PRELIMINARY PREPRINT VERSION – DO NOT CITE
This is not the version of record. The final version will appear in the Proceedings of IJCAI-ECAI 2026.

# FairTCD: Dual-Teacher Temporal Contrastive Distillation for Twofold Fair Dynamic Graph Embedding

**Yuxuan Gu**[1] , **Yicong Li**[1,2] , **Weiwei Yuan**[1*] , **Tianzi Zang**[1] , **Donghai Guan**[1*] , **Jason J. Jung**[3] , **Jie Zhao**[1] and **Yongbo Ma**[1]
[1]College of Computer Science and Technology,
Nanjing University of Aeronautics and Astronautics, P.R. China
[2]State Key Laboratory of Novel Software Technology, Nanjing University, P.R. China
[3]Department of Computer Engineering, Chung-Ang University, Republic of Korea
{guyuxuan, liyicong, yuanweiwei, zangtianzi, dhguan, pycro6815, mayongbo}@nuaa.edu.cn,
j3ung@cau.ac.kr

## Abstract

Fair dynamic graph embedding is crucial for real-world systems, such as recommendation and social networks. Prior studies impose a single-axis fairness formulation, treating attribute and structural bias as separable artifacts. This overlooks their coupling relationship, under which debiasing along one axis can induce cross-axis amplification. This coupling further introduces opposing gradient constraints under joint optimization, leading to optimization conflicts. Furthermore, the evolution of dynamic graphs causes shifts in bias distribution, leading to unstable optimization and exacerbating these conflicts. To address these issues, we propose **FairTCD**, a **Fair** Dual-Teacher **T**emporal **C**ontrastive **D**istillation framework. FairTCD employs two adversarial fairness teachers to decouple attribute and structural fairness representations. To reconcile dynamic conflicts between two fairness objectives, we introduce a temporal contrastive distillation to induce consistency between attribute and structural fairness representations across time while retaining temporal semantics. A unified student model distills complementary knowledge from both teachers to achieve twofold fairness. Experiments on three real-world benchmarks demonstrate that FairTCD preserves performance while improving twofold fairness metrics by at least 4.53%.

## 1 Introduction

Dynamic graph embedding has become an important technique in modeling temporal relational data and has found widespread applications in recommendation systems, social network analysis, and related domains [Barros *et al.*, 2021; Anand and Maurya, 2024; Pham *et al.*, 2025]. Despite their success, these embeddings often inherit and even amplify bias from graph data. To address this problem, research on fair graph embedding has attracted considerable attention [Dong *et al.*, 2023; Chen *et al.*, 2024a].

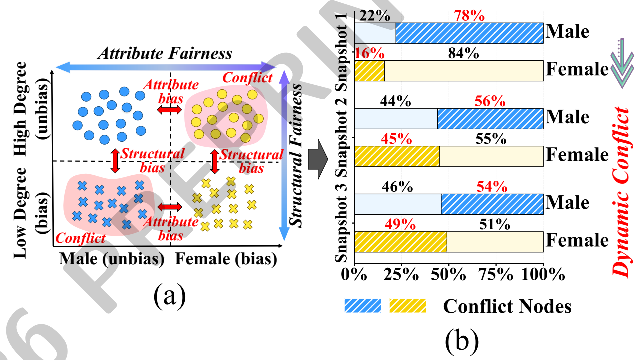


Figure 1: (a) Drawbacks of single fairness objective. (b) Statistical results of proportions of conflict nodes across three snapshots on the MovieLens-1M dataset.

The fairness in graph embedding is typically defined as **attribute fairness** and **structural fairness**. Attribute fairness aims to balance the model's performance among sensitive demographic groups (e.g., gender, race), while structural fairness aims to balance the model's performance among high-degree and low-degree node groups. To achieve attribute fairness, most studies introduce an adversarial network to obscure sensitive attribute information in node representations [Khajehnejad *et al.*, 2020; Dai and Wang, 2021; Zhang *et al.*, 2024]. To achieve structural fairness, some studies propose debiasing functions or contrastive learning to reduce the gap between representations of high-degree nodes and low-degree nodes [Liu *et al.*, 2023; Wang *et al.*, 2022].

