# Graph Contrastive Learning: An Odyssey towards Generalizable, Scalable and Principled Representation Learning on Graphs

Yan Han[†]    Yuning You[‡]    Wenqing Zheng[†]    Scott Hoang[†]    Tianxin Wei[§]
Majdi Hassan[¶]    Tianlong Chen[†]    Ying Ding[†]    Yang Shen[‡]    Zhangyang Wang[†]

[†]University of Texas at Austin
{yh9442, w.zheng, hoangd, tianlong.chen, atlaswang}@utexas.edu
ying.ding@austin.utexas.edu

[‡]Texas A&M University
{yuning.you, yshen}@tamu.edu

[§]University of Illinois Urbana-Champaign
twei10@illinois.edu

[¶]AbbVie
majdi.mhas@gmail.com

## Abstract

Graph Contrastive Learning (GCL), an uprising regime of learning representations of graph-structured data, has gained significant attention in recent years. At its core, GCL leverages the idea of comparing different views of a graph to learn representations that capture desirable characteristics of the graph structures. GCL has been applied to a wide range of graph-structured data, including attributed graphs, multi-relational graphs, temporal graphs, hierarchical graphs, heterogeneous graphs, and hypergraphs. The learned graph representations yield predictive performance that generalizes well in various downstream tasks at the node, link, and graph levels, and can scale up to graphs with millions of nodes. In this paper, we present a review of representative GCL approaches with a major emphasis on our own recent efforts. Beginning with the original GCL approach with ad-hoc view generation and simple homogeneous graphs, we demonstrate how the framework can be further extended to more complex heterogeneous graphs and hypergraphs, as well as improved via principled view generation towards generalizability, fairness, interpretability, and other aspects. Theoretical explorations are covered at the end. In conclusion, we discuss the future prospects and ongoing challenges in the field of GCL.

## 1 Introduction

Graph-structured data are ubiquitous in various real-world applications, including social networks, biological networks, transportation systems, and recommender systems [25, 17, 57]. The analysis and learning from graph data have gained increasing importance as they can unveil hidden patterns and relationships, thereby enhancing decision-making in practical scenarios. In recent years, graph contrastive learning (GCL) has emerged as a promising approach for graph representation learning. GCL has demonstrated remarkable success in capturing the underlying structural properties of graphs and achieving generalized predictive performance.

The key concept behind GCL, inspired by image representation learning [5], is simple yet effective: comparing different views of a graph to learn representations that encapsulate its desirable structural properties, which can scale up to large graphs. However, determining what and how to contrast presents a non-trivial challenge when it comes to graph data. Unlike images, graph data exhibit high heterogeneity across applications in terms of both semantics and graph types (e.g., social networks versus molecular scaffolds). Consequently, achieving a universally beneficial design is exceedingly difficult.

In this paper, we present a comprehensive review of various GCL approaches, with a particular focus on our recent contributions. We start from introducing the vanilla GCL approach, known as vanilla GCL, which involves the utilization of ad-hoc view generation and primarily targets simple homogeneous graphs. Subsequently, we systematically explore the extensions of the GCL framework to accommodate more intricate heterogeneous graphs and hypergraphs. This involves the integration of principled view generation techniques to enhance generalizability, fairness, interpretability, and other important aspects. We in addition cover the preliminary but critical investigations of the theoretical aspects, providing valuable insights into the underlying mechanisms and potential limitations of GCL. We conclude the paper by discussing future prospects and open challenges in the GCL field, including potential directions for further research and development in this area.

## 2 GCL Approaches and Graph Types

### 2.1 Homogeneous Graphs

A homogeneous graph, denoted as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\} \in \mathbb{G}$, is characterized by a set of vertices $\mathcal{V} = \{v_1, ..., v_{|\mathcal{V}|}\}$ and a set of edges $\mathcal{E} = \{(v_i, v_j) | v_i, v_j \in \mathcal{V}\}$. In different contexts, each vertex and edge can be associated with specific attributes for feature representation. For example, in social networks [36], vertices may correspond to user side-information and edges to connections, while in molecular graphs [7], vertices may represent atoms and edges may represent bond types. These attributes can be mapped to vectorial representations of dimension $D$ using graph neural networks (GNNs) [24, 48, 64]. Thus, GNNs can be represented as $f_\theta : \mathbb{G} \to \mathbb{R}^D$, where $\theta$ denotes the parameters of the GNN, enabling downstream utility.

In the supervised learning setting, the annotated dataset $\mathcal{D}_{\text{lab}} = \{(\mathcal{G}_1, \mathcal{Y}_1), ...\}$ is provided, where machine learning models, such as GNNs, can be trained to make predictions on new, unseen data. However, the effectiveness of supervised learning is hindered by the challenge of limited graph labels [37]. One dominant solution to address the challenge is to employ self-supervised learning, where models are pre-trained on large-scale unlabeled data [5, 9]. The rationale behind this approach is well-appropriate principled objectives could enhance their generalizability. Fortunately, accessing unlabeled datasets $\mathcal{D}_{\text{unlab}} = \{\mathcal{G}_1, ...\}$ for graphs is often viable. A key question then arises: *how can self-supervised objectives be designed specifically for graph-structured data?*

Among the various self-supervised tasks for graphs (for a more comprehensive review, please refer to [62, 34]), graph contrastive learning (GCL) [68, 50, 79, 20, 43, 40] stands out due to its consistently generalizable performance across diverse applications. The fundamental concept of GCL is to compare different views of a graph to learn representations that capture the desired structural properties of the graph (refer to Figure 1 for the overall pipeline). In GCL, prior knowledge is explicitly incorporated into graph neural networks (GNNs) through the construction of graph views. Consequently, the primary focus of GCL research is to investigate effective methods for constructing graph views that incorporate appropriate inductive biases. The training of GCL with a batch of graph samples $\{\mathcal{G}_1, ..., \mathcal{G}_N\}$ is then formulated using the NT-Xent loss [5] as follows:

$$\min_{\theta, \phi} \mathcal{L}_{\text{GCL}}(\theta, \phi) = \min_{\theta, \phi} -\frac{1}{N} \sum_{n=1}^{N} \log \frac{\exp\left(\text{sim}\left(g_\phi \circ f_\theta(\tilde{\mathcal{G}}_{n,1}), g_\phi \circ f_\theta(\tilde{\mathcal{G}}_{n,2})\right)/\tau\right)}{\sum_{n'=1, n' \neq n}^{N} \exp\left(\text{sim}\left(g_\phi \circ f_\theta(\tilde{\mathcal{G}}_{n,1}), g_\phi \circ f_\theta(\tilde{\mathcal{G}}_{n',2})\right)/\tau\right)} \quad (25)$$

