A Survey of Deep Graph Clustering: Taxonomy, Challenge, and Application

Yue Liu\(^1\), Jun Xia\(^2\), Sihang Zhou\(^3\), Siwei Wang\(^1\), Xifeng Guo\(^1\), Xihong Yang\(^1\), Ke Liang\(^1\), Wenxuan Tu\(^1\), Stan Z. Li\(^2\), Xinwang Liu\(^1\)†

\(^1\)College of Computer Science and Technology, National University of Defense Technology
\(^2\)School of Engineering, Westlake University
\(^3\)College of Intelligence Science and Technology, National University of Defense Technology

yueliu@nudt.edu.cn

Abstract

Graph clustering, which aims to divide the nodes in the graph into several distinct clusters, is a fundamental and challenging task. In recent years, deep graph clustering methods have been increasingly proposed and achieved promising performance. However, the corresponding survey paper is scarce and it is imminent to make a summary in this field. From this motivation, this paper makes the first comprehensive survey of deep graph clustering. Firstly, the detailed definition of deep graph clustering and the important baseline methods are introduced. Besides, the taxonomy of deep graph clustering methods is proposed based on four different criteria including graph type, network architecture, learning paradigm, and clustering method. In addition, through the careful analysis of the existing works, the challenges and opportunities from five perspectives are summarized. At last, the applications of deep graph clustering in four domains are presented. It is worth mentioning that a collection of state-of-the-art deep graph clustering methods including papers, codes, and datasets is available on GitHub\(^1\). We hope this work will serve as a quick guide and help researchers to overcome challenges in this vibrant field.

1 Introduction

Graph clustering is a fundamental and challenging task to separate the nodes into different groups in an unsupervised manner. In recent years, benefiting from the strong graph representation capability of deep graph neural networks (GNNs) [Kipf and Welling, 2017; Velicković et al., 2017; Kipf and Welling, 2016], deep graph clustering has witnessed fruitful advances. However, unlike the deep clustering area [Zhou et al., 2022; Ren et al., 2022; Aljalbout et al., 2018; Min et al., 2018], the survey paper of deep graph clustering is scarce. In order to assist researchers in reviewing, summarizing, and planning for the future, a comprehensive survey of deep graph clustering is expected. In this paper, we begin by introducing the general pipeline and important baselines in this fast-growing field. And then, the taxonomy, challenges, and applications will be described in detail.

Firstly, the general pipeline of deep graph clustering is demonstrated in Figure 1. To be specific, the encoding neural network \(\mathcal{F}\) is trained in a self-supervised manner and embed the nodes into the latent space. Subsequently, the designed clustering method \(C\) separates the node embeddings \(\mathbf{Z}\) into several disjoint clusters. The detailed formulations and the important baselines can be found in Section 2.1.

As shown in Figure 2, we contribute a structured taxonomy to provide a broad overview of this field, which categorizes existing works from four perspectives: graph type, network architecture, learning paradigm, and clustering method. More specifically, the input graph type can be classified into...
four distinct categories: pure structure graph, attribute graph, heterogeneous graph, and dynamic graph. The characteristic of each graph type is analyzed and the corresponding processing method is introduced. In addition, for the network architecture, the existing deep graph clustering methods are grouped into multi-layer perceptron based (MLP-based) methods, graph neural network based (GNN-based) methods, and the hybrid methods. The benefits and drawbacks of each type are carefully discussed. Moreover, the learning paradigms are classified into reconstructive, adversarial, contrastive, and hybrid ones. For each learning paradigm, the general pipeline is summarized in detail. At last, the clustering methods are separated into traditional clustering methods and neural clustering methods. Here, the advantages and disadvantages of traditional clustering are simply analyzed and the technological evolution of the neural clustering is summarized in depth. We elaborate on the taxonomy in Section 3.

Despite the remarkable progress, this fast-growing field is still fraught with several crucial challenges. Thus, this section summarizes the challenges of deep graph clustering in Figure 3. Specifically, the main challenges contain the problem of graph data quality, stability, scalability, discriminative capability, and unknown cluster number. The detailed analysis and the potential solutions are provided in Section 4.

• The difficult challenges in the deep graph clustering field are summarized from five perspectives. Besides, through the careful analysis, the possible technical solutions are provided.
• A collection of state-of-the-art deep graph clustering methods including papers, codes, and datasets is shared on the GitHub repository.

2 Deep Graph Clustering

In this section, we first introduce the basic notation and definition of deep graph clustering. Then the important deep graph clustering baselines and the related applications are detailed.

2.1 Notation & Definition

The basic notations in this paper are summarized in Table 1. Take the attribute graph $G_A = (V, E, X)$ as the example input of deep graph clustering, where $V = \{v_1, v_2, \ldots, v_N\}$ is the node set of $N$ nodes with $K$ classes, $E$ is a set of edges, and $X \in \mathbb{R}^{N \times D}$ is the node attribute matrix. Also, with matrix form, $G_A$ is represented by the adjacency matrix $A \in \mathbb{R}^{N \times N}$ and the attribute matrix $X \in \mathbb{R}^{N \times D}$. For other types of the input graph, please check in Section 3.1.

As shown in Figure 1, the target of deep graph clustering is to encode nodes with neural networks and divide them into different clusters. To be specific, the self-supervised neural network $F$ first encodes the nodes in the attribute graph $G_A$ as follows.

$$Z = F(G_A) = F(X, A).$$

Besides, the self-supervised neural network $F$ is trained with pre-text tasks like reconstruction, contrast, etc. The details of network architecture and network learning paradigm can be found in Section 3.2 and Section 3.3, respectively. After encoding, the clustering method $C$ groups the nodes into several disjoint clusters as follows.

$$\Phi = C(Z, K),$$

where $K$ denotes cluster number and $\Phi \in \mathbb{R}^{N \times K}$ is the clustering assignment matrix. The details of the clustering method $C$ can be found in Section 3.4.

2.2 Important Baselines

Figure 4 summarizes the important baselines in the field of deep graph clustering. These papers have been published in influential international conferences or high-quality journals in the domains of data mining, artificial intelligence, machine learning, computer vision, and multimedia, etc. Besides, until this survey is finished, these articles are at least cited by ten times per year and the corresponding code is available and solid. Next, we introduce these important baselines in detail.

