 2017 French presidential election. 2017, arXiv preprint arXiv: 1707.00086
- [398] Ashmore B, Chen L. HOVER: homophilic oversampling via edge removal for class-imbalanced bot detection on graphs. In: Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. 2023, 3728–3732
- [399] Zhou M, Feng W, Zhu Y, Zhang D, Dong Y, Tang J. Semi-supervised social bot detection with initial residual relation attention networks. In: Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases. 2023, 207–224
- [400] Li S, Qiao B, Li K, Lu Q, Lin M, Zhou W. Multi-modal social bot detection: learning homophilic and heterophilic connections adaptively. In: Proceedings of the 31st ACM International Conference on Multimedia. 2023, 3908–3916
- [401] Ye S, Tan Z, Lei Z, He R, Wang H, Zheng Q, Luo M. HOFA: twitter bot detection with homophily-oriented augmentation and frequency adaptive attention. 2023, arXiv preprint arXiv: 2306.12870
- [402] Shi S, Qiao K, Wang Z, Yang J, Song B, Chen J, Yan B. Multi-scale graph neural network with signed-attention for social bot detection: a frequency perspective. 2023, arXiv preprint arXiv: 2307.01968
- [403] Wu Q, Yang Y, He B, Liu H, Wang X, Liao Y, Yang R, Zhou P. Heterophily-aware social bot detection with supervised contrastive learning. 2023, arXiv preprint arXiv: 2306.07478
- [404] Chen X, Zhou F, Trajcevski G, Bonsangue M. Multi-view learning with distinguishable feature fusion for rumor detection. *Knowledge-Based Systems*, 2022, 240: 108085
- [405] Yang X, Lyu Y, Tian T, Liu Y, Liu Y, Zhang X. Rumor detection on social media with graph structured adversarial learning. In: Proceedings of the 29th International Conference on International Joint Conferences on Artificial Intelligence. 2021, 197
- [406] Yan Y, Wang Y, Zheng P. A graph-based pivotal semantic mining framework for rumor detection. *Engineering Applications of Artificial Intelligence*, 2023, 118: 105613
- [407] Nguyen T T, Ren Z, Nguyen T T, Jo J, Nguyen Q V H, Yin H. Portable graph-based rumour detection against multi-modal heterophily. *Knowledge-Based Systems*, 2024, 284: 111310
- [408] Wang C, Lin Z, Yang X, Sun J, Yue M, Shahabi C. HAGEN: homophily-aware graph convolutional recurrent network for crime forecasting. In: Proceedings of the 36th AAAI Conference on Artificial Intelligence. 2022, 4193–4200
- [409] He X, Deng K, Wang X, Li Y, Zhang Y, Wang M. LightGCN: simplifying and powering graph convolution network for recommendation. In: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. 2020, 639–648
- [410] Wang X, He X, Wang M, Feng F, Chua T S. Neural graph collaborative filtering. In: Proceedings of the 42nd international ACM SIGIR conference on Research and development in Information Retrieval. 2019, 165–174
- [411] Sun J, Guo W, Zhang D, Zhang Y, Regol F, Hu Y, Guo H, Tang R, Yuan H, He X, Coates M. A framework for recommending accurate and diverse items using Bayesian graph convolutional neural networks. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020, 2030–2039

- [412] Zhao Z, Zhang X, Zhou H, Li C, Gong M, Wang Y. HetNERec: heterogeneous network embedding based recommendation. *Knowledge-Based Systems*, 2020, 204: 106218
- [413] Wu J, Wang X, Feng F, He X, Chen L, Lian J, Xie X. Self-supervised graph learning for recommendation. In: *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2021, 726–735
- [414] Jiang W, Jiao Y, Wang Q, Liang C, Guo L, Zhang Y, Sun Z, Xiong Y, Zhu Y. Triangle graph interest network for click-through rate prediction. In: *Proceedings of the 15th ACM International Conference on Web Search and Data Mining*. 2022, 401–409
- [415] Jiang W, Gao X, Xu G, Chen T, Yin H. Challenging low homophily in social recommendation. In: *Proceedings of the ACM Web Conference 2024*. 2024, 3476–3484
- [416] Gholinejad N, Chehreghani M H. Heterophily-aware fair recommendation using graph convolutional networks. 2024, arXiv preprint arXiv: 2402.03365
- [417] Zoidi O, Fotiadou E, Nikolaidis N, Pitas I. Graph-based label propagation in digital media: a review. *ACM Computing Surveys (CSUR)*, 2015, 47(3): 48