where $g_\phi(\cdot)$ represents the projection head implemented as a multi-layer perceptron [5], $\tilde{\mathcal{G}}_{n,i} = h_i(\mathcal{G}_n)$ denotes
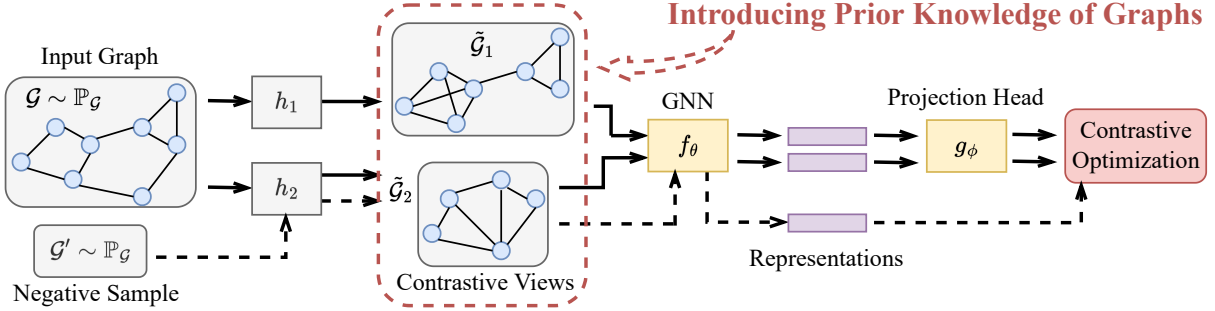
Figure 1: The generic pipeline of homogeneous GCL.

the contrastive view of $\mathcal{G}_n$, $\mathcal{G}'_n$ is treated as the negative sample, $\text{sim}(\cdot, \cdot)$ is the similarity function (typically implemented using cosine similarity $\text{sim}(z_1, z_2) = z_1^\mathsf{T} z_2 / \|z_1\|\|z_2\|$), and $\tau \in \mathbb{R}_{>0}$ is the temperature parameter.

**Graph data augmentations for view construction.** A well-established approach to construct contrastive views is through graph data augmentations [68, 80, 10, 76, 77, 35], which effectively introduce inductive biases specific to the target applications. For example, in one of the earliest works [68], four generic graph augmenta-

Table 1: Four generic augmentation strategies (in the spatial domain) with the corresponding underlying inductive bias.

| Data Augmentation | Type | Underlying Prior |
|---|---|---|
| Node dropping | Nodes, edges | Vertex missing does not alter semantics. |
| Edge perturbation | Edges | Semantic robustness against connectivity variations. |
| Attribute masking | Nodes | Semantic robustness against partial attribute loss. |
| Subgraph | Nodes, edges | Local structure can provide hints to full semantics. |

tions are designed, as shown in Table 1, each with an underlying prior incorporated. These augmentations operate in the spatial domain and include node dropping, edge perturbation, attribute masking, and subgraph transformations, the resulting views of which capture different aspects of the graph's semantics.

In addition to augmentations in the spatial domain, spectral augmentations have been proposed for graph data, leveraging the graph spectrum where the semantic information is more accessible [16, 72, 31]. Furthermore, some works explore augmentations in the latent space implicitly learned by the model [61, 26, 51], which is believed to better capture the underlying semantics. Other approaches focus on augmenting specific graph model-based parameters, such as graphons [18, 41] or contextual stochastic block models (CSBMs) [55], based on the explicit downstream data generation assumptions.

Overall, the choice of graph data augmentations plays a crucial role in constructing effective contrastive views and capturing the desired inductive biases for the specific graph-structured data. In certain designated applications, augmentations can be specifically designed to cater to the unique characteristics of the data. For example, small molecules have been shown to benefit from motif-based [54, 13] or energy-guided [33] perturbations. Similarly, protein structures can be augmented through cropping while preserving consecutive amino-acid sequences [73, 71]. The construction of effective graph views for contrastive learning, which are robust across diverse domains or exhibit strong generalizability in specific applications, remains an active area of research.

## 2.2 Heterogeneous Graphs

A heterogeneous graph is defined as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{T}_\mathcal{V}, \mathcal{T}_\mathcal{E}\} \in \mathbb{G}$, where $\mathcal{V} = \{v_1, ..., v_{|\mathcal{V}|}\}$ represents the set of vertices, $\mathcal{E} = \{(v_i, v_j)|v_i, v_j \in \mathcal{V}\}$ denotes the set of edges, $\mathcal{T}_\mathcal{V}$ denotes the set of different vertex types, and $\mathcal{T}_\mathcal{E}$ represents the set of different edge types. In various contexts, each vertex and edge in a heterogeneous graph may be associated with specific attributes (e.g., user profiles and relationships in social networks [44], or gene expressions and regulatory interactions in biological networks [39]), which serve as features for graph learning.

Differing from homogeneous graphs, the complicated relations in heterogeneous graphs are more effec-
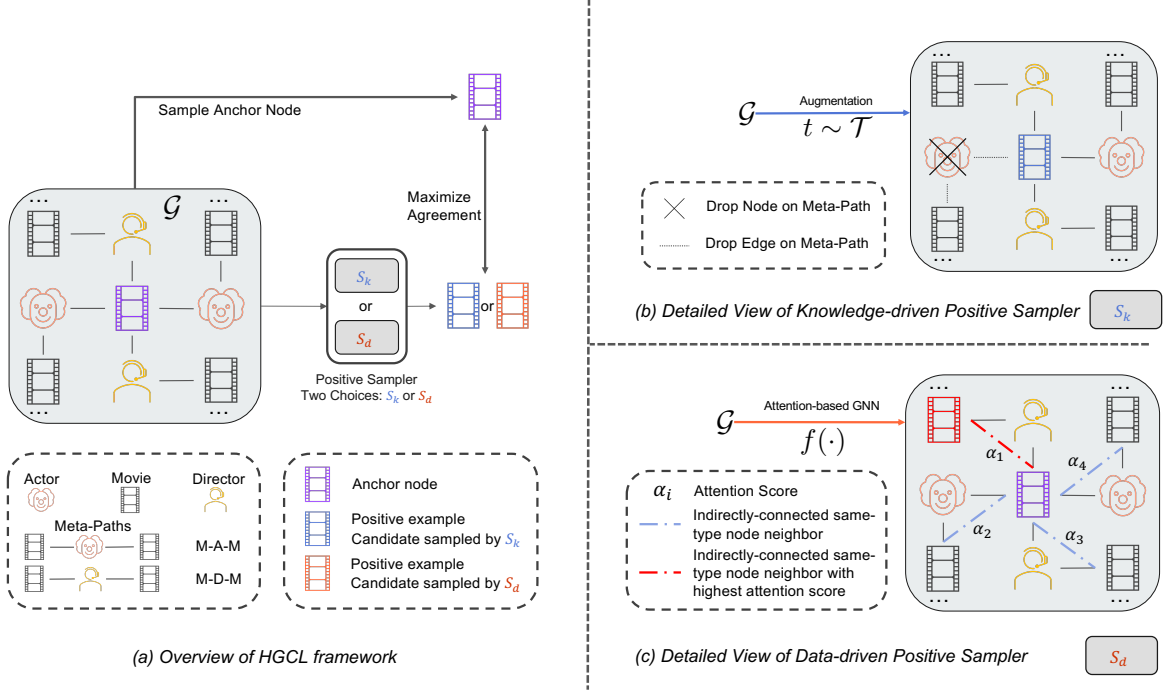
Figure 2: HGCL framework: (a) **Overview**: Anchor node $u$ and positive example node $v$ in a heterogeneous graph $\mathcal{G}$ are processed with a GNN backbone and projection head using a contrastive loss. (b) $S_k$ (**knowledge-driven**): Pre-defined meta-paths guide positive example node $v_k$ generation. (c) $S_d$ (**data-driven**): GNN-based encoder with attention module generates positive example node $v_d$.