At early stage, motivated by the great success of deep learning, researchers aim to endow the graph clustering methods with strong representation capability of deep neural networks. Concretely, the pioneers [Tian et al., 2014] adopt the sparse auto-encoder [Ng and others, 2011] to learn the non-linear node representations and then perform K-means [Hartigan and Wong, 1979] to separate the node embeddings into disjoint clusters in the GraphEncoder model. After that, DNGR [Cao et al., 2016] is developed to capture
GAE/VGAE [Kipf and Welling, 2016] is proposed with the attributes. To learn both attribute and structural information, graph structural information by the random surfing model. Subsequently, SDCN [Bo et al., 2020] and DFCN [Tu et al., 2020] verify the effectiveness of integrating structural information and attribute information by the delivery operator and the information fusion module, respectively. Then, to avoid the expensive costs of spectral clustering, [Bianchi et al., 2020] formulate a continuous relaxation of the normalized minCut problem and optimize the clustering objective with the GNN. It is worth mentioning that some graph pooling methods [Tsitsulin et al., 2020; Ying et al., 2018; Lee et al., 2019] also contribute to this field. Then, O2MAC [Fan et al., 2020], and MAGCN [Cheng et al., 2021] make attempts to exploit deep neural networks for attributed multi-view graph clustering via the multiple graph view reconstruction, and the view-consistency information learning, respectively. Besides, a self-supervised method SGCMC [Xia et al., 2022] utilizes the clustering labels to guide the network learning, thus improving clustering performance.

More recently, contrastive learning has become the mainstream paradigm in the domains of vision [Hjelm et al., 2018; He et al., 2020; Chen et al., 2020a; Grill et al., 2020; Zbontar et al., 2021] and graph [Velickovic et al., 2019; Zhu et al., 2020; Xia et al., 2022a; You et al., 2020; Thakoor et al., 2021; Liang et al., 2022; Zhu et al., 2021], and contrastive deep graph clustering methods are increasingly proposed. To be specific, in the AGE model, [Cui et al., 2020] firstly filters the high-frequency noises in node attributes and then trains the encoder by adaptively discriminating the positive and negative samples. In the same year, MVGRL [Hassani and Khasahmadi, 2020] generates an augmented structural view and contrasts node embeddings from one view with graph embeddings of another view and vice versa. Following up, MCGC [Pan and Kang, 2021] and HeCo [Wang et al., 2021] extend the contrastive paradigm to multi-view clustering and heterogeneous graph learning.

Although the effectiveness of the contrastive learning paradigm has been verified, there are still many open technical problems. Specifically, [Zhao et al., 2021] proposes GDCL to correct the bias of sampling in contrastive deep graph clustering. Besides, to avoid the semantic drift caused by inappropriate data augmentations, AFGRL [Lee et al., 2021] is proposed by replacing data augmentations with node discovery. Then, to refine the noisy connections in the graph, [Liu et al., 2022a] propose SUBLIME by generating the graph structural information by the random surfing model.
sketched graph view with unsupervised structure learning. Moreover, [Liu et al., 2022b; Liu et al., 2022d] design the dual correlation reduction strategy in the DCRN model to alleviate the representation collapse problem in deep graph clustering.

3 Taxonomy

In this section, we introduce the taxonomy of deep graph clustering methods from the following perspectives.

3.1 Graph Type

As shown in Figure 5, the input graphs of the existing deep graph clustering methods are categorized into four types: pure structure graph, attribute graph, heterogeneous graph, and dynamic graph. The definitions of these graph types are formulated as follows.

**Definition 1. Pure Structure Graph.** In a pure structure graph $G_S = \{V, E\}$, $V = \{v_1, v_2, \ldots, v_N\}$ is the set of $N$ nodes with $K$ classes, and $E$ is a set of $M$ edges. With the matrix form, the pure structure graph can be represented by $A = R^{N \times N}$. Here, $A_{ij} = 1$ if $i$-th node is linked to $j$-th node, and $A_{ij} = 0$ if $i$-th node is not linked to $j$-th node.

**Definition 2. Attribute Graph.** Compared with the pure structure graph $G_S$, the attribute graph $G_A = \{V, E, X\}$ attaches the attributes of nodes. $X \in R^{N \times D}$ is denoted as the attribute matrix, where $D$ is the dimension number of node attributes. With matrix form, $G_A$ is represented by the adjacency matrix $A \in R^{N \times N}$ and attribute matrix $X$.

**Definition 3. Heterogeneous Graph.** In a graph, we firstly denote the type number of nodes and edges as $T_n$ and $T_e$. A heterogeneous graph $G_H$ satisfies $T_n + T_e > 2$, i.e., it contains multiple types of nodes or/and multiple types of edges. Otherwise, the graph is a homogeneous graph.

**Definition 4. Dynamic Graph.** Different from the static graph, the dynamic graph $G_D = \{G_D^{(1)}, \ldots, G_D^{(t)}, \ldots, G_D^{(T)}\}$ will dynamically change over time, where $t$ is the time step.

The first type of graph, the pure structure graph is relatively easy to process since it merely contains structural information. For example, in early works, [Tian et al., 2014] encodes the structural embeddings with the sparse auto-encoders [Ng and others, 2011]. Besides, [Cao et al., 2016] apply random surfing in the graph and embed the graph structure with the stacked denoising auto-encoder. For attribute graph, the additional attribute information always brings more process operations and performance improvement.

**MLP-based Methods.** The MLP-based methods utilize the multi-layer perceptrons (MLPs) [Pal and Mitra, 1992] to extract the informative features in the graphs. For example, GraphEncoder [Tian et al., 2014] and DNGR [Cao et al., 2016] encode the graph structure with the auto-encoders. Subsequently, in ProGAN [Gao et al., 2019] and CommunityGAN [Jia et al., 2019], the authors adopt MLPs to design the generative adversarial networks. Moreover, based on MLPs, AGE [Cui et al., 2020] and SCGC [Liu et al., 2022c] design the adaptive encoder and the parameter un-shared encoders to embed the smoothed node features into the latent space. Although the effectiveness of these methods has been demonstrated, they are inefficient for MLPs to capture non-euclidean structural information in graphs. Thus, GNN-based methods are increasingly proposed in recent years.

**GNN-based Methods.** The GNN-based methods model the non-euclidean graph data with the GNN encoders like graph convolutional network [Kipf and Welling, 2017], graph attention network [Veličković et al., 2017], graph auto-encoder [Kipf and Welling, 2016], etc. Benefiting from the strong graph structure representation capability, GNN-based methods have achieved promising performance. For instance, MGAE [Wang et al., 2017] is proposed to learn the node attribute and graph structure with the designed graph auto-encoder. In addition, [Park et al., 2019] design a novel...
Table 2: Taxonomy of deep graph clustering methods. The criteria contain graph type, network architecture, learning paradigm, and clustering method.