- [418] Taelman C, Chlailly S, Khachatryan E, van der Sommen F, Marinoni A. On the exploitation of heterophily in graph-based multimodal remote sensing data analysis. In: *Proceedings of the 29th CIKM 2021 Workshops Co-Located with 30th ACM International Conference on Information and Knowledge Management*. 2021
- [419] Eswaran D, Günemann S, Faloutsos C, Makhija D, Kumar M. ZooBP: belief propagation for heterogeneous networks. *Proceedings of the VLDB Endowment*, 2017, 10(5): 625–636
- [420] Han K, Wang Y, Guo J, Tang Y, Wu E. Vision GNN: an image is worth graph of nodes. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. 2022, 603
- [421] Peng W, Hong X, Chen H, Zhao G. Learning graph convolutional network for skeleton-based human action recognition by neural searching. In: *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. 2020, 2669–2676
- [422] Chen Y, Rohrbach M, Yan Z, Shuicheng Y, Feng J, Kalantidis Y. Graph-based global reasoning networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, 433–442
- [423] Yang J, Lu J, Lee S, Batra D, Parikh D. Graph R-CNN for scene graph generation. In: *Proceedings of the 15th European Conference on Computer Vision (ECCV)*. 2018, 690–706
- [424] Li H, Zhu G, Zhang L, Jiang Y, Dang Y, Hou H, Shen P, Zhao X, Shah S A A, Bennamoun M. Scene graph generation: a comprehensive survey. *Neurocomputing*, 2024, 566: 127052
- [425] Chang X, Ren P, Xu P, Li Z, Chen X, Hauptmann A. A comprehensive survey of scene graphs: Generation and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(1): 1–26
- [426] Lin X, Ding C, Zhan Y, Li Z, Tao D. HL-Net: heterophily learning network for scene graph generation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, 19454–19463
- [427] Shuman D I, Narang S K, Frossard P, Ortega A, Vandergheynst P. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 2013, 30(3): 83–98
- [428] Chen L, Song Y, Lin S, Wang C, He G. Kumaraswamy wavelet for heterophilic scene graph generation. In: *Proceedings of the 38th AAAI Conference on Artificial Intelligence*. 2024, 1138–1146
- [429] Kumaraswamy P. A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, 1980, 46(1-2): 79–88
- [430] Nguyen A, Le B. 3D point cloud segmentation: a survey. In: *Proceedings of the 6th IEEE Conference on Robotics, Automation and Mechatronics (RAM)*. 2013, 225–230
- [431] Wang L, Huang Y, Hou Y, Zhang S, Shan J. Graph attention convolution for point cloud semantic segmentation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, 10288–10297
- [432] Lei H, Akhtar N, Mian A. Spherical kernel for efficient graph convolution on 3D point clouds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(10): 3664–3680
- [433] Pan X, Xia Z, Song S, Li L E, Huang G. 3D object detection with pointformer. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, 7459–7468
- [434] Chen S, Wei L, Liang L, Lang C. Joint homophily and heterophily relational knowledge distillation for efficient and compact 3D object detection. In: *Proceedings of the 32nd ACM International Conference on Multimedia*. 2024, 2127–2135
- [435] Wang Y, Wang J, Cao Z, Barati Farmani A. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 2022, 4(3): 279–287
- [436] Merchant A, Batzner S, Schoenholz S S, Aykol M, Cheon G, Cubuk E D. Scaling deep learning for materials discovery. *Nature*, 2023, 624(7990): 80–85
- [437] Fang X, Liu L, Lei J, He D, Zhang S, Zhou J, Wang F, Wu H, Wang H. Geometry-enhanced molecular representation learning for property prediction. *Nature Machine Intelligence*, 2022, 4(2): 127–134
- [438] Li H, Han Z, Sun Y, Wang F, Hu P, Gao Y, Bai X, Peng S, Ren C, Xu X, Liu Z, Chen H, Yang Y, Bo X. CGMega: explainable graph neural network framework with attention mechanisms for cancer gene module dissection. *Nature Communications*, 2024, 15(1): 5997
- [439] Wu L, Wen Y, Leng D, Zhang Q, Dai C, Wang Z, Liu Z, Yan B, Zhang Y, Wang J, He S, Bo X. Machine learning methods, databases and tools for drug combination prediction. *Briefings in Bioinformatics*, 2022, 23(1): bbab355
- [440] Cheng F, Kovács I A, Barabási A L. Network-based prediction of drug combinations. *Nature Communications*, 2019, 10(1): 1197