tively captured by meta-paths. A meta-path [45, 11] $P$ is a path that connects multiple nodes in the form of $A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \ldots \xrightarrow{R_l} A_{l+1}$ (abbreviated as $A_1 A_2 \ldots A_{l+1}$). It describes a composition of different relations $R_1, R_2, \ldots, R_l$ that link objects $A_1$ and $A_{l+1}$. For example, in the heterogeneous graph shown in Figure 2(a), there are two different meta-paths: Movie-Director-Movie (MDM), which represents the co-director relation between two movies, and Movie-Actor-Movie (MAM), which represents the co-actor relation between two movies. Different meta-paths reveal diverse semantics and guide the connection of even distant objects based on their semantic similarities. Utilizing meta-paths as prior knowledge in positive sampling ensures that the selected candidates are semantically related to some extent.

Heterogeneous graph contrastive learning (HGCL) has witnessed significant progress through various methods. HeCo [53] utilizes meta-path-based random walks for self-supervised learning, while STENCIL [78] employs structural templates to encode neighborhood information. HGCLR [4] learns embeddings by contrasting multiple meta-path-derived views, and CPT-HG [23] captures node relations through pairwise contrastive learning. MVSE [75] combines intra-view and inter-view contrastive learning tasks using different meta-path-based views. Generally, contrastive learning methods can be classified into two categories: **knowledge-driven** and **data-driven**, with a primary focus on generating positive views.

**Knowledge-driven views.** As illustrated in Figure 2(b), the framework leverages meta-paths to drop nodes or perturb edges, resulting in related views. This strategy shares similarities with homogeneous graphs, with the key difference being that the augmentation in heterogeneous graphs depends on meta-paths.

**Data-driven views.** As demonstrated in Figure 2(c), data-driven methods rely on attention mechanisms employed in graph neural networks [52, 49, 15]. The attention score, $e_{ij}$, of node $j$'s embedding to node $i$ serves
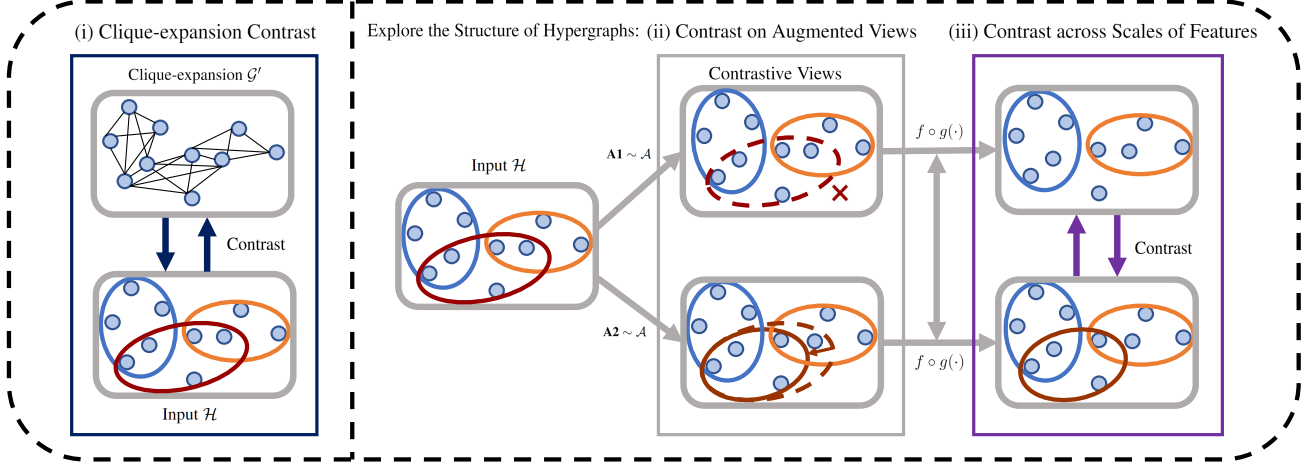
Figure 3: Three major lines of HyperGCL pipelines: (i) clique-expansion contrast, (ii) contrasting on augmented hypergraph views, and (iii) contrasting across scales of augmented hypergraph features.

for the learnable sampling distribution as:

$$e_{ij} = \text{LeakyReLU}(a(Wh_i||Wh_j)), \quad \alpha_{ij} = \text{softmax}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}i} \exp(e_{ik})} \tag{26}$$

To ensure differentiability in selecting the positive example $\mathbf{v}^+$, the Gumbel-Softmax trick [22] is introduced, using the softmax function as a continuous, differentiable approximation (Gumbel-Softmax trick), facilitating joint updates of the learned sampling distribution $p_d$ during training, resulting in a dynamic and adaptive contrastive learning framework:

$$v_j^+ = \frac{\exp((\log(\alpha_{ij}) + g_j)/\tau)}{\sum_{m=1,m\neq i}^{k} \exp(\log(\alpha_{im}) + g_m)/\tau)}, \text{ for } j = 1, \cdots, i\!\!\!/, \cdots, k \tag{27}$$

## 2.3 Hypergraphs

Hypergraphs extend the concept of homogeneous and heterogeneous graphs by allowing many-body interactions across nodes, represented by hyperedges. They have attracted significant attention in the research community [14, 65, 6]. A hypergraph is denoted as $\mathcal{H} = \{\mathcal{V}, \mathcal{E}\} \in \mathbb{H}$, where $\mathcal{V} = \{v_1, ..., v_{|\mathcal{V}|}\}$ is the set of vertices and $\mathcal{E} = \{e_1, ..., e_{|\mathcal{E}|}\}$ is the set of hyperedges. Each hyperedge $e_n = \{v_1, ..., v_{|e_n|}\}$ represents a higher-order interaction among a set of vertices. Hypergraph neural networks (HyperGNNs) [14, 65, 6] have been proposed as state-of-the-art approaches to encode such complex structures, mapping the hypergraph to a $D$-dimensional latent space via $f : \mathbb{H} \to \mathbb{R}^D$ using higher-order message passing. For contrastive learning, a projection head $h(\cdot)$ is applied to $f(\cdot)$. Currently, most hypergraph contrastive learning (HyperGCL) methods focus on node-level applications.