| Methods                  | Graph Type       | Network Architecture | Learning Paradigm       | Clustering Method     |
|--------------------------|------------------|----------------------|-------------------------|-----------------------|
| GraphEncoder [Yuan et al., 2014] | Pure Structure Graph | MLP                 | Adversarial + Contrastive | Traditional Clustering |
| DGC [Cao et al., 2016]   | Pure Structure Graph | MLP                 | Adversarial + Contrastive | Traditional Clustering |
| CommunityGAN [Jia et al., 2019] | Pure Structure Graph | MLP                 | Adversarial + Contrastive | Traditional Clustering |
| NetVAE [Jin et al., 2019] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| ProGAN [Gao et al., 2019] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| AGE [Cui et al., 2020]   | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| GALA [Park et al., 2019] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| MGA [Wang et al., 2017]  | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| CGCN [Hui et al., 2020]  | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| GCLN [Hu et al., 2020a]  | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| AHGAE [Hu et al., 2021]  | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| SCAE [Zhao et al., 2022a] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| DAEGC [Wang et al., 2019a] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| DGVAE [Li et al., 2020]  | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| CDSS [Zhu et al., 2022]  | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| GGC [Zhong et al., 2021] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| JANE [Yang et al., 2020] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| SAIL [Yu et al., 2022a]  | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| AGF [Lee et al., 2021]   | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| SGC [Devvitt et al., 2022] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| SCAGC [Xia et al., 2022e] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| CommDGI [Zhang et al., 2020b] | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| AGAE [Tao et al., 2019]  | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| ARL [Pan et al., 2019]   | Attribute Graph  | MLP                 | Contrastive             | Traditional Clustering |
| SDGC [Bo et al., 2020]   | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| DFCN [Tu et al., 2020]   | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| AGCN [Peng et al., 2021] | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| R-GAE [Mrabah et al., 2021] | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| MVGRL [Hassani and Khasahmadi, 2020] | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| SUBLME [Li et al., 2022a] | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| GDC [Zhao et al., 2021]  | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| DCRN [Li et al., 2022b]  | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| AGC-DRR [Gong et al., 2022] | Attribute Graph  | MLP+GNN             | Reconstructive + Contrastive | Traditional Clustering |
| DMGC [Luo et al., 2020]  | Heterogeneous Graph | MLP                 | Reconstructive + Contrastive | Traditional Clustering |
| MAGCN [Cheng et al., 2021] | Heterogeneous Graph | MLP                 | Reconstructive + Contrastive | Traditional Clustering |
| O2MAC [Fan et al., 2020] | Heterogeneous Graph | MLP                 | Reconstructive + Contrastive | Traditional Clustering |
| HeCo [Wang et al., 2021] | Heterogeneous Graph | MLP                 | Reconstructive + Contrastive | Traditional Clustering |
| SGCMC [Xia et al., 2022f] | Heterogeneous Graph | MLP                 | Reconstructive + Contrastive | Traditional Clustering |
| VaCA-HINE [Khan and Kleinsteuber, 2022] | Heterogeneous Graph | MLP                 | Reconstructive + Contrastive | Traditional Clustering |
| HNE [Chang et al., 2015] | Heterogeneous Graph | MLP                 | Reconstructive + Contrastive | Traditional Clustering |
| VQGMM [Li et al., 2022]  | Dynamic Graph    | MLP                 | Reconstructive + Contrastive | Traditional Clustering |
| CGC [Park et al., 2022]  | Dynamic Graph    | MLP                 | Reconstructive + Contrastive | Traditional Clustering |

Gong et al., 2022] also adopt the hybrid architecture of MLPs and GNNs as the backbones.

3.3 Learning paradigm

Based on the learning paradigm, the surveyed methods contain four categories: reconstructive methods, adversarial methods, contrastive methods, and hybrid methods. Take the attribute graph input for example, we elaborate on these learning paradigms of deep graph clustering methods as follows.

Reconstructive Methods. The core idea of the reconstructive methods is forcing the network to encode the features in the graph and then recovering (part of) the graph information from the learned embeddings. Thus, the reconstructive methods focus on the intra-data information in the graph and adopt the node attributes and graph structure as the self-supervision signals. The general pipeline of reconstructive deep graph clustering methods is illustrated in Figure 6.

Adversarial Methods. The adversarial deep graph clustering methods aim to improve the quality of features by creating a zero-sum game between the generator and the discrimina-
Figure 6: The general pipeline of the reconstructive deep graph clustering methods. Firstly, the nodes of the input graph are encoded into the node embeddings $Z$ via a designed Encoder. Subsequently, with the reconstruction pre-text tasks like attribute reconstruction or link reconstruction, the decoder aims to recover the graph information from the learned embeddings $Z$. Eventually, in the latent space, the clustering method $C$ groups the nodes into different clusters.

To be specific, the discriminator is trained to recognize whether learned features are from the real data distribution while the generator aims to generate confusing embeddings to cheat the discriminator. In these settings, the generation operation and the discrimination operation provide effective self-supervision signals, guiding the network to learn high-quality embeddings. Figure 7 demonstrates the general pipeline of the adversarial deep graph clustering methods.

Figure 7: The general pipeline of adversarial deep graph clustering. Firstly, the Generator aims to generate high-quality node embeddings $Z$ from the input graph. Subsequently, the Discriminator is trained to distinguish the fake information and the learned features. After that, the clustering method $C$ separates the learned node embeddings into several clusters.

Contrastive Methods. The key idea in contrastive deep graph clustering methods is to improve the discriminativeness of features by pulling together the positive samples while pushing away the negative ones. Thus, the contrastive methods focus on the inter-data information and construct the self-supervision signals via the meaningful relationships between samples. The general pipeline of contrastive methods can be found in Figure 8.

Hybrid Methods. In recent years, some papers demonstrate the effectiveness of the combination of different learning paradigms. For example, in ARGA [Pan et al., 2019], Pan et al. adopt both the reconstructive and adversarial learning paradigms. Besides, researchers [Xia et al., 2022f; Liu et al., 2022b] also verified the effectiveness of the combination of reconstructive and contrastive learning paradigms. Moreover, in AGC-DRR [Gong et al., 2022], the researchers show that an adversarial mechanism is a new option for data augmentation in contrastive learning.

3.4 Clustering Method

The clustering methods in deep graph clustering aim to separate the learned node embeddings into different clusters in an unsupervised manner. They can be categorized into two classes: traditional clustering and neural clustering.

Traditional Clustering. The traditional clustering methods [Hartigan and Wong, 1979; Von Luxburg, 2007; Reynolds, 2009; Zhang et al., 2022b] can be directly performed on the learned node embeddings to group them into disjoint clusters in many early deep graph clustering methods [Tian et al., 2014; Cao et al., 2016; Wang et al., 2017; Pan et al., 2019; Park et al., 2019; Tao et al., 2019; Hasnani and Khasahmadi, 2020; Cui et al., 2020]. Although they achieve promising performance in an efficient manner, the traditional clustering methods have two drawbacks as follows: 1) Firstly, the clustering process can not benefit from the strong representation capability of deep neural networks, thus limiting the discriminative capability of the samples. 2) Besides, in these methods, the node representation learning and the clustering learning cannot be jointly optimized in a unified framework (end-to-end manner?).