- [441] Jia Y, Yun C H, Park E, Ercan D, Manuia M, Juarez J, Xu C, Rhee K, Chen T, Zhang H, Palakurthi S, Jang J, Lelais G, Didonato M, Bursulaya B, Michellys P Y, Epple R, Marsilje T H, McNeill M, Lu W, Harris J, Bender S, Wong K K, Jänne P A, Eck M J. Overcoming EGFR (T790M) and EGFR (C797S) resistance with mutant-selective allosteric inhibitors. *Nature*, 2016, 534(7605): 129–132
- [442] Chen H, Lu Y, Yang Y, Rao Y. A drug combination prediction framework based on graph convolutional network and heterogeneous information. *IEEE/ACM Transactions on Computational Biology and*

Bioinformatics, 2023, 20(3): 1917–1925

- [443] Liu B M, Gao Y L, Li F, Zheng C H, Liu J X. SLGCN: structure-enhanced line graph convolutional network for predicting drug–disease associations. *Knowledge-Based Systems*, 2024, 283: 111187
- [444] Xue H, Li J, Xie H, Wang Y. Review of drug repositioning approaches and resources. *International Journal of Biological Sciences*, 2018, 14(10): 1232–1244
- [445] Jarada T N, Rokne J G, Alhadj R. A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions. *Journal of Cheminformatics*, 2020, 12(1): 46
- [446] Lotfi Shahreza M, Ghadiri N, Mousavi S R, Varshosaz J, Green J R. A review of network-based approaches to drug repositioning. *Briefings in Bioinformatics*, 2018, 19(5): 878–892
- [447] Kang S, Cho K. Conditional molecular design with deep generative models. *Journal of Chemical Information and Modeling*, 2019, 59(1): 43–52
- [448] Yang M, Sun H, Liu X, Xue X, Deng Y, Wang X. CMGN: a conditional molecular generation net to design target-specific molecules with desired properties. *Briefings in Bioinformatics*, 2023, 24(4): bbad185
- [449] Rigoni D, Navarin N, Sperduti A. Conditional constrained graph variational autoencoders for molecule design. In: *Proceedings of 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2020, 729–736
- [450] Wang H, Solin A, Garg V. Molecule generation by heterophilous triple flows. In: *Proceedings of the Twelfth International Conference on Learning Representations*. 2024
- [451] Wang Z, Zhu X, Adeli E, Zhu Y, Nie F, Munsell B, Wu G. Multi-modal classification of neurodegenerative disease by progressive graph-based transductive learning. *Medical Image Analysis*, 2017, 39: 218–230
- [452] Zhu Y, Zhu X, Kim M, Yan J, Kaufer D, Wu G. Dynamic hypergraph inference framework for computer-assisted diagnosis of neurodegenerative diseases. *IEEE Transactions on Medical Imaging*, 2019, 38(2): 608–616
- [453] Wang M, Shao W, Huang S, Zhang D. Hypergraph-regularized multimodal learning by graph diffusion for imaging genetics based Alzheimer’s disease diagnosis. *Medical Image Analysis*, 2023, 89: 102883
- [454] Qu Z, Yao T, Liu X, Wang G. A graph convolutional network based on univariate neurodegeneration biomarker for Alzheimer’s disease diagnosis. *IEEE Journal of Translational Engineering in Health and Medicine*, 2023, 11: 405–416
- [455] Xu J, Yang Y, Huang D, Gururajapathy S S, Ke Y, Qiao M, Wang A, Kumar H, McGeown J, Kwon E. Data-driven network neuroscience: on data collection and benchmark. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. 2023, 955
- [456] Hammond D K, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 2011, 30(2): 129–150
- [457] Xu B, Shen H, Cao Q, Qiu Y, Cheng X. Graph wavelet neural network. 2019, arXiv preprint arXiv: 1904.07785
- [458] Nair A, Roy A, Meinke K. funcGNN: a graph neural network approach to program similarity. In: *Proceedings of the 14th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*. 2020, 10
- [459] Xu S, Shen J, Li Y, Yao Y, Yu P, Xu F, Ma X. On the heterophily of program graphs: a case study of graph-based type inference. In: *Proceedings of the 15th Asia-Pacific Symposium on Internetware*. 2024, 1–10
- [460] Yan M, Xia X, Fan Y, Hassan A E, Lo D, Li S. Just-in-time defect identification and localization: a two-phase framework. *IEEE Transactions on Software Engineering*, 2022, 48(1): 82–101
- [461] Qiu F, Yan M, Xia X, Wang X, Fan Y, Hassan A E, Lo D. JITO: a tool for just-in-time defect identification and localization. In: *Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*. 2020, 1586–1590