Table 2: Five generic augmentation operations for HyperGCL.

| Hypergraph Augmentations |
| --- |
| Naïve Hyperedge Perturbation |
| Generalized Hyperedge Perturbation |
| Vertex Dropping |
| Attribute Masking |
| Subgraph |
| Generative Augmentation |

When dealing with complex relationships in hypergraphs, the challenge is how to construct contrastive views for hypergraphs. However, building effective hypergraph views is non-trivial due to the overly complicated topology of hypergraphs. Unlike graphs, where there are $\binom{N}{2}$ possibilities for one edge with $N$ vertices, hyperedges in hypergraphs can have $2^N$ possibilities. To address this challenge, three lines of methods have

emerged in HyperGCL research, which focus on (i) clique-expansion contrast, (ii) hypergraph view generation, and (iii) hypergraph objective augmentation, as summarized in Figure 3.

**Clique-expansion contrast.** The first line of HyperGCL methods contrasts the representations of hypergraphs with clique-expansion views [59, 3]. This method is intuitive but computationally expensive in terms of time and memory, as it requires optimizing multiple neural networks of different modalities. Additionally, contrasting between clique expansion poses the risk of losing higher-order awareness by bringing the representations of hypergraphs and graphs closer together [56].

**Contrasting on augmented hypergraph views.** The second line of methods explores the structure of hypergraphs itself to construct contrastive views [56]. They assess whether generic augmentations are suitable for HyperGCL. Since hypergraphs are composed of hyperedges and vertices, they propose two strategies to augment hyperedges: direct perturbation on hyperedges and perturbation on the "edges" between hyperedges and vertices in the converted bipartite graph. To augment vertices, they adopt three schemes: vertex dropping, attribute masking, and subgraph, which are borrowed from graph-structured data augmentations [68]. Their findings show that while vertex augmentations benefit graphs more, hypergraphs mostly benefit from hyperedge augmentations, revealing that higher-order information encoded in hyperedges is usually more downstream-relevant than information in vertices.

**Contrasting across scales of augmented hypergraph features.** The last line of methods [29, 42] focuses on contrasting augmented features beyond the node-level. The main idea is to perform multi-level contrast, aiming to maximize the agreement between the same node, the node members of the same hyperedge, and each hyperedge and its node members in two augmented views. This approach expects that complementary information can be captured from different views to enhance HyperGCL.

## 2.4 Principled Graph Views for Contrastive Learning

The construction of appropriate contrastive views is crucial in GCL, but it often relies on empirical rules of thumb, which can vary significantly depending on the nature of the graph dataset. Therefore, the question arises: *could we develop principled ways to construct graph views for contrastive learning?* The answer lies in two folds: defining the space of GCL views and formulating principles to search within that space (Figure 4).

**Space of graph contrastive views.** Defining a good search space is essential for well-behaved search algorithms. The view constructor can be represented by a mapping function $h_\psi : \mathbb{G} \to \tilde{\mathbb{G}}$, where $\tilde{\mathcal{G}} = h_\psi(\mathcal{G})$ and $\psi$ represents the parameters of the mapping function. The search space is defined on this family of functions, specifically on the parameters $\psi$. Various approaches have been proposed to construct the function space, such as using learnable sampling distributions combined with prefabricated graph augmentation functions [67], masking operators on



Figure 4: Two indispensable steps for principled graph view construction: defining view space and formulating search principles.

topology or node features [80, 46], and training graph generative models to define the augmentation space in a data-driven manner [69, 66]. While the construction of search spaces is domain-agnostic, there is a need for further research to explore how to construct search spaces tailored to specific applications, e.g., in molecular graph analysis, chemical knowledge can be incorporated to define the graph augmentation functions. Future research can focus on customizing the search space construction to the characteristics of the target application.

**Principles for view searching.** The choice of graph contrastive views can be sensitive to different datasets in
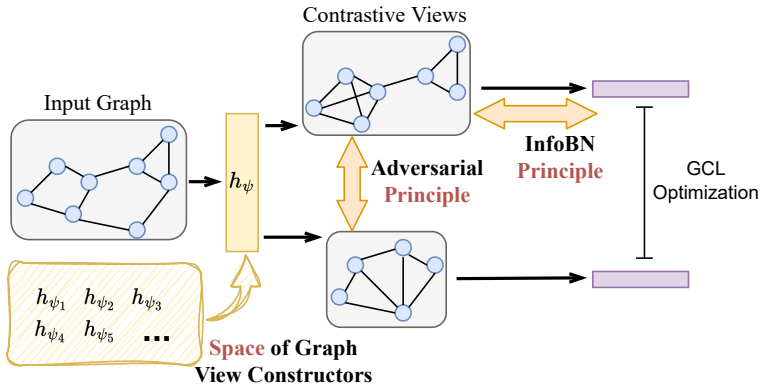
GCL. However, evidence shows that views derived from certain principles, conditioned on the dataset, provide more robust benefits. View searching can be formulated as a bi-level optimization problem, where the upper-level optimization aims to minimize the GCL loss with respect to the parameters $\theta$ and $\phi$, while the lower-level optimization finds the optimal parameters $\psi^*$ that minimize a principled objective $\mathcal{L}_{\text{Principle}}$. Various principles have been proposed for view construction. For example, adversarial training can enforce GCL to optimize on "difficult" views [67], leveraging graph properties can guide the view construction [80], and the information bottleneck principle can eliminate superficial features from raw graph data [69, 46, 63]. Future research can focus on unifying different principles and understanding the relationships among them. Additionally, theoretical analysis to bridge principled view construction and downstream performance is needed for further development.

## 2.5 Principled View towards Interpretability and Fairness

While graph contrastive learning has shown success in various tasks, fairness and interpretability are important aspects that have received less attention in this area. In this section, we discuss the principled view for fairness and interpretability in graph contrastive learning, highlighting recent studies.