Neural Clustering. To alleviate the aforementioned issues of traditional clustering methods, the neural clustering methods are designed to group the samples into different clusters with deep neural networks. Concretely, in the neural clustering methods, the clustering process and the deep neural network are jointly optimized by the gradient descent algorithms [Kingma and Ba, 2014; Bottou, 2010].
Figure 9: The clustering distribution alignment loss $\mathcal{L}_{KL}$ in the neural clustering methods.

For example, in many two-stage neural clustering methods [Wang et al., 2019a; Cheng et al., 2021; Fan et al., 2020; Bo et al., 2020; Peng et al., 2021; Zhao et al., 2021; Tu et al., 2020; Liu et al., 2022b; Zhu et al., 2022], a widely-used clustering distribution alignment loss is introduced to network optimization as shown in Figure 9. Specifically, in the first stage, these methods pre-train the encoders $\mathcal{F}$ and initialize the cluster center embeddings $\mathbf{C} \in \mathbb{R}^{K \times d}$ by the traditional clustering algorithms on the learned node embeddings $\mathbf{Z} \in \mathbb{R}^{N \times d}$. Here, the cluster center embeddings $\mathbf{C}$ are set to the learnable parameters of the deep neural networks. Then, in the second stage, a soft assignment between the node embeddings $\mathbf{Z}$ and the cluster center embeddings $\mathbf{C}$ is calculated as formulated:

$$
Q_{ij} = \frac{(1 + ||Z_i - C_j||^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j'}(1 + ||Z_i - C_{j'}||^2/\alpha)^{-\frac{\alpha+1}{2}}}.
$$

(3)

Here, the similarity between $Z_i$ and $C_j$ is measured by the Student’s $t$-distribution kernel. Besides, $\alpha$ is the freedom degree of Student’s $t$-distribution. $Q_{ij}$ can be considered as the probability of assigning $i$-th node to $j$-th cluster, namely a soft assignment. Subsequently, the clustering is promoted by aligning the soft assignment $Q_{ij}$ with the high confidence assignments $P_{ij}$ as follows.

$$\mathcal{L}_{KL} = KL(P||Q) = \sum_i \sum_j P_{ij} \log \frac{P_{ij}}{Q_{ij}},$$

(4)

where the high confidence (target) assignments $P_{ij}$ can be calculated by first raising $Q_{ij}$ to the second pow and then normalizing by the frequency per cluster as formulated:

$$P_{ij} = \frac{Q_{ij}^2 / \sum_i Q_{ij}}{\sum_{j'} Q_{ij'} / \sum_i Q_{ij'}}.$$  

(5)

In this manner, the clusters are iteratively refined by learning from the sharpened clustering distribution, thus improving clustering performance.

Similarly, in other two-stage neural clustering methods [Zhao et al., 2021; Xia et al., 2022a; Xia et al., 2022c; Zhu et al., 2022; Liu et al., 2022d], the high confidence clustering pseudo labels are adopted as the supervision signal for better clustering as shown in Figure 10. For instance, in GDCL, [Zhao et al., 2021] debias the false negative sampling of contrastive learning with the pseudo clustering labels. Besides, [Xia et al., 2022c] proposes SCAGC to maximize the similarities of intra-cluster nodes while minimizing the similarities of inter-cluster nodes. Moreover, in IDCNR, [Liu et al., 2022d] utilizes the high confidence pseudo labels to refine the adjacency matrix and guide the learned embeddings to recover the affinity matrix even across views. In addition, [Xia et al., 2022f] design a cross-entropy based objective function to guide the network to classify the clustering pseudo labels in the SGCMC model. Furthermore, [Zhu et al., 2022] collaborates the pseudo node classification task with clustering and augments a pseudo-label set to further improve the clustering performance.

However, the promising performance of the two-stage neural clustering methods is heavily dependent on the good initialization of cluster centers, leading to the manual trial-and-error network pre-training and the auxiliary of traditional clustering methods. To overcome these shortcomings, some one-stage methods [Liu et al., 2022c; Gong et al., 2022; Zhong et al., 2021] are gradually proposed recently. For example, [Liu et al., 2022c] proposes a new method termed SCGC with the Laplacian smoothing and neighborhood contrastive learning. Besides, in AGC-DRR [Gong et al., 2022], the clustering sub-network is developed to directly predict the probabilities of cluster-ID for each sample. Moreover, rather than dealing with clustering as a downstream task, [Zhong et al., 2021] proposes a unified framework termed GCC by minimizing the difference between node embeddings and reconstructed cluster embeddings. The one-stage neural clustering will be an interesting future direction. In the next section, the difficult challenges will be summarized and analyzed in detail.

4 Challenge & Opportunity

In this section, we first summarize and analyze the urge challenges in deep graph clustering. Subsequently, the corresponding potential solutions and opportunities are discussed.

4.1 Clustering with Low Quality Graph Data

The real-world graph data are always with low quality. Firstly, the graph data may contain many noises on both the node attributes and the graph structures, e.g., the wrong node attributes and error connections between nodes. Secondly, there might be lots of missing information in the graph data, e.g., missing part node attributes, missing connections, or even missing whole node attributes.
For the noise problem, some denoising technologies [Chen and Eldar, 2021; Do et al., 2020; Xia et al., 2020] can be the potential solutions. Besides, the imputation networks [Tu et al., 2022; Chen et al., 2020b] can help alleviate the information missing issue.

### 4.2 Stability

The stability of graph learning algorithms is very important, especially in some sensitive domains, such as financial risk control, anomaly detection in social networks, etc. However, the stability of deep graph clustering models is weaker than other supervised models. The reason is that these methods are purely unsupervised and easily bring more randomness to the optimization process. The randomness mainly consists of two parts as follows. 1) The initialization and training of the encoders. 2) The initialization and optimization of the clustering methods. Making the deep graph clustering methods more robust will be a promising direction.

Thus, the existing works always fix the random seed and run algorithms ten times to alleviate the influence of randomness. Some optimization strategies [Zheng et al., 2016; Rozsa et al., 2016] might enhance the stability of deep graph clustering.

### 4.3 Scalability

Although achieving promising performance, most previous deep graph clustering methods cannot scale to the large-scale graph datasets. The heavy time and memory costs mainly come from two parts as follows. 1) Train the encoding network on the large graph. 2) Group the large number of nodes in the graph into separate clusters.