***However, existing fairness studies typically optimize only one type of fairness, ignoring the co-existing twofold bias.*** As shown in Figure 1(a), when only attribute fairness along the horizontal axis is considered, structural bias along the vertical axis cannot be addressed. In contrast, focusing solely on structural fairness along the vertical axis also fails to mitigate attribute bias. This issue is further exacerbated in dynamic graphs, where evolving topologies continuously reshape node

*Corresponding authors.

degrees and neighborhood structures, leading to twofold bias that may drift across snapshots. This non-stationary twofold bias renders static and single-objective debiasing strategies ineffective. Therefore, to address these issues, this study aims to achieve both attribute and structural fairness (i.e., twofold fairness) in dynamic graph embedding.

Naively optimizing attribute and structural fairness simultaneously raises three fundamental challenges. **(i) Coupling relationship between two fairness objectives.** In practice, sensitive attribute groups and structural groups often exhibit substantial overlap, leading to interdependencies between the two fairness objectives. For example, the statistical results in Figure 1(b) show that nodes within the same sensitive attribute group can belong to different structural groups. **(ii) Optimization conflicts induced by the coupling relationship.** Drawing on the empirical results [Mansoury *et al.*, 2020], Figure (a) shows low-degree male users and high-degree female users exhibit opposite bias directions under the structural and attribute fairness objectives, which leads the two fairness objectives may impose conflicting gradient signals on the same node and hinder model convergence. Moreover, the statistical result of the third snapshot in Figure 1(b) shows that 54% of male users are low-degree nodes and 49% of female users are high-degree nodes, indicating that such conflicts are widespread rather than isolated phenomena. **(iii) Conflicts drift in dynamic graphs.** As bias distributions evolve in dynamic graphs, the optimization conflicts exhibit temporal drift. As shown in Figure 1(b), the proportions of these groups change over time, reflecting the non-stationary fairness signals and the evolving coupling relationship between the two fairness objectives.

To address these challenges, this study proposes a novel Fair Dual-Teacher Temporal Contrastive Distillation (FairTCD) framework for twofold fair dynamic graph embedding. To tackle the first challenge, we construct a dual-teacher adversarial Graph Neural Network (GNN) framework, in which attribute and structural fairness are learnt by two independent teachers via adversarial debiasing. This design separates the optimization processes and gradient flows of the two fairness objectives, decoupling the coupling relationship between two fairness objectives. To tackle the second challenge, we employ contrastive distillation to align structural and attribute fairness representations of the same node, thereby reconciling conflicting gradient directions induced by competing fairness objectives. To tackle the third challenge, we introduce temporal contrastive distillation (TCD), which extends contrastive distillation to the temporal dimension and generalizes fairness representations from snapshot-level to time-aware modeling, enabling adaptation to the conflict drift between twofold fairness objectives over time. Finally, a student model distills knowledge from both teachers for achieving twofold fairness. In summary, the primary contributions of this study are listed below.

- **A new problem and formulation.** We formally define the research problem and challenges of twofold fairness in dynamic graph embedding, in which attribute and structural fairness are coupled and conflict over time.
- **A novel and effective framework.** We propose FairTCD, a novel dynamic graph embedding framework that decouples attribute and structural objectives and reconciles their conflicts to achieve twofold fairness.
- **Superior Twofold Fairness.** Experiment results show that FairTCD significantly reduces both attribute and structural biases, while maintaining competitive embedding performance.

## 2 Problem Definition

In this section, we formally define the dynamic graphs, twofold fairness metrics, and the problem of twofold fair dynamic graph embedding.