**Fairness.** Fairness in graph representation learning is crucial to prevent biased results, especially towards underrepresented groups. Graph contrastive learning methods can benefit from incorporating fairness-aware data augmentations to promote fair and unbiased representations. Studies such as Graphair [32], as well as other approaches [28, 27], propose fairness-aware data augmentations that are learned from data. These augmentations, which can be integrated into graph contrastive learning frameworks, aim to mitigate sensitive information while preserving other useful information. By doing so, they improve the fairness-accuracy trade-off performance in various node classification datasets, leading to more equitable graph contrastive learning outcomes.

**Interpretability.** Interpretability is another important aspect of graph contrastive learning, as it helps users understand and trust the decisions made by graph neural networks (GNNs). Task-Agnostic GNN Explainer (TAGE) [60] is a self-supervised, task-independent explanation approach that can be applied to GNNs used in graph contrastive learning. TAGE enables the explanation of GNN embedding models with unseen downstream tasks and allows efficient explanation of multitask models. Integrating TAGE into graph contrastive learning frameworks can significantly enhance their explanation efficiency while achieving similar or better explanation quality than existing state-of-the-art GNN explanation methods.

Another approach to improve the interpretability of GNNs in graph contrastive learning is to focus on their reasoning capabilities. Existing neural reasoners often struggle with out-of-distribution (OOD) test data featuring larger input sizes. A recent study [1] proposes data augmentation procedures that leverage causal frameworks to develop self-supervised objectives, which can also be applied to graph contrastive learning. By incorporating these data augmentation procedures into graph contrastive learning, the OOD generalization capabilities of the reasoner can be improved, resulting in better performance on OOD test data.

In summary, enhancing fairness and interpretability in graph contrastive learning is crucial for developing trustworthy and unbiased node representations. By incorporating fairness-aware data augmentations, self-supervised objectives, and interpretable explanation methods, we can create more robust and equitable graph contrastive learning frameworks that remain interpretable and generalizable across various application domains.

## 2.6 Theoretical Exploration

Although numerous algorithms and principles have been developed for GCL pretraining, the explicit theoretical connection between pretraining and downstream fine-tuning performance still lags behind. The gap between pretraining and fine-tuning performance can be attributed to several factors, including the misalignment between the optimization objectives of pretraining and fine-tuning, the discrepancy between the data distributions used in pretraining and fine-tuning, and the inductive biases encoded in neural networks. While the latter two factors have been more heavily studied in graph out-of-distribution generation [2, 70], little work has been done to analyze

the sources of error arising from the misalignment between pretraining and fine-tuning optimization objectives, especially in the context of GCL.

One major challenge in analyzing the misalignment between pretraining and fine-tuning in GCL is the diverse range of downstream applications for graph-structured data. Unlike Euclidean data, which has more standardized downstream tasks such as image classification [12, 19], graph-structured data is used in a wide range of applications, from molecule generation to social network analysis. This diversity makes it difficult to establish appropriate assumptions about downstream data distributions for analysis.

In a recent study, Trivedi et al. [47] attempted to bridge the gap between pretraining and fine-tuning in GCL by assuming a highly generic "label-preserving" behavior of graph views. They justified this assumption through the use of graph edit distances between contrastive views and label-preserved samples, which measure the similarity between two graphs based on the number of operations required to transform one into the other. While this approach provides a starting point for analyzing the misalignment between pretraining and fine-tuning in GCL, more fine-grained analyses are needed to fully understand the relationship between these two processes.

# 3    Applications and Benchmarks

## 3.1    Graph Types

The benchmarks and datasets used to evaluate graph contrastive learning methods come from a variety of domains and cover a range of graph types and tasks. These graph types include bioinformatics, social networks, molecules, computer vision, and synthetic graphs. Each of these types is characterized by unique properties that affect the tasks they are suited for and the evaluation metrics used to assess contrastive learning methods.

**TUDataset.** TUDataset [38] is extensively used in the evaluation of graph contrastive learning methods. It contains a diverse set of graph data from various domains, including small molecules and proteins, computer vision, and social networks. The datasets within TUDataset cover different types of relational networks, including graphs with discrete or continuous node and edge attributes. The small molecules datasets contain class labels representing toxicity or biological activity, with the graphs representing molecules and nodes representing atoms, while edges represent chemical bonds. The bioinformatics datasets in TUDataset represent macromolecules such as proteins and use a graph model where nodes represent secondary structure elements, annotated by their type, and several physical and chemical information, with edges connecting neighboring nodes. The computer vision datasets contain graphs representing various tasks such as image processing, fingerprint recognition, and letter recognition. Finally, the social network datasets within TUDataset include Reddit discussion threads, scientific collaboration networks, actor collaborations, and GitHub users, each with different tasks, such as distinguishing between discussion-based and question-answer-based subreddits, predicting the research field of researchers, predicting the genre of actor collaborations, and identifying GitHub users who starred popular repositories.

**Pokec-z and Pokec-n**. Pokec-z and Pokec-n [8] are two social graphs sampled from a larger Facebook-like social network in Slovakia called Pokec. The nodes in these graphs correspond to users living in two major regions, with the region information being used as the sensitive attribute. The Recidivism graph is built upon the information of defendants who got released on bail at the US state courts, where the edges are created based on the similarity of past criminal records and demographics. The sensitive attribute for this graph is race, where the node classification task is built upon classifying defendants into bail or no bail. Similarly, the Credit defaulter graph is generated by creating links between people based on the similarity of their spending and payment patterns, with labels for node classification corresponding to whether a person will handle the credit card payment or not, and age being used as the sensitive attribute.

**MoleculeNet**. MoleculeNet [58] is a collection of molecular graph datasets used for the prediction of different molecule properties. Each atom in the molecule is considered a node in the graph, with each bond considered an edge. The prediction of molecule properties is a graph-level task, and three graph classification tasks from MoleculeNet are used in the evaluation of contrastive learning methods.

**PPI**. The Protein-Protein Interaction (PPI) [81] dataset documents the physical interactions between proteins in 24 different human tissues. In PPI graphs, each protein is considered as a node with its motif and immunological features, and there is an edge between two proteins if they interact with each other. The prediction of each protein function is considered an individual task instead of a multi-class classification, and hence typical approaches require individual explainers for the 121 tasks.

**ACM**. The ACM [74] heterogeneous graph dataset comprises academic papers divided into three classes based on research areas. Each paper is associated with multiple attributes, such as authors, subjects, and publication venues. With an average of 3.33 authors per paper and one subject, the ACM dataset captures the collaboration and knowledge-sharing dynamics among researchers in various disciplines.

**DBLP**. In DBLP [15] heterogeneous graph dataset, the target nodes represent authors classified into four research areas. The dataset captures the relationships between authors, their publications, and the venues in which they are published. With an average of 4.84 papers per author, the DBLP dataset offers insights into the academic productivity and research contributions of authors across different fields.