For the first part, batch training [Li et al., 2014] will be a good solution. Concretely, researchers can partition the large graph into a mini-batch graph via sub-graph extraction technology like random walk [Lovász, 1993] and then train the networks on the sub-graphs. Here, different from images or texts, one most important characteristics of graph data are the relational connections between nodes. Thus, keeping the original connection information after the graph partition is important for the model’s performance. For the second part, for the learned node embeddings $Z \in \mathbb{R}^{N \times d}$, the classical $K$-Means clustering algorithm will take $O(tKNd)$ time complexity and $O(Nd + Kd)$ space complexity. Here, $N$, $K$, and $d$ are the node number, cluster number, and feature dimension number, respectively. Besides, $t$ is the iteration number of $K$-Means. On the large graph, the node number $N$ is very large, such as 111M node number on ogbn-papers100M dataset [Hu et al., 2020b], easily leading to out-of-memory on GPU or the long calculation time on CPU. The potential solutions include the mini-batch $K$-Means [Newling and Fleuret, 2016], finding cluster centers via the neural network and calculating the clustering assignments in a mini-batch manner, bi-part graph clustering [Wang et al., 2022a; Xia et al., 2022d], etc.

### 4.4 Discriminative Capability

The indiscriminative samples in the graph easily confuse the network and lead to a representation collapse problem, i.e., the network tends to embed all sample into the same cluster. To demonstrate this problem, we visualize the similarity matrix of latent sample features as shown in Figure 11. From these results, we find that the discriminative capability of SDCN [Bo et al., 2020] is relatively weak compared with IDCRN [Liu et al., 2022d], leading to worse clustering performance. Thus, the sample discriminative capability of the network is one key factor to achieve promising clustering performance.

![Figure 11: The visualization of sample similarity matrix on ACM dataset.](image)

Recently, more and more methods [Liu et al., 2022b; Gong et al., 2022; Xia et al., 2022e] utilize the contrastive learning paradigm to pull together the positive samples while pushing away the negative ones, thus improving the sample discriminative capability of the network. To further enhance the discriminative capability, hard sample mining [Xia et al., 2022b] will be a possible future direction.

### 4.5 Unknown Cluster Number

The remarkable success of recent deep graph clustering algorithms relies on the pre-defined cluster number. To verify this, experiments are conducted in Figure 12. Concretely, we adopt DCRN [Liu et al., 2022b] as the baseline and test its clustering performance with different cluster numbers on DBLP and CiteSeer datasets. From the experimental results, we observe that the wrong cluster number will lead to performance degeneration. Thus, designing the deep graph clustering methods without the pre-defined cluster number will be an open question in the future.

![Figure 12: Clustering performance of DCRN with different cluster number $K \in [2, 10]$ on two datasets.](image)
cretely, the clustering network should be designed to assign the high-density data area as the cluster centers and regard the low-density data area as the decision borders. Besides, another potential solution technology is deep reinforcement learning [Arulkumaran et al., 2017]. To be specific, in an unsupervised scenario, recognizing the cluster number in the graph data can be modeled as the Markov decision process and be handled with the deep reinforcement neural networks. In the next section, we will introduce the applications of deep graph clustering in four domains.

5 Application
Thanks to the strong data partitioning capability, deep graph clustering has been applied to different domains of real-world applications, such as social network analysis, computer vision, natural language processing, bioinformatics, etc. As shown in Figure 13, the specific applications of deep graph clustering contain face analysis [Wang et al., 2019b; Wang et al., 2022b], anomaly detection [Ahmed et al., 2021; Ma et al., 2021], co-saliency detection [Zhang et al., 2020a], document mining [Chiu et al., 2020], speech separation [Qin et al., 2020], community detection [Cavallari et al., 2017; Rozemberczki et al., 2019; Tu et al., 2018; Liu et al., 2020], molecular mining [Grunig et al., 2022; Xia et al., 2022c], metagenomic binning [Xue et al., 2022], single-cell RNA sequencing [Yu et al., 2022b], and etc.

Figure 13: Applications of deep graph clustering.

6 Conclusion
In this work, we make a comprehensive survey about deep graph clustering. Firstly, we introduce the detailed definition and applications of deep graph clustering. Besides, this paper summarizes the important baselines and the taxonomy of deep graph clustering methods based on four criteria. Moreover, we summarize and analyze the challenges, and also provide some potential technical solutions for researchers. Besides, the collection of papers, codes, and datasets of the survey methods is shared on the GitHub platform. We hope this work will be a quick guide for the researchers and motivate them to solve important problems in the deep graph clustering domain.

References
[Ahmed et al., 2021] Imtiaz Ahmed, Travis Galoppo, Xia Hu, and Yu Ding. Graph regularized autoencoder and its application in unsupervised anomaly detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.
[Aljalbout et al., 2018] Elie Aljalbout, Vladimir Golkov, Yawar Siddiqui, Maximilian Strobel, and Daniel Cremers. Clustering with deep learning: Taxonomy and new methods. arXiv preprint arXiv:1801.07648, 2018.
[Arulkumaran et al., 2017] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6):26–38, 2017.
[Bianchi et al., 2020] Filippo Maria Bianchi, Daniele Gratton, and Cesare Alippi. Spectral clustering with graph neural networks for graph pooling. In Proc. of ICML, 2020.
[Bo et al., 2020] Deyu Bo, Xiao Wang, Chuan Shi, Meiqi Zhu, Emiao Lu, and Peng Cui. Structural deep clustering network. In Proc. of WWW, 2020.
[Bottou, 2010] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In Proceedings of COMPSTAT’2010, pages 177–186. Springer, 2010.
[Cao et al., 2016] Shaosheng Cao, Wei Lu, and Qiongkai Xu. Deep neural networks for learning graph representations. In Proc. of AAAI, 2016.
[Cavallari et al., 2017] Sandro Cavallari, Vincent W Zheng, Hongyun Cai, Kevin Chen-Chuan Chang, and Erik Cambria. Learning community embedding with community detection and node embedding on graphs. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017.
[Chang et al., 2015] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. Heterogeneous network embedding via deep architectures. In Proc. of KDD, 2015.
[Chen and Eldar, 2021] Siheng Chen and Yonina C Eldar. Graph signal denoising via unrolling networks. In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5290–5294. IEEE, 2021.
[Chen et al., 2020a] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Proc. of ICML, 2020.
[Chen et al., 2020b] Xu Chen, Siheng Chen, Jiangchao Yao, Huangjie Zheng, Ya Zhang, and Ivor W Tsang. Learning on attribute-missing graphs. IEEE transactions on pattern analysis and machine intelligence, 2020.
[Cheng et al., 2021] Jiafeng Cheng, Qianqian Wang, Zhiqiang Tao, Deyan Xie, and Quanxue Gao. Multi-view attribute graph convolution networks for clustering. In


Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, 2021.

[Chiu et al., 2020] Billy Chiu, Sunil Kumar Sahu, Derek Thomas, Neha Sengupta, and Mohammady Mahdy. Autoencoding keyword correlation graph for document clustering. In Proc. of ACL, 2020.

[Chung et al., 2014] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555, 2014.