**Definition** (Dynamic Graph and Twofold Fairness Metrics)**.** *A dynamic graph is defined as* $\mathcal{G} = \{\mathcal{G}_t\}_{t=1}^{T}$*, where each snapshot* $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t, \mathbf{X}_t)$ *consists of nodes, edges, and node attributes. Let* $\mathcal{V} = \bigcup_{t=1}^{T} \mathcal{V}_t$ *denote the set of all nodes across time. We adopt* Equal Opportunity (EO) *and* Statistical Parity (SP) *to define structural and attribute fairness. Given groups* $\{G_1, \dots, G_k\}$*, the EO-based and the SP-based fairness metrics are defined as:*

$$\mathbb{P}(\hat{y} = 1 \mid y = 1, G_1) = \cdots = \mathbb{P}(\hat{y} = 1 \mid y = 1, G_k), \quad (1)$$

$$\mathbb{P}(\hat{y} = 1 \mid G_1) = \cdots = \mathbb{P}(\hat{y} = 1 \mid G_k), \quad (2)$$

*where* $\hat{y}$ *and* $y$ *denote the predicted and true labels,* $\mathbb{P}$ *is the probability, and* $\{G_1, \dots, G_k\}$ *is either high and low degree structural groups or sensitive attribute groups. In practice, the disparities of EO and SP (i.e., DEO and DSP) across different groups are used as the corresponding fairness metrics. Following prior multi-objective optimization studies [Guerreiro* et al.*, 2021], we further adopt Hypervolume (HV) as a unified twofold fairness metric, which is defined as:*

$$\mathrm{HV}(\mathbf{f}) = \mathrm{Vol}\left(\bigcup_{\mathbf{u} \preceq \mathbf{f}} [\mathbf{u}, \mathbf{r}]\right) \quad (3)$$

*where* $\mathbf{f} = [\mathrm{DEO}_{\mathrm{att}}, \mathrm{DEO}_{\mathrm{str}}, \mathrm{DSP}_{\mathrm{att}}, \mathrm{DSP}_{\mathrm{str}}]$ *denotes the four-dimensional fairness objective vector consisting of the DEO and DSP for both attribute and structural fairness,* $\mathbf{r} = (1, 1, 1, 1)$ *is the reference point, and* $\mathbf{u} \preceq \mathbf{f}$ *indicates the Pareto-dominant region with respect to* $\mathbf{f}$*. Larger HV implies better twofold fairness.*

**Problem** (Twofold Fair Dynamic Graph Embedding)**.** *The objective is to learn a mapping function* $f_\theta : \{\mathcal{G}_t\}_{t=1}^{T} \mapsto \{\widetilde{\mathbf{H}}_{\mathrm{DF}}^{(T)} \in \mathbb{R}^{d \times \mathcal{V}}\}$ *that produces a* $d$*-dimensional embedding* $\widetilde{\mathbf{H}}_{\mathrm{DF}}^{(T)}$ *for the* $T$*-th snapshot, such that this embedding jointly optimizes downstream task performance, structural fairness, and the attribute fairness.*

## 3 Methodology

Figure 2 presents the FairTCD framework. First, the dual-teacher fairness model learns structural and attribute fairness features respectively. Then, a TCD module aligns their cross-snapshot representations to reconcile dynamic conflicts while preserving fairness and temporal features. Finally, a twofold fairness student model inherits both fairness objectives.