- [462] Qiu F, Gao Z, Xia X, Lo D, Grundy J, Wang X. Deep just-in-time defect localization. *IEEE Transactions on Software Engineering*, 2021, 48(12): 5068–5086
- [463] Zhang H, Min W, Wei Z, Kuang L, Gao H, Miao H. A just-in-time software defect localization method based on code graph representation. In: *Proceedings of the 32nd IEEE/ACM International Conference on Program Comprehension*. 2024, 293–303
- [464] Shi C, Li Y, Zhang J, Sun Y, Yu P S. A survey of heterogeneous information network analysis. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(1): 17–37
- [465] Wang X, Ji H, Shi C, Wang B, Ye Y, Cui P, Yu P S. Heterogeneous graph attention network. In: *Proceedings of the World Wide Web Conference*. 2019, 2022–2032
- [466] Li X, Wu Y, Ester M, Kao B, Wang X, Zheng Y. Semi-supervised clustering in attributed heterogeneous information networks. In: *Proceedings of the 26th International Conference on World Wide Web*. 2017, 1621–1629
- [467] Li X, Ding D, Kao B, Sun Y, Mamoulis N. Leveraging meta-path contexts for classification in heterogeneous information networks. In: *Proceedings of 2021 IEEE 37th International Conference on Data Engineering (ICDE)*. 2021, 912–923
- [468] Li X, Kao B, Zheng Y, Huang Z. On transductive classification in heterogeneous information networks. In: *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. 2016, 811–820
- [469] Zhao J, Wang X, Shi C, Hu B, Song G, Ye Y. Heterogeneous graph structure learning for graph neural networks. In: *Proceedings of the 35th AAAI Conference on Artificial Intelligence*. 2021, 4697–4705
- [470] Hu B, Fang Y, Shi C. Adversarial learning on heterogeneous information networks. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2019, 120–129
- [471] Ji H, Wang X, Shi C, Wang B, Yu P S. Heterogeneous graph propagation network. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(1): 521–532
- [472] Guo J, Du L, Bi W, Fu Q, Ma X, Chen X, Han S, Zhang D, Zhang Y. Homophily-oriented heterogeneous graph rewiring. In: *Proceedings of the ACM Web Conference 2023*. 2023, 511–522
- [473] Li J, Wei Z, Dan J, Zhou J, Zhu Y, Wu R, Wang B, Zhen Z, Meng C, Jin H, Zheng Z, Chen L. Hetero2Net: heterophily-aware representation

- learning on heterogeneous graphs. 2023, arXiv preprint arXiv: 2310.11664
- [474] Shen Z, Kang Z. When heterophily meets heterogeneous graphs: latent graphs guided unsupervised representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2025
- [475] Lin J, Guo X, Zhang S, Zhou D, Zhu Y, Shun J. When heterophily meets heterogeneity: new graph benchmarks and effective methods. 2024, arXiv preprint arXiv: 2407.10916
- [476] Longa A, Lachi V, Santin G, Bianchini M, Lepri B, Liò P, Scarselli F, Passerini A. Graph neural networks for temporal graphs: state of the art, open challenges, and opportunities. 2023, arXiv preprint arXiv: 2302.01018
- [477] Al Sahili Z, Awad M. Spatio-temporal graph neural networks: a survey. 2023, arXiv preprint arXiv: 2301.10569
- [478] Zhou Z, Huang Q, Lin G, Yang K, Bai L, Wang Y. GReto: remedying dynamic graph topology-task discordance via target homophily. In: *Proceedings of the 11th International Conference on Learning Representations*. 2023
- [479] Antelmi A, Cordasco G, Polato M, Scarano V, Spagnuolo C, Yang D. A survey on hypergraph representation learning. *ACM Computing Surveys*, 2024, 56(1): 24
- [480] Veldt N, Benson A R, Kleinberg J. Combinatorial characterizations and impossibilities for higher-order homophily. *Science Advances*, 2023, 9(1): eabq3200
- [481] Wang P, Yang S, Liu Y, Wang Z, Li P. Equivariant hypergraph diffusion neural operators. In: *Proceedings of the 11th International Conference on Learning Representations*. 2023
- [482] Nguyen B, Sani L, Qiu X, Liò P, Lane N D. Sheaf hypernetworks for personalized federated learning. 2024, arXiv preprint arXiv: 2405.20882
- [483] Zou M, Gan Z, Wang Y, Zhang J, Sui D, Guan C, Leng S. UniG-Encoder: a universal feature encoder for graph and hypergraph node classification. *Pattern Recognition*, 2024, 147: 110115