[Cui et al., 2020] Ganqu Cui, Jie Zhou, Cheng Yang, and Zhiyuan Liu. Adaptive graph encoder for attributed graph embedding. In Proc. of KDD, 2020.

[Devvrit et al., 2022] Fnu Devvrit, Aditya Sinha, Inderjit Dhillon, and Prateek Jain. S3gc: Scalable self-supervised graph clustering. 2022.

[Do et al., 2020] Tien Huu Do, Duc Minh Nguyen, and Nikos Deligiannis. Graph auto-encoder for graph signal denoising. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 3322–3326. IEEE, 2020.

[Fan et al., 2020] Shaohua Fan, Xiao Wang, Chuan Shi, Emiao Lu, Ken Lin, and Bai Wang. One2multi graph autoencoder for multi-view graph clustering. In Proc. of WWW, 2020.

[Gao et al., 2019] Hongchang Gao, Jian Pei, and Heng Huang. Progan: Graph clustering via proximity generative adversarial network. In Proc. of KDD, 2019.

[Gong et al., 2022] Lei Gong, Sihang Zhou, Xinwang Liu, and Wenxuan Tu. Attributed graph clustering with dual redundancy reduction. In Proc. of IJCAI, 2022.

[Goodfellow et al., 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Proc. of NeurIPS, 2014.

[Grill et al., 2020] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhao-han Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. arXiv preprint arXiv:2006.07733, 2020.

[Grunig et al., 2022] Gabriele Grunig, Nedim Durmus, Yian Zhang, Yuting Lu, Sultan Pehlivan, Yuyan Wang, Kathleen Doo, Maria L Cotrina-Vidal, Roberta Goldring, Kenneth I Berger, et al. Molecular clustering analysis of blood biomarkers in world trade center exposed community members with persistent lower respiratory symptoms. International journal of environmental research and public health, 2022.

[Guo et al., 2017] Xifeng Guo, Long Gao, Xinwang Liu, and Jianping Yin. Improved deep embedded clustering with local structure preservation. In Proc. of IJCAI, 2017.

[Hartigan and Wong, 1979] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. Journal of the royal statistical society. series c (applied statistics), 1979.

[Hassani and Khasahmadi, 2020] Kaveh Hassani and Amir Hosein Khasahmadi. Contrastive multi-view representation learning on graphs. In Proc. of ICML, 2020.

[He et al., 2020] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In Proc. of CVPR, 2020.

[Hjelm et al., 2018] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. In Proc. of ICLR, 2018.

[Hu et al., 2020a] Ruiqi Hu, Shirui Pan, Guodong Long, Qinghua Lu, Liming Zhu, and Jing Jiang. Going deep: Graph convolutional ladder-shape networks. In Proc. of AAAI, 2020.

[Hu et al., 2020b] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Castasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. Advances in neural information processing systems, 33:22118–22133, 2020.

[Hu et al., 2021] Youpeng Hu, Xunkai Li, Yujie Wang, Yixuan Wu, Yining Zhao, Chenggang Yan, Jian Yin, and Yue Gao. Adaptive hypergraph auto-encoder for relational data clustering. IEEE Transactions on Knowledge and Data Engineering, 2021.

[Hui et al., 2020] Binyuan Hui, Pengfei Zhu, and Qinghua Hu. Collaborative graph convolutional networks: Unsupervised learning meets semi-supervised learning. In Proc. of AAAI, 2020.

[Jia et al., 2019] Yuting Jia, Qinqin Zhang, Weinan Zhang, and Xinbing Wang. Communitygan: Community detection with generative adversarial nets. In Proc. of WWW, 2019.

[Jin et al., 2019] Di Jin, Bingyi Li, Pengfei Jiao, Dongxiao He, and Weixiong Zhang. Network-autoencoder for embedding in attribute networks. In Proc. of IJCAI, 2019.

[Khan and Kleinsteuber, 2022] Rayyan Ahmad Khan and Martin Kleinsteuber. Cluster-aware heterogeneous information network embedding. In Proc. of WSDM, 2022.

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[Kipf and Welling, 2016] Thomas N Kipf and Max Welling. Variational graph auto-encoders. arXiv preprint arXiv:1611.07308, 2016.

[Kipf and Welling, 2017] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In Proc. of ICLR, 2017.
[Lee et al., 2019] Junhyun Lee, Inyeop Lee, and Jaewoo Kang. Self-attention graph pooling. In Proc. of ICML, 2019.

[Lee et al., 2021] Namkyeong Lee, Junseok Lee, and Chanyoung Park. Augmentation-free self-supervised learning on graphs. arXiv preprint arXiv:2112.02472, 2021.

[Li et al., 2014] Mu Li, Tong Zhang, Yuqiang Chen, and Alexander J Smola. Efficient mini-batch training for stochastic optimization. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 661–670, 2014.

[Li et al., 2020] Jia Li, Jianwei Yu, Jiajin Li, Honglei Zhang, Kangfei Zhao, Yu Rong, Hong Cheng, and Junzhou Huang. Dirichlet graph variational autoencoder. Proc. of NeurIPS, 2020.

[Li et al., 2022] Tianpeng Li, Wenjun Wang, Pengfei Jiao, Yinhui Wang, Ruomeng Ding, Huaming Wu, Lin Pan, and Di Jin. Exploring temporal community structure via network embedding. IEEE Transactions on Cybernetics, 2022.

[Liang et al., 2022] Ke Liang, Yue Liu, Sihang Zhou, Xinwang Liu, and Wenxuan Tu. Relational symmetry based knowledge graph contrastive learning, 2022.

[Liu et al., 2020] Fanzhen Liu, Shan Xue, Jia Wu, Chuan Zhou, Wenbin Hu, Cecile Paris, Surya Nepal, Jian Yang, and Philip S Yu. Deep learning for community detection: progress, challenges and opportunities. arXiv preprint arXiv:2005.08225, 2020.

[Liu et al., 2022a] Yixin Liu, Yu Zheng, Daokun Zhang, Hongxu Chen, Hao Peng, and Shirui Pan. Towards unsupervised deep graph structure learning. In Proceedings of the ACM Web Conference 2022, 2022.

[Liu et al., 2022b] Yue Liu, Wenxuan Tu, Sihang Zhou, Xinwang Liu, Linxuan Song, Xihong Yang, and En Zhu. Deep graph clustering via dual correlation reduction. In Proc. of AAAI, 2022.

[Liu et al., 2022c] Yue Liu, Xihong Yang, Sihang Zhou, and Xinwang Liu. Simple contrastive graph clustering. arXiv preprint arXiv:2205.07865, 2022.

[Liu et al., 2022d] Yue Liu, Sihang Zhou, Xinwang Liu, Wenxuan Tu, and Xihong Yang. Improved dual correlation reduction network. arXiv preprint arXiv:2202.12533, 2022.