**Freebase**. The Freebase [30] heterogeneous graph dataset contains information about movies, which are categorized into three genres. The dataset captures various relationships, such as the ones between movies, actors, directors, and writers. With an average of 18.7 actors, 1.07 directors, and 1.83 writers per movie, the Freebase dataset provides a comprehensive view of the complex interactions among entities in the film industry.

**AMiner**. The AMiner dataset [21] is a heterogeneous graph that focuses on academic papers extracted from a subset of the original dataset, divided into four research areas. In addition to capturing relationships between papers, authors, and references, the dataset includes an average of 2.74 authors and 8.96 references per paper. This comprehensive dataset offers valuable insights into the citation patterns, research trends, and collaborations in the academic world.

## 3.2   Special Tasks

In addition to the normal evaluation setting, GCL methods are evaluated in various special learning settings [68, 67, 69, 56, 28, 27, 32, 60], including semi-supervised learning, unsupervised representation learning, transfer learning, adversarial robustness, and the accuracy-fairness trade-off. These settings are important because they reflect real-world scenarios and challenges that GCL methods may encounter in practical applications, such as limited labeled data, unsupervised learning, transferability, robustness, and fairness. By evaluating the performance of GCL methods in these special settings, we can gain a better understanding of their strengths and limitations and improve their applicability and robustness in real-world scenarios.

**Semi-supervised learning**. Semi-supervised learning addresses the issue of limited labeled data in graph classification. Existing works on graph contrastive learning evaluate their models in this setting using small social network benchmarks and large-scale graph datasets. They compare their models' performance against conventional pre-training schemes such as adjacency information reconstruction and local and global representation consistency enforcement.

**Unsupervised learning**. Unsupervised learning is another important setting for graph contrastive learning because it allows for the learning of representations without any labeled data, which is useful for many applications where labeled data is scarce. Existing works on graph contrastive learning evaluate their models in unsupervised learning tasks using graph embeddings generated by unsupervised methods, which are then fed into a downstream SVM classifier.

**Transfer learning**. Transfer learning enables the evaluation of a model's transferability across different datasets, tasks, and domains. Existing works on graph contrastive learning evaluate their models in transfer learning tasks on molecular property prediction in chemistry and protein function prediction in biology. They pre-train and fine-tune their models on different datasets and evaluate the transferability of their pre-training schemes.

**Adversarial robustness**. Adversarial robustness is an important setting for graph contrastive learning because

it addresses the issue of adversarial attacks on graph data, which are becoming increasingly common in many applications. Existing works on graph contrastive learning evaluate their models' robustness on synthetic data to classify the component number in graphs, facing the RandSampling, GradArgmax, and RL-S2V attacks. They show that graph contrastive learning boosts GNN robustness compared to training from scratch under these evasion attacks.

**Accuracy and fairness trade-off**. The trade-off between accuracy and fairness is a crucial setting for graph contrastive learning because it enables the evaluation of a model's performance with respect to both accuracy and fairness metrics. Existing works on graph contrastive learning compare the accuracy-fairness trade-off performance of their models with several baselines using demographic parity as the fairness metric. They show that their model achieves the best ACC-DP trade-off compared to all fairness-aware baselines on three datasets.

# 4   Conclusion

The paper provides an overview of GCL, an emerging pipeline for generating generalizable graph representations that are crucial in real-world graph applications. The paper first introduces the vanilla formulations of GCL on various graph types and then discusses advanced variants guided by different principles, such as robustness, interpretability, and fairness. The paper also covers routine assessment and datasets commonly used for evaluating GCL methods. Based on the reviewed foundations, the paper summarizes several future perspectives and open challenges in the field of GCL.

**Tailored pipeline for focused applications.** While the reviewed methods are generally designed to be task-agnostic, real-world applications often require more focused approaches that leverage domain-specific knowledge. Designing tailored GCL pipelines that address domain-specific problems can leverage algorithmic advancements and incorporate domain priors. Examples of such tailored pipelines include applications in the fields of molecules and proteins. It is expected that more sophisticated and effective designs will be developed in the future.

**Explicit theoretical bridge between pre-training and downstream.** A notable phenomenon in graph pre-training is negative transfer, where inappropriate pre-training strategies can lead to performance degradation. This is due to the heterogeneous nature of graph data and the absence of a universal good prior for downstream tasks. In addition to being guided by implicit principles, there is a need for an explicit understanding of the relationship between GCL pre-training and downstream performance in theory. This will help answer questions about why and when to use pre-trained graph representations.

**Handling more complicated and composed graph structures.** With the rise of artificial general intelligence and digital medicine, the field of graph learning is entering an era of multi-modality learning, where in-silico models serve as proxies for real-world intelligent and biological systems. In this context, data structures are becoming more complex, with multi-modal graphs that are heterogeneous and contain multi-body relations. This presents a new frontier for graph learning, particularly in the area of generalizable representation learning that can handle extremely complex and composed graph data structures.

**Unifying graph pre-training strategies under the umbrella of GCL.** One advantage of the GCL pipeline is its simplicity and flexibility, allowing for the incorporation of various strategies into a unified framework. A future ambition is to further unify different lines of graph pre-training strategies, including both predictive and generative approaches, within this framework. This will provide a better understanding of the theoretical underpinnings and the optimal approach for specific applications by exploring the full space of graph pre-training methods within the GCL framework.

# References

[1] Beatrice Bevilacqua, Kyriacos Nikiforou, Borja Ibarz, Ioana Bica, Michela Paganini, Charles Blundell, Jovana Mitrovic, and Petar Veličković. Neural algorithmic reasoning with causal regularisation. arXiv preprint arXiv:2302.10258, 2023.

[2] Beatrice Bevilacqua, Yangze Zhou, and Bruno Ribeiro. Size-invariant graph representations for graph classification extrapolations. In International Conference on Machine Learning, pages 837–851. PMLR, 2021.

[3] Derun Cai, Chenxi Sun, Moxian Song, Baofeng Zhang, Shenda Hong, and Hongyan Li. Hypergraph contrastive learning for electronic health records. In Proceedings of the 2022 SIAM International Conference on Data Mining (SDM), pages 127–135. SIAM, 2022.

[4] Mengru Chen, Chao Huang, Lianghao Xia, Wei Wei, Yong Xu, and Ronghua Luo. Heterogeneous graph contrastive learning for recommendation. In Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, pages 544–552, 2023.

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In International conference on machine learning, pages 1597–1607. PMLR, 2020.

[6] Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. You are allset: A multiset function framework for hypergraph neural networks. ICLR, 2022.

[7] David P Clark and Nanette J Pazdernik. Molecular biology. Elsevier, 2012.