- [484] Wang J, Guo Y, Yang L, Wang Y. Understanding heterophily for graph neural networks. In: *Proceedings of the 41st International Conference on Machine Learning*. 2024, 2067
- [485] Zhu Q, Jiao Y, Ponomareva N, Han J, Perozzi B. Explaining and adapting graph conditional shift. 2023, arXiv preprint arXiv: 2306.03256
- [486] Mao H, Chen Z, Jin W, Han H, Ma Y, Zhao T, Shah N, Tang J. Demystifying structural disparity in graph neural networks: can one size fit all? In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. 2023, 1610
- [487] Yang J, Chen Z, Xiao T, Zhang W, Lin Y, Kuang K. Leveraging invariant principle for heterophilic graph structure distribution shifts. 2024, arXiv preprint arXiv: 2408.09490
- [488] Loveland D, Zhu J, Heimann M, Fish B, Schaub M T, Koutra D. On performance discrepancies across local homophily levels in graph neural networks. In: *Proceedings of the 2nd Learning on Graphs Conference*. 2023, 6
- [489] Chen J, Chen S, Gao J, Huang Z, Zhang J, Pu J. Exploiting neighbor effect: conv-agnostic GNN framework for graphs with heterophily. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(10): 13383–13396
- [490] Oono K, Suzuki T. Graph neural networks exponentially lose expressive power for node classification. In: *Proceedings of the 8th International Conference on Learning Representations*. 2020
- [491] Liu M, Gao H, Ji S. Towards deeper graph neural networks. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020, 338–348
- [492] Rusch T K, Bronstein M M, Mishra S. A survey on oversmoothing in graph neural networks. 2023, arXiv preprint arXiv: 2303.10993
- [493] Guo K, Cao X, Liu Z, Chang Y. Taming over-smoothing representation on heterophilic graphs. *Information Sciences*, 2023, 647: 119463
- [494] Topping J, Di Giovanni F, Chamberlain B P, Dong X, Bronstein M M. Understanding over-squashing and bottlenecks on graphs via curvature. In: *Proceedings of the 10th International Conference on Learning Representations*. 2022
- [495] Rubin J, Loomba S, Jones N S. Geodesic distributions reveal how heterophily and bottlenecks limit the expressive power of message passing neural networks. In: *Proceedings of the 2nd Learning on Graphs Conference*. 2023
- [496] Pei H, Li Y, Deng H, Hai J, Wang P, Ma J, Tao J, Xiong Y, Guan X. Multi-track message passing: tackling oversmoothing and oversquashing in graph learning via preventing heterophily mixing. In: *Proceedings of 41st International Conference on Machine Learning*. 2024
- [497] Yang C, Wu Q, Wipf D, Sun R, Yan J. How graph neural networks learn: lessons from training dynamics in function space. 2023, arXiv preprint arXiv: 2310.05105
- [498] Cui G, Wei Z. MGNN: graph neural networks inspired by distance geometry problem. In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2023, 335–347
- [499] Shi C, Pan L, Hu H, Dokmanić I. Homophily modulates double descent generalization in graph convolution networks. *Proceedings of the National Academy of Sciences of the United States of America*, 2024, 121(8): e2309504121
- [500] Zhu Y, Du Y, Wang Y, Xu Y, Zhang J, Liu Q, Wu S. A survey on deep graph generation: methods and applications. In: *Proceedings of the 1st Learning on Graphs Conference*. 2022, 47
- [501] Chanpuriya S, Rossi R A, Rao A, Mai T, Lipka N, Song Z, Musco C. An interpretable graph generative model with heterophily. 2021, arXiv preprint arXiv: 2111.03030
- [502] Jin W, Zhao L, Zhang S, Liu Y, Tang J, Shah N. Graph condensation for graph neural networks. In: *Proceedings of the 10th International Conference on Learning Representations*. 2022
- [503] Gao X, Yu J, Jiang W, Chen T, Ye G, Zhang W, Yin H. Graph condensation: a survey. *IEEE Transactions on Knowledge and Data Engineering*, 2025, 37(4): 1819–1837
- [504] Zheng X, Zhang M, Chen C, Nguyen Q V H, Zhu X, Pan S. Structure-free graph condensation: from large-scale graphs to condensed graph-free data. In: *Proceedings of the 37th International Conference on Neural Information Processing Systems*. 2023, 264
- [505] Munikoti S, Das L, Natarajan B. Scalable graph neural network-based framework for identifying critical nodes and links in complex networks. *Neurocomputing*, 2022, 468: 211–221