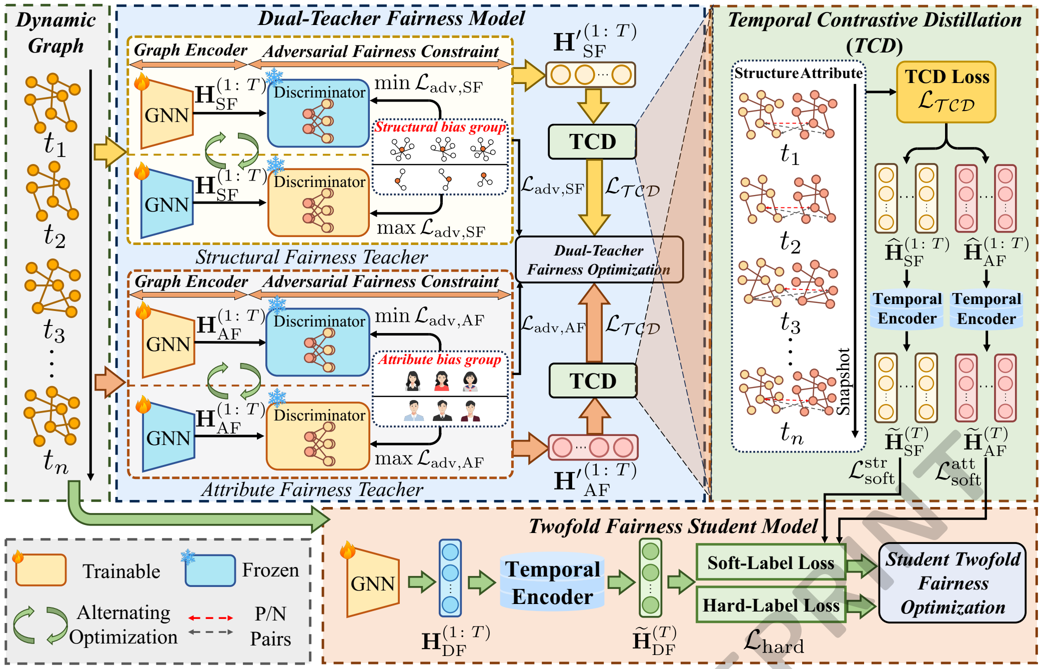


Figure 2: Overall framework of FairTCD. P/N pairs denote positive and negative samples, shown by red and black dotted lines, respectively.



### 3.1 Dual-Teacher Fairness Model

Since the structural and attribute fairness teachers are built upon the same adversarial debiasing paradigm, we present a unified group-adversarial fairness teacher. In the following, we define $* \in \{\text{SF}, \text{AF}\}$ to index the instantiated structural or attribute fairness teacher.

**Snapshot-wise Graph Encoder.** Given a dynamic graph sequence $\{\mathcal{G}_t\}_{t=1}^{T}$, the fairness teacher aims to learn node representations that are informative for downstream tasks while being invariant to group membership. At each snapshot, a GNN encoder is first employed to capture local structural and feature-based dependencies. As our primary focus lies in fairness modeling rather than architectural design, we adopt the widely used GCN [Kipf and Welling, 2017] as the graph encoder:

$$\mathbf{H}_*^{(1:T)} = \text{GCN}(\{\mathcal{G}_t\}_{t=1}^{T}), \tag{4}$$

where $\mathbf{H}_*^{(1:T)} \in \mathbb{R}^{|\mathcal{V}_t| \times d}$ denotes the node embeddings from snapshot 1 to snapshot $T$.

**Adversarial Fairness Constraint.** To remove bias encoded in the learned representations, the fairness teacher incorporates an adversarial discriminator that attempts to infer sensitive group membership from node embeddings, while the GNN encoder learns to suppress sensitive information.

Let $\mathbf{Y}^{(t)} = \{0,1\}^{|\mathcal{V}_t| \times 2}$ denote the group labels at snapshot $t$, where $|\mathcal{V}_t|$ denotes the number of nodes in the $t$-th snapshot. For structural fairness, $\mathbf{Y}^{(t)}$ denotes high-degree and low-degree group . For attribute fairness, $\mathbf{Y}^{(t)}$ corresponds to sensitive attributes. At each snapshot, the discriminator maps node embeddings to group membership probabilities via a Multi-Layer Perceptron (MLP):

$$\mathbf{P}_*^{(t)} = \text{softmax}(\text{MLP}(\mathbf{H}_*^{(t)})). \tag{5}$$

The discriminator is trained to maximize its ability to infer group membership from node embeddings, quantified by the cross-entropy loss:

$$\mathcal{L}_{\text{adv},*}^{(t)} = -\frac{1}{|\mathcal{V}_t|}\mathbf{1}^\top\left(\mathbf{Y}^{(t)} \odot \log \mathbf{P}_*^{(t)}\right)\mathbf{1}, \tag{6}$$

where $\mathbf{1}$ denotes the all-ones vector and $\odot$ denotes element-wise product. By aggregating over all snapshots, we obtain the overall adversarial objective:

$$\min_{\theta_G} \max_{\theta_D} \ \mathcal{L}_{\text{adv},*} = \phi(\mathcal{L}_{\text{adv},*}^{(1:T)}), \tag{7}$$

where $\theta_G$ and $\theta_D$ denote parameter set of GCN encoder and discriminator of teacher, $\mathcal{L}_{\text{adv,o}}^{(1:T)}$ denotes the adversarial loss from snapshot 1 to $T$, and $\phi(\cdot)$ denotes mean pooling. Through this minimax optimization, the discriminator continuously amplifies bias-related signals, while the encoder is forced to eliminate group-discriminative information. As a result, the encoder generates fairness representations $\mathbf{H}_*'^{(1:T)}$.

**Dual-Teacher Fairness Optimization.** While adversarial debiasing effectively mitigates bias within each snapshot, two fairness objectives in dynamic graphs are inherently non-stationary and may evolve and conflict over time. To address this temporal conflict between two fairness objectives, we further introduce a TCD module to guide the evolution of fairness representations:

$$\widetilde{\mathbf{H}}_*^{(T)} = \mathcal{TCD}(\mathbf{H}_*'^{(1:T)}), \tag{8}$$

where $\widetilde{\mathbf{H}}_*^{(T)}$ denotes the final fairness feature of teacher model at snapshot $T$. The complete training objective of a fairness teacher is thus defined as:

$$\mathcal{L}_{\text{teacher},*} = \mathcal{L}_{\text{adv},*} + \mathcal{L}_{\mathcal{TCD}}. \tag{9}$$

where $\mathcal{L}_{\mathcal{TCD}}$ is the TCD loss that enforces consistency in the evolution of fairness representations across snapshots. The details of TCD module will be explained in Section 3.2.

### 3.2 Temporal Contrastive Distillation

Since structural and attribute fairness objectives impose opposite gradient signals on certain nodes, this optimization conflict makes direct joint distillation ineffective. To address this, we propose TCD to reconcile the dynamic conflicts by aligning the representations of the same node from both teachers across snapshots while preserving their diversity.

First, the structural and attribute fairness representations, denoted as $\mathbf{H'}^{(1:T)}_{\text{SF}}$ and $\mathbf{H'}^{(1:T)}_{\text{AF}}$, are obtained by instantiating the unified fairness representation $\mathbf{H}'^{(1:T)}_{*}$ with different group definitions. Then, we construct positive pairs and negative pairs in the $t$-th snapshot. The positive pairs are composed of structural and attribute fairness embedding from the same node. The negative pairs consist of the embeddings of different nodes from different teachers. Specifically, the positive pair at snapshot $t$ is defined as $\mathcal{P}_t = \{(u^S, u^A)\}$, where $u^S$ and $u^A$ denote the embedding of node $u$ from the structural and attribute teacher, respectively. The negative pair is defined as $\mathcal{N}_t = \{(u^S, v^A) \mid u \neq v, v \in \mathcal{V}_t\}$. Let $h'^{\text{SF}}_{t,u} \in \mathbf{H'}^{(1:T)}_{\text{SF}}$ and $h'^{\text{AF}}_{t,v} \in \mathbf{H'}^{(1:T)}_{\text{AF}}$ denote the structural fairness embedding of node $u$ and attribute fairness embedding of node $v$ at snapshot $t$, respectively. We average the contrastive distillation loss across all nodes and snapshots to obtain $\mathcal{L}_{\mathcal{TCD}}$, which is defined as:

$$\begin{aligned} \mathcal{L}_{\mathcal{TCD}} &= g\Big(-\log \tfrac{\langle u^S, u^A\rangle^t}{\langle u^S, u^A\rangle^t + \sum_{v\neq u}(\langle u^S, v^A\rangle^t + \langle u^A, v^S\rangle^t)}\Big), \\ \langle u^S, v^A\rangle^t &= \exp\big(\text{sim}(h'^{\text{SF}}_{t,u}, h'^{\text{AF}}_{t,v})/\tau\big), \end{aligned} \tag{10}$$

where $g(\xi) = \frac{1}{T|\mathcal{V}|}\sum_{t=1}^{T}\sum_{u\in\mathcal{V}} \xi$, $\text{sim}(\cdot)$ denotes cosine similarity, and $\tau$ is the scaling parameter. The $\mathcal{L}_{\mathcal{TCD}}$ loss aligns embeddings produced by the attribute and structural teachers to reconcile conflicts. Theoretical proof is presented in Appendix A.