[8] Enyan Dai and Suhang Wang. 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, pages 680–688, 2021.

[9] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[10] Kaize Ding, Zhe Xu, Hanghang Tong, and Huan Liu. Data augmentation for deep graph learning: A survey. ACM SIGKDD Explorations Newsletter, 24(2):61–77, 2022.

[11] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pages 135–144, 2017.

[12] Simon S Du, Wei Hu, Sham M Kakade, Jason D Lee, and Qi Lei. Few-shot learning via learning the representation, provably. arXiv preprint arXiv:2002.09434, 2020.

[13] Yin Fang, Qiang Zhang, Haihong Yang, Xiang Zhuang, Shumin Deng, Wen Zhang, Ming Qin, Zhuo Chen, Xiaohui Fan, and Huajun Chen. Molecular contrastive learning with chemical element knowledge graph. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 3968–3976, 2022.

[14] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 3558–3565, 2019.

[15] Xinyu Fu, Jiani Zhang, Ziqiao Meng, and Irwin King. Magnn: Metapath aggregated graph neural network for heterogeneous graph embedding. Proceedings of The Web Conference 2020, Apr 2020.

[16] Amur Ghose, Yingxue Zhang, Jianye Hao, and Mark Coates. Spectral augmentations for graph contrastive learning. In International Conference on Artificial Intelligence and Statistics, pages 11213–11266. PMLR, 2023.

[17] William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. arXiv preprint arXiv:1706.02216, 2017.

[18] Xiaotian Han, Zhimeng Jiang, Ninghao Liu, and Xia Hu. G-mixup: Graph data augmentation for graph classification. In International Conference on Machine Learning, pages 8230–8248. PMLR, 2022.

[19] Jeff Z HaoChen, Colin Wei, Adrien Gaidon, and Tengyu Ma. Provable guarantees for self-supervised deep learning with spectral contrastive loss. Advances in Neural Information Processing Systems, 34:5000–5011, 2021.

[20] Kaveh Hassani and Amir Hosein Khasahmadi. Contrastive multi-view representation learning on graphs. In International conference on machine learning, pages 4116–4126. PMLR, 2020.

[21] Binbin Hu, Yuan Fang, and Chuan Shi. Adversarial learning on heterogeneous information networks. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 120–129, 2019.

[22] Eric Jang, Shixiang Gu, and Ben Poole. Categorical reparameterization with gumbel-softmax. arXiv preprint arXiv:1611.01144, 2016.

[23] Xunqiang Jiang, Yuanfu Lu, Yuan Fang, and Chuan Shi. Contrastive pre-training of gnns on heterogeneous graphs. In Proceedings of the 30th ACM International Conference on Information & Knowledge Management, pages 803–812, 2021.

[24] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016.

[25] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In ICLR. OpenReview.net, 2017.

[26] Kezhi Kong, Guohao Li, Mucong Ding, Zuxuan Wu, Chen Zhu, Bernard Ghanem, Gavin Taylor, and Tom Goldstein. Flag: Adversarial data augmentation for graph neural networks. arXiv preprint arXiv:2010.09891, 2020.

[27] O Deniz Kose and Yanning Shen. Fair node representation learning via adaptive data augmentation. arXiv preprint arXiv:2201.08549, 2022.

[28] Oyku Deniz Kose and Yanning Shen. Fair contrastive learning on graphs. IEEE Transactions on Signal and Information Processing over Networks, 8:475–488, 2022.

[29] Dongjin Lee and Kijung Shin. I'm me, we're us, and i'm us: Tri-directional contrastive learning on hypergraphs. arXiv preprint arXiv:2206.04739, 2022.

[30] Xiang Li, Danhao Ding, Ben Kao, Yizhou Sun, and Nikos Mamoulis. Leveraging meta-path contexts for classification in heterogeneous information networks. In 2021 IEEE 37th International Conference on Data Engineering (ICDE), pages 912–923. IEEE, 2021.

[31] Lu Lin, Jinghui Chen, and Hongning Wang. Spectral augmentation for self-supervised learning on graphs. arXiv preprint arXiv:2210.00643, 2022.

[32] Hongyi Ling, Zhimeng Jiang, Youzhi Luo, Shuiwang Ji, and Na Zou. Learning fair graph representations via automated data augmentations. In The Eleventh International Conference on Learning Representations, 2023.

[33] Shengchao Liu, Hongyu Guo, and Jian Tang. Molecular geometry pretraining with se (3)-invariant denoising distance matching. arXiv preprint arXiv:2206.13602, 2022.

[34] Yixin Liu, Ming Jin, Shirui Pan, Chuan Zhou, Yu Zheng, Feng Xia, and Philip Yu. Graph self-supervised learning: A survey. IEEE Transactions on Knowledge and Data Engineering, 2022.

[35] Maria Marrium and Arif Mahmood. Data augmentation for graph data: Recent advancements. arXiv preprint arXiv:2208.11973, 2022.

[36] J Clyde Mitchell. Social networks. Annual review of anthropology, 3(1):279–299, 1974.

[37] Tom Michael Mitchell et al. Machine learning, volume 1. McGraw-hill New York, 2007.

[38] Christopher Morris, Nils M Kriege, Franka Bause, Kristian Kersting, Petra Mutzel, and Marion Neumann. Tudataset: A collection of benchmark datasets for learning with graphs. arXiv preprint arXiv:2007.08663, 2020.

[39] Jiajie Peng, Yuxian Wang, Jiaojiao Guan, Jingyi Li, Ruijiang Han, Jianye Hao, Zhongyu Wei, and Xuequn Shang. An end-to-end heterogeneous graph representation learning-based framework for drug–target interaction prediction. Briefings in bioinformatics, 22(5):bbaa430, 2021.

[40] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. Gcc: Graph contrastive coding for graph neural network pre-training. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pages 1150–1160, 2020.

[41] Luana Ruiz, Luiz Chamon, and Alejandro Ribeiro. Graphon neural networks and the transferability of graph neural networks. Advances in Neural Information Processing Systems, 33:1702–1712, 2020.

[42] Yumeng Song, Yu Gu, Tianyi Li, Jianzhong Qi, Zhenghao Liu, Christian S Jensen, and Ge Yu. Chgnn: A semi-supervised contrastive hypergraph learning network. arXiv preprint arXiv:2303.06213, 2023.

[43] Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. arXiv preprint arXiv:1908.01000, 2019.

[44] Yizhou Sun and Jiawei Han. Mining heterogeneous information networks: principles and methodologies. Synthesis Lectures on Data Mining and Knowledge Discovery, 3(2):1–159, 2012.

[45] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S Yu, and Tianyi Wu. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. Proceedings of the VLDB Endowment, 4(11):992–1003, 2011.