- [506] Ling C, Jiang J, Wang J, Thai M T, Xue L, Song J, Qiu M, Zhao L. Deep graph representation learning and optimization for influence

- maximization. In: Proceedings of the 40th International Conference on Machine Learning. 2023, 880
- [507] Panagopoulos G, Tziortziotis N, Vazirgiannis M, Malliaros F. Maximizing influence with graph neural networks. In: Proceedings of the International Conference on Advances in Social Networks Analysis and Mining. 2023, 237–244
- [508] Feng Y, Tan V Y F, Cautis B. Influence maximization via graph neural bandits. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024, 771–781
- [509] Zhou F, Cao C, Zhang K, Trajcevski G, Zhong T, Geng J. Meta-GNN: on few-shot node classification in graph meta-learning. In: Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019, 2357–2360
- [510] Wang S, Dong Y, Ding K, Chen C, Li J. Few-shot node classification with extremely weak supervision. In: Proceedings of the 16th ACM International Conference on Web Search and Data Mining. 2023, 276–284
- [511] Wan S, Zhan Y, Liu L, Yu B, Pan S, Gong C. Contrastive graph Poisson networks: semi-supervised learning with extremely limited labels. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. 2021, 483
- [512] Zhao T, Zhang X, Wang S. GraphSMOTE: imbalanced node classification on graphs with graph neural networks. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining. 2021, 833–841
- [513] Yu C, Zhu J, Li X. GraphCBAL: class-balanced active learning for graph neural networks via reinforcement learning. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024, 3022–3031
- [514] Yun S, Kim K, Yoon K, Park C. LTE4G: long-tail experts for graph neural networks. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022, 2434–2443
- [515] Wu X, Wu H, Wang R, Li D, Zhou X, Lu K. Leveraging free labels to power up heterophilic graph learning in weakly-supervised settings: an empirical study. In: Proceedings of European Conference on Machine Learning and Knowledge Discovery in Databases. 2023, 140–156
- [516] Liu Y, Ding K, Wang J, Lee V, Liu H, Pan S. Learning strong graph neural networks with weak information. In: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2023, 1559–1571
- [517] Peng T, Wu W, Yuan H, Bao Z, Zhao P, Yu X, Lin X, Liang Y, Pu Y. GraphRARE: reinforcement learning enhanced graph neural network with relative entropy. In: Proceedings of the 40th IEEE International Conference on Data Engineering (ICDE). 2024, 2489–2502
- [518] Chen J, Chen S, Bai M, Gao J, Zhang J, Pu J. SA-MLP: distilling graph knowledge from GNNs into structure-aware MLP. 2022, arXiv preprint arXiv: 2210.09609
- [519] Wu L, Lin H, Huang Y, Li S Z. Knowledge distillation improves graph structure augmentation for graph neural networks. In: Proceedings of the 36th International Conference on Neural Information Processing Systems. 2022, 858
- [520] Wang F, Zhao T, Xu J, Wang S. HC-GST: heterophily-aware distribution consistency based graph self-training. In: Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. 2024, 2326–2335
- [521] Wei L, He Z, Zhao H, Yao Q. Searching heterophily-agnostic graph neural networks. SSRN 4825405, 2024
- [522] Wei L, He Z, Zhao H, Yao Q. Enhancing intra-class information extraction for heterophilous graphs: one neural architecture search approach. 2022, arXiv preprint arXiv: 2211.10990
- [523] Wei L, Zhao H, He Z. Designing the topology of graph neural networks: a novel feature fusion perspective. In: Proceedings of the ACM Web Conference 2022. 2022, 1381–1391
- [524] Zheng X, Zhang M, Chen C, Zhang Q, Zhou C, Pan S. Auto-HeG: automated graph neural network on heterophilic graphs. In: Proceedings of the ACM Web Conference 2023. 2023, 611–620
- [525] Liu Y, Li M, Li X, Giunchiglia F, Feng X, Guan R. Few-shot node classification on attributed networks with graph meta-learning. In: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. 2022, 471–481
- [526] Zhang Y, Yang Q. A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(12): 5586–5609