[Lovász, 1993] László Lovász. Random walks on graphs. Combinatorics, Paul erdos is eighty, 2(1-46):4, 1993.

[Luo et al., 2020] Dongsheng Luo, Jingchao Ni, Suhang Wang, Yuchen Bian, Xiong Yu, and Xiang Zhang. Deep multi-graph clustering via attentive cross-graph association. In Proceedings of the 13th international conference on web search and data mining, 2020.

[Ma et al., 2021] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z Sheng, Hui Xiong, and Leman Akoglu. A comprehensive survey on graph anomaly detection with deep learning. IEEE Transactions on Knowledge and Data Engineering, 2021.

[Min et al., 2018] Erxue Min, Xifeng Guo, Qiang Liu, Gen Zhang, Jianjng Cui, and Jun Long. A survey of clustering with deep learning: From the perspective of network architecture. IEEE Access, 6:39501–39514, 2018.

[Mrabah et al., 2021] Nairouz Mrabah, Mohmed Bouguessa, Mohamed Fawzi Touati, and Riadh Ksantini. Rethinking graph auto-encoder models for attributed graph clustering. arXiv preprint arXiv:2107.08562, 2021.

[Newling and Fleuret, 2016] James Newling and François Fleuret. Nested mini-batch k-means. Advances in neural information processing systems, 29, 2016.

[Ng and others, 2011] Andrew Ng et al. Sparse autoencoder. CS294A Lecture notes, 72(2011):1–19, 2011.

[Pal and Mitra, 1992] Sankar K Pal and Sushmita Mitra. Multilayer perceptron, fuzzy sets, classification. 1992.

[Pan and Kang, 2021] Erlin Pan and Zhao Kang. Multi-view contrastive graph clustering. Proc. of NeurIPS, 2021.

[Pan et al., 2018] Shirui Pan, Ruiqi Hu, Guodong Long, Jing Jiang, Lina Yao, and Chengqi Zhang. Adversarially regularized graph autoencoder for graph embedding. In Proc. of IJCAI, 2018.

[Pan et al., 2019] Shirui Pan, Ruiqi Hu, Sai-fu Fung, Guodong Long, Jing Jiang, and Chengqi Zhang. Learning graph embedding with adversarial training methods. IEEE transactions on cybernetics, 2019.

[Park et al., 2019] Jiwoong Park, Minsik Lee, Hyung Jin Chang, Kyuweung Lee, and Jin Young Choi. Symmetric graph convolutional autoencoder for unsupervised graph representation learning. In Proc. of ICCV, 2019.

[Park et al., 2022] Namkyeong Park, Ryan Rossi, Eunyee Koh, Iftikhar Ahamath Burhanuddin, Sungchul Kim, Fan Du, Nesreen Ahmad, and Christos Faloutsos. Cgc: Contrastive graph clustering forcummmunity detection and tracking. In Proceedings of the ACM Web Conference 2022, 2022.

[Peng et al., 2021] Zhihao Peng, Hui Liu, Yuheng Jia, and Junhui Hou. Attention-driven graph clustering network. In Proc. of ACM MM, 2021.

[Qin et al., 2020] Shan Qin, Ting Jiang, Sheng Wu, Ning Wang, and Xinran Zhao. Graph convolution-based deep clustering for speech separation. IEEE Access, 2020.

[Ren et al., 2022] Yazhou Ren, Jingyu Pu, Zhimeng Yang, Jie Xu, Guofeng Li, Xiaorong Pu, Philip S Yu, and Lifang He. Deep clustering: A fast search and find of density peaks. science, 344(6191):1492–1496, 2014.

[Rozemberczki et al., 2019] Benedek Rozemberczki, Ryan Davies, Rik Sarkar, and Charles Sutton. Gemsec: Graph embedding with self clustering. In Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, 2019.
[Rozsa et al., 2016] Andras Rozsa, Manuel Gunther, and Terrance E Boult. Towards robust deep neural networks with bang. arXiv preprint arXiv:1612.00138, 2016.

[Tao et al., 2019] Zhiqiang Tao, Hongfu Liu, Jun Li, Zhaowen Wang, and Yun Fu. Adversarial graph embedding for ensemble clustering. In International Joint Conferences on Artificial Intelligence Organization, 2019.

[Thakoor et al., 2021] Shantanu Thakoor, Corentin Tallec, Mohammad Gheshlaghi Azar, Rémi Munos, Petar Veličković, and Michal Valko. Bootstrapped representation learning on graphs. In ICLR 2021 Workshop on Geometrical and Topological Representation Learning, 2021.

[Tian et al., 2014] Fei Tian, Bin Gao, Qing Cui, Enhong Chen, and Tie-Yan Liu. Learning deep representations for graph clustering. In Proc. of AAAI, 2014.

[Tsitsulin et al., 2020] Anton Tsitsulin, John Palowitch, Bryan Perozzi, and Emmanuel Müller. Graph clustering with graph neural networks. arXiv preprint arXiv:2006.16904, 2020.

[Tu et al., 2018] Cunchao Tu, Xiangkai Zeng, Hao Wang, Zhengyan Zhang, Zhiyuan Liu, Maosong Sun, Bo Zhang, and Leyu Lin. A unified framework for community detection and network representation learning. IEEE Transactions on Knowledge and Data Engineering, 2018.

[Tu et al., 2020] Wenxuan Tu, Sihang Zhou, Xinwang Liu, Xifeng Guo, Zhiping Cai, Jieren Cheng, et al. Deep fusion clustering network. arXiv preprint arXiv:2012.09600, 2020.

[Tu et al., 2022] Wenxuan Tu, Sihang Zhou, Xinwang Liu, Yue Liu, Zhiping Cai, En Zhu, Zhang Changwang, and Jieren Cheng. Initializing then refining: A simple graph attribute imputation network. IJCAI, pages 3494–3500, 2022.

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

[Velickovic et al., 2019] Petar Velickovic, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. ICLR (Poster), 2019.

[Von Luxburg, 2007] Ulrike Von Luxburg. A tutorial on spectral clustering. Statistics and computing, 2007.

[Wang et al., 2017] Chun Wang, Shirui Pan, Guodong Long, Xingquan Zhu, and Jing Jiang. Mgae: Marginalized graph autoencoder for graph clustering. In Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017.

[Wang et al., 2019a] Chun Wang, Shirui Pan, Ruiqi Hu, Guodong Long, Jing Jiang, and Chengqi Zhang. Attributed graph clustering: A deep attentional embedding approach. arXiv preprint arXiv:1906.06532, 2019.

[Wang et al., 2019b] Zhongdao Wang, Liang Zheng, Yali Li, and Shengjin Wang. Linkage based face clustering via graph convolution network. In Proc. of CVPR, 2019.