[46] Susheel Suresh, Pan Li, Cong Hao, and Jennifer Neville. Adversarial graph augmentation to improve graph contrastive learning. Advances in Neural Information Processing Systems, 34:15920–15933, 2021.

[47] Puja Trivedi, Ekdeep S Lubana, Mark Heimann, Danai Koutra, and Jayaraman Thiagarajan. Analyzing data-centric properties for graph contrastive learning. Advances in Neural Information Processing Systems, 35:14030–14043, 2022.

[48] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. arXiv preprint arXiv:1710.10903, 2017.

[49] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks, 2018.

[50] Petar Velickovic, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. ICLR (Poster), 2(3):4, 2019.

[51] Vikas Verma, Meng Qu, Kenji Kawaguchi, Alex Lamb, Yoshua Bengio, Juho Kannala, and Jian Tang. Graphmix: Improved training of gnns for semi-supervised learning. In Proceedings of the AAAI conference on artificial intelligence, volume 35, pages 10024–10032, 2021.

[52] Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. Heterogeneous graph attention network. In The World Wide Web Conference, pages 2022–2032, 2019.

[53] Xiao Wang, Nian Liu, Hui Han, and Chuan Shi. Self-supervised heterogeneous graph neural network with co-contrastive learning. In Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining, pages 1726–1736, 2021.

[54] Yuyang Wang, Rishikesh Magar, Chen Liang, and Amir Barati Farimani. Improving molecular contrastive learning via faulty negative mitigation and decomposed fragment contrast. Journal of Chemical Information and Modeling, 62(11):2713–2725, 2022.

[55] Rongzhe Wei, Haoteng Yin, Junteng Jia, Austin R Benson, and Pan Li. Understanding non-linearity in graph neural networks from the bayesian-inference perspective. arXiv preprint arXiv:2207.11311, 2022.

[56] Tianxin Wei, Yuning You, Tianlong Chen, Yang Shen, Jingrui He, and Zhangyang Wang. Augmentations in hypergraph contrastive learning: Fabricated and generative. arXiv preprint arXiv:2210.03801, 2022.

[57] Shiwen Wu, Wentao Zhang, Fei Sun, and Bin Cui. Graph Neural Networks in Recommender Systems: A Survey. arXiv:2011.02260 [cs], 2020.

[58] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. Chemical science, 9(2):513–530, 2018.

[59] Lianghao Xia, Chao Huang, Yong Xu, Jiashu Zhao, Dawei Yin, and Jimmy Xiangji Huang. Hypergraph contrastive collaborative filtering. arXiv preprint arXiv:2204.12200, 2022.

[60] Yaochen Xie, Sumeet Katariya, Xianfeng Tang, Edward Huang, Nikhil Rao, Karthik Subbian, and Shuiwang Ji. Task-agnostic graph explanations. arXiv preprint arXiv:2202.08335, 2022.

[61] Yaochen Xie, Zhao Xu, and Shuiwang Ji. Self-supervised representation learning via latent graph prediction. In International Conference on Machine Learning, pages 24460–24477. PMLR, 2022.

[62] Yaochen Xie, Zhao Xu, Jingtun Zhang, Zhengyang Wang, and Shuiwang Ji. Self-supervised learning of graph neural networks: A unified review. IEEE transactions on pattern analysis and machine intelligence, 2022.

[63] Dongkuan Xu, Wei Cheng, Dongsheng Luo, Haifeng Chen, and Xiang Zhang. Infogcl: Information-aware graph contrastive learning. Advances in Neural Information Processing Systems, 34:30414–30425, 2021.

[64] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? arXiv preprint arXiv:1810.00826, 2018.

[65] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. Advances in neural information processing systems, 32, 2019.

[66] Yihang Yin, Qingzhong Wang, Siyu Huang, Haoyi Xiong, and Xiang Zhang. Autogcl: Automated graph contrastive learning via learnable view generators. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 8892–8900, 2022.

[67] Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. Graph contrastive learning automated. In International Conference on Machine Learning, pages 12121–12132. PMLR, 2021.

[68] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. Advances in neural information processing systems, 33:5812–5823, 2020.

[69] Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. Bringing your own view: Graph contrastive learning without prefabricated data augmentations. In Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, pages 1300–1309, 2022.

[70] Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. Graph domain adaptation via theory-grounded spectral regularization. In The Eleventh International Conference on Learning Representations, 2023.

[71] Yuning You and Yang Shen. Cross-modality and self-supervised protein embedding for compound–protein affinity and contact prediction. Bioinformatics, 38(Supplement_2):ii68–ii74, 2022.

[72] Yifei Zhang, Hao Zhu, Zixing Song, Piotr Koniusz, and Irwin King. Spectral feature augmentation for graph contrastive learning and beyond. arXiv preprint arXiv:2212.01026, 2022.

[73] Zuobai Zhang, Minghao Xu, Arian Jamasb, Vijil Chenthamarakshan, Aurelie Lozano, Payel Das, and Jian Tang. Protein representation learning by geometric structure pretraining. arXiv preprint arXiv:2203.06125, 2022.

[74] Jianan Zhao, Xiao Wang, Chuan Shi, Zekuan Liu, and Yanfang Ye. Network schema preserving heterogeneous information network embedding. In International Joint Conference on Artificial Intelligence (IJCAI), 2020.

[75] Jianan Zhao, Qianlong Wen, Shiyu Sun, Yanfang Ye, and Chuxu Zhang. Multi-view self-supervised heterogeneous graph embedding. In Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part II 21, pages 319–334. Springer, 2021.

[76] Tong Zhao, Yozen Liu, Leonardo Neves, Oliver Woodford, Meng Jiang, and Neil Shah. Data augmentation for graph neural networks. In Proceedings of the aaai conference on artificial intelligence, volume 35, pages 11015–11023, 2021.

[77] Jiajun Zhou, Jie Shen, and Qi Xuan. Data augmentation for graph classification. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pages 2341–2344, 2020.

[78] Yanqiao Zhu, Yichen Xu, Hejie Cui, Carl Yang, Qiang Liu, and Shu Wu. Structure-enhanced heterogeneous graph contrastive learning. In Proceedings of the 2022 SIAM International Conference on Data Mining (SDM), pages 82–90. SIAM, 2022.

[79] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep graph contrastive representation learning. arXiv preprint arXiv:2006.04131, 2020.

[80] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph contrastive learning with adaptive augmentation. In Proceedings of the Web Conference 2021, pages 2069–2080, 2021.

[81] Marinka Zitnik and Jure Leskovec. Predicting multicellular function through multi-layer tissue networks. Bioinformatics, 33(14):i190–i198, 2017.