- [527] Wu Y, Yao J, Han B, Yao L, Liu T. Unraveling the impact of heterophilic structures on graph positive-unlabeled learning. In: Proceedings of the 41st International Conference on Machine Learning. 2024
- [528] Xi Z, Chen W, Guo X, He W, Ding Y, Hong B, Zhang M, Wang J, Jin S, Zhou E, Zheng R, Fan X, Wang X, Xiong L, Zhou Y, Wang W, Jiang C, Zou Y, Liu X, Yin Z, Dou S, Weng R, Qin W, Zheng Y, Qiu X, Huang X, Zhan Q. The rise and potential of large language model based agents: a survey. Science China Information Sciences, 2025, 68(2): 121101
- [529] Gong C, Li X. Agents with foundation models: advance and vision. Frontiers of Computer Science, 2025, 19(4): 194330
- [530] Lee J, Stevens N, Han S C, Song M. A survey of large language models in finance (FinLLMs). 2024, arXiv preprint arXiv: 2402.02315
- [531] Zhang X, Wang L, Helwig J, Luo Y, Fu C, et al. Artificial intelligence for science in quantum, atomistic, and continuum systems. 2023, arXiv preprint arXiv: 2307.08423
- [532] Wu X, Shen Y, Shan C, Song K, Wang S, Zhang B, Feng J, Cheng H, Chen W, Xiong Y, Li D. Can graph learning improve task planning? 2024, arXiv preprint arXiv: 2405.19119
- [533] Yang S, Nachum O, Du Y, Wei J, Abbeel P, Schuurmans D. Foundation models for decision making: problems, methods, and opportunities. 2023, arXiv preprint arXiv: 2303.04129
- [534] Soleymani F, Paquet E. Deep graph convolutional reinforcement learning for financial portfolio management–DeepPocket. Expert Systems with Applications, 2021, 182: 115127
- [535] Bayraktar Z, Molla S, Mahavadi S. Graph neural network generated metal-organic frameworks for carbon capture. In: Proceedings of the 11th International Conference on Learning Representations. 2023
- [536] Wang M, Wang E, Liu X, Wang C. Topological graph representation of stratigraphic properties of spatial-geological characteristics and compression modulus prediction by mechanism-driven learning. Computers and Geotechnics, 2023, 153: 105112

- [537] Yang Q, Wang X, Zhang X, Zheng J, Ke Y, Wang L, Guo H. A novel deep learning method for automatic recognition of coseismic landslides. *Remote Sensing*, 2023, 15(4): 977
- [538] Liu Z, Wan G, Prakash B A, Lau M S, Jin W. A review of graph neural networks in epidemic modeling. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024, 6577–6587
- [539] Hamilton J D. State-space models. *Handbook of Econometrics*, 1994, 4: 3041–3080
- [540] Gu A, Dao T. Mamba: linear-time sequence modeling with selective state spaces. 2023, arXiv preprint arXiv: 2312.00752
- [541] Behrouz A, Hashemi F. Graph mamba: towards learning on graphs with state space models. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024, 119–130
- [542] Liu Z, Wang Y, Vaidya S, Ruehle F, Halverson J, Soljačić M, Hou T Y, Tegmark M. KAN: Kolmogorov-Arnold networks. 2024, arXiv preprint arXiv: 2404.19756
- [543] Ahmed T, Sifat M H R. GraphKAN: graph Kolmogorov Arnold network for small molecule-protein interaction predictions. In: *Proceedings of the 1st Machine Learning for Life and Material Sciences Workshop*. 2024
- [544] Zhang F, Zhang X. GraphKAN: enhancing feature extraction with graph Kolmogorov Arnold networks. 2024, arXiv preprint arXiv: 2406.13597
- [545] Yu J, Ren Y, Gong C, Tan J, Li X, Zhang X. Empower text-attributed graphs learning with large language models (LLMs). 2023, arXiv preprint arXiv: 2310.09872
- [546] He X, Bresson X, Laurent T, Perold A, LeCun Y, Hooi B. Harnessing explanations: LLM-to-LM interpreter for enhanced text-attributed graph representation learning. In: *Proceedings of the 12th International Conference on Learning Representations*. 2024
- [547] Wang Y, Zhu Y, Zhang W, Zhuang Y, Li Y, Tang S. Bridging local details and global context in text-attributed graphs. In: *Proceedings of 2024 Conference on Empirical Methods in Natural Language Processing*. 2024, 14830–14841
- [548] Huang X, Han K, Yang Y, Bao D, Tao Q, Chai Z, Zhu Q. GNNs as adapters for LLMs on text-attributed graphs. In: *Proceedings of the Web Conference 2024*. 2024
- [549] Chen Z, Mao H, Li H, Jin W, Wen H, Wei X, Wang S, Yin D, Fan W, Liu H, Tang J. Exploring the potential of large language models (LLMs) in learning on graphs. *ACM SIGKDD Explorations Newsletter*, 2024, 25(2): 42–61