[Wang et al., 2021] Xiao Wang, Nian Liu, Hui Han, and Chuan Shi. Self-supervised heterogeneous graph neural network with co-contrastive learning. In Proc. of KDD, 2021.

[Wang et al., 2022a] Siwei Wang, Xinwang Liu, Li Liu, Wenxuan Tu, Xinzhong Zhu, Jiyuan Liu, Sihang Zhou, and En Zhu. Highly-efficient incomplete large-scale multi-view clustering with consensus bipartite graph. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 9776–9785, 2022.

[Wang et al., 2022b] Yaohua Wang, Yaobin Zhang, Fangyi Zhang, Zhenzhang Wang, Ming Lin, YuQi Zhang, and Xi- uyu Sun. Ada-nets: Face clustering via adaptive neighbour discovery in the structure space. arXiv preprint arXiv:2202.03800, 2022.

[Xia et al., 2020] Jun Xia, Haitao Lin, Yongjie Xu, Lirong Wu, Zhangyang Gao, Siyuan Li, and Stan Z Li. Towards robust graph neural networks against label noise. 2020.

[Xia et al., 2022a] Jun Xia, Lirong Wu, Jintao Chen, Bozhen Hu, and Stan Z Li. Simgrace: A simple framework for graph contrastive learning without data augmentation. arXiv preprint arXiv:2202.03104, 2022.

[Xia et al., 2022b] Jun Xia, Lirong Wu, Ge Wang, Jintao Chen, and Stan Z Li. Progcl: Rethinking hard negative mining in graph contrastive learning. In Proc. of ICML, 2022.

[Xia et al., 2022c] Jun Xia, Yanqiao Zhu, Yuanqi Du, Yue Liu, and Stan Z Li. A systematic survey of molecular pre-trained models. arXiv preprint arXiv:2210.16484, 2022.

[Xia et al., 2022d] Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. Tensorized bipartite graph learning for multi-view clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.

[Xia et al., 2022e] Wei Xia, Qianqian Wang, Quanxue Gao, Ming Yang, and Xinbo Gao. Self-consistent contrastive attributed graph clustering with pseudo-label prompt. IEEE Transactions on Multimedia, 2022.

[Xia et al., 2022f] Wei Xia, Qianqian Wang, Quanxue Gao, Xiangdong Zhang, and Xinbo Gao. Self-supervised graph convolutional network for multi-view clustering. IEEE Trans. Multim., 24:3182–3192, 2022.

[Xie et al., 2016] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In Proc. of ICML, 2016.

[Xue et al., 2022] Hansheng Xue, Vijini Mallawaarachchi, Yujia Zhang, Vaibhav Rajan, and Yu Lin. Repbin: Constraint-based graph representation learning for metagenomic binning. In Proc. of AAAI, 2022.

[Yang et al., 2020] Liang Yang, Yuexue Wang, Junhua Gu, Chuan Wang, Xiaochn Cao, and Yuanfang Guo. Jane: Jointly adversarial network embedding. In Proc. of IJCAI, 2020.

[Ying et al., 2018] Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec.
Hierarchical graph representation learning with differentiable pooling. *Proc. of NeurIPS*, 2018.

[You et al., 2020] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Proc. of NeurIPS*, 2020.

[Yu et al., 2022a] Lu Yu, Shichao Pei, Lihong Ding, Jun Zhou, Longfei Li, Chuxu Zhang, and Xiangliang Zhang. Sail: Self-augmented graph contrastive learning. In *Proc. of AAAI*, 2022.

[Yu et al., 2022b] Zhuohan Yu, Yifu Lu, Yunhe Wang, Fan Tang, Ka-Chun Wong, and Xiangtao Li. Zinb-based graph embedding autoencoder for single-cell ma-seq interpretations. In *Proc. of AAAI*, 2022.

[Zbontar et al., 2021] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. *arXiv preprint arXiv:2103.03230*, 2021.

[Zhang et al., 2019] Xiaotong Zhang, Han Liu, Qimai Li, and Xiao-Ming Wu. Attributed graph clustering via adaptive graph convolution. In *Proc. of IJCAI*, 2019.

[Zhang et al., 2020a] Kaiming Zhang, Tengpeng Li, Shiwen Shen, Bo Liu, Jin Chen, and Qingshan Liu. Adaptive graph convolutional network with attention graph clustering for co-saliency detection. In *Proc. of CVPR*, 2020.

[Zhang et al., 2020b] Tianqi Zhang, Yun Xiong, Jiawei Zhang, Yao Zhang, Yizhu Jiao, and Yangyong Zhu. Commdgi: community detection oriented deep graph infomax. In *Proc. of CIKM*, 2020.

[Zhang et al., 2022a] Hongyuan Zhang, Pei Li, Rui Zhang, and Xuelong Li. Embedding graph auto-encoder for graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[Zhang et al., 2022b] Junpu Zhang, Liang Li, Siwei Wang, Jiyuan Liu, Yue Liu, Xinwang Liu, and En Zhu. Multiple kernel clustering with dual noise minimization. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3440–3450, 2022.

Zhao et al., 2021] Han Zhao, Xu Yang, Zhenru Wang, Erkun Yang, and Cheng Deng. Graph debiased contrastive learning with joint representation clustering. In *Proc. of IJCAI*, 2021.

[Zheng et al., 2016] Stephan Zheng, Yang Song, Thomas Leung, and Ian Goodfellow. Improving the robustness of deep neural networks via stability training. In *Proceedings of the ieee conference on computer vision and pattern recognition*, pages 4480–4488, 2016.

[Zhong et al., 2021] Huasong Zhong, Jianlong Wu, Chong Chen, Jianqiang Huang, Minghua Deng, Liqiang Nie, Zhouchen Lin, and Xian-Sheng Hua. Graph contrastive clustering. In *Proc. of ICCV*, 2021.

[Zhou et al., 2022] Sheng Zhou, Hongjia Xu, Zhuoran Zheng, Jiawei Chen, Jiajun Bu, Jia Wu, Xin Wang, Wenwu Zhu, Martin Ester, et al. A comprehensive survey on deep clustering: Taxonomy, challenges, and future directions. *arXiv preprint arXiv:2206.07579*, 2022.

[Zhu et al., 2020] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Deep Graph Contrastive Representation Learning. In *ICML Workshop on Graph Representation Learning and Beyond*, 2020.

[Zhu et al., 2021] Yanqiao Zhu, Yichen Xu, Qiang Liu, and Shu Wu. An empirical study of graph contrastive learning. *arXiv preprint arXiv:2109.01116*, 2021.

[Zhu et al., 2022] Pengfei Zhu, Jialu Li, Yu Wang, Bin Xiao, Shuai Zhao, and Qinghua Hu. Collaborative decision-reinforced self-supervision for attributed graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.