- [550] Mao Q, Liu Z, Liu C, Li Z, Sun J. Advancing graph representation learning with large language models: a comprehensive survey of techniques. 2024, arXiv preprint arXiv: 2402.05952
- [551] Wu Y, Li S, Fang Y, Shi C. Exploring the potential of large language models for heterophilic graphs. 2024, arXiv preprint arXiv: 2408.14134
- [552] Ren X, Tang J, Yin D, Chawla N, Huang C. A survey of large language models for graphs. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2024, 6616–6626
- [553] Liu J, Yang C, Lu Z, Chen J, Li Y, Zhang M, Bai T, Fang Y, Sun L, Yu P S, Shi C. Graph foundation models: concepts, opportunities and challenges. 2023, arXiv preprint arXiv: 2310.11829
- [554] Galkin M, Yuan X, Mostafa H, Tang J, Zhu Z. Towards graph foundation models for knowledge graph reasoning. In: *Proceedings of the 12th International Conference on Learning Representations*. 2024
- [555] Mao H, Chen Z, Tang W, Zhao J, Ma Y, Zhao T, Shah N, Galkin M, Tang J. Position: graph foundation models are already here. In: *Proceedings of the 41st International Conference on Machine Learning*. 2024
- [556] Tang J, Yang Y, Wei W, Shi L, Su L, Cheng S, Yin D, Huang C. GraphGPT: graph instruction tuning for large language models. In: *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2024, 491–500
- [557] Xia L, Kao B, Huang C. OpenGraph: towards open graph foundation models. In: *Proceedings of the Association for Computational Linguistics*. 2024, 2365–2379
- [558] Xia L, Huang C. AnyGraph: graph foundation model in the wild. 2024, arXiv preprint arXiv: 2408.10700
- [559] Alabdulmohsin I, Neyshabur B, Zhai X. Revisiting neural scaling laws in language and vision. In: *Proceedings of the 36th International Conference on Neural Information Processing Systems*. 2022, 1620



Cheng-Hua GONG received his Bachelor's degree from Nanjing Normal University, China in 2022 and he is now a master student at East China Normal University, China. His research interests include graph neural networks and graph data mining.



Yao CHENG received her Master's degree from Dalian Maritime University, China in 2021 and she is now a PhD student at East China Normal University, China. Her general research interests include graph neural networks and graph foundation models.



Jianxiang YU received his Master's degree from East China Normal University in 2024 and he is now a PhD student at East China Normal University, China. His research interests include graph neural networks and natural language processing.



Can XU received his Master's degree from Zhejiang Gongshang University, China in 2024 and he is now a PhD student at East China Normal University, China. His research interests include graph neural networks and generative models.



Cai-Hua SHAN is a Senior Researcher at Microsoft Research Asia (Shanghai). She received her PhD degree in Computer Science from The University of Hong Kong, China in 2020. Her primary research interests include graph neural network and molecular generation. She has published over 30 papers at top-tier conferences such as ICML, NeurIPS, ICLR, KDD, WWW, SIGMOD, VLDB, and ICDE.



Si-Qiang LUO is a Nanyang Assistant Professor at the College of Computing and Data Science, Nanyang Technological University, Singapore. He received his Ph.D. degree in computer science from The University of Hong Kong, China, in 2019. His research interest lies in graph data management, such as graph neural networks, PageRanks, and community search. His research has been regularly published in top venues such as SIGMOD, PVLDB, and ICDE.



Xiang LI received his PhD degree from the University of Hong Kong, China in 2018. From 2018 to 2020, he worked as a research scientist in the Data Science Lab at JD.com and a research associate at The University of Hong Kong, China respectively. He is currently a research professor at the School of Data Science and Engineering in East China Normal University, China. His general research interests include Data Mining and Machine Learning.