**Temporal Encoder.** By minimizing the structural and attribute teacher loss (Eq. 9), we can induce contrasted structural fairness feature $\widehat{\mathbf{H}}^{(1:T)}_{\text{SF}}$ and contrasted attribute fairness feature $\widehat{\mathbf{H}}^{(1:T)}_{\text{AF}}$ for each snapshot. Then, a temporal encoder is designed to aggregate dynamic fairness features. To reduce computational cost, we adopt a decay factor to encode fairness features of historical snapshots for obtaining final fairness features $\widetilde{\mathbf{H}}^{(T)}_{\text{SF}}$ and $\widetilde{\mathbf{H}}^{(T)}_{\text{AF}}$:

$$\widetilde{\mathbf{H}}^{(T)}_{\text{SF}} = \sum_{d=0}^{T-1} \frac{\gamma^{T-d}}{\sum_{j=1}^{T}\gamma^j}\widehat{\mathbf{H}}^{(T-d)}_{\text{SF}}, \tag{11}$$

$$\widetilde{\mathbf{H}}^{(T)}_{\text{AF}} = \sum_{d=0}^{T-1} \frac{\gamma^{T-d}}{\sum_{j=1}^{T}\gamma^j}\widehat{\mathbf{H}}^{(T-d)}_{\text{AF}}, \tag{12}$$

where $\widehat{\mathbf{H}}^{(T-d)}_{\text{SF}}$ and $\widehat{\mathbf{H}}^{(T-d)}_{\text{AF}}$ denote the structural fairness feature and the attribute fairness feature of the $(T-d)$-th snapshot obtained after updating via the two teacher losses, and $\gamma$ denotes the decay factor.

### 3.3 Twofold Fairness Student Model

The goal of the student model is to jointly achieve twofold fairness by distilling complementary fairness knowledge from two teacher models, while preserving downstream performance. The student model is supervised by both *hard label loss* and *soft label loss*. Specifically, the student model first learns dynamic graph representations using a GCN with temporal decay to integrate evolving information, which is defined as:

$$\mathbf{H}^{(1:T)}_{\text{DF}} = \text{GCN}_{\text{DF}}(\{\mathcal{G}_t\}_{t=1}^{T}), \tag{13}$$

$$\widetilde{\mathbf{H}}^{(T)}_{\text{DF}} = \sum_{d=0}^{T-1} \frac{\gamma^{T-d}}{\sum_{j=1}^{T}\gamma^j}\mathbf{H}^{(T-d)}_{\text{DF}}, \tag{14}$$

where $\text{GCN}_{\text{DF}}$ denotes the GCN of student model.

**Hard Label Loss.** Hard labels provide ground-truth supervision for the downstream task. In this study, we employ link prediction as the downstream task and adopt the Bayesian Personalized Ranking (BPR) loss as the student model's hard-label loss:

$$\mathcal{L}_{\text{hard}} = -\frac{1}{|\mathcal{I}|}\sum_{(e^+,e^-)\in\mathcal{I}} \log\, \sigma\,(S_{e^+} - S_{e^-}), \tag{15}$$

where $|\mathcal{I}|$ denotes the size of the set of sample pairs formed by a positive edge $e^+$ and negative edge $e^-$, $S_{e^+}$ denotes the scores of positive edges, $S_{e^-}$ denotes the scores of negative edges, and $\sigma$ indicates the sigmoid function. The scores of the edges between two nodes are calculated by the dot product of the two node embeddings.

**Soft Label Loss.** To align the features of the student model with the two teacher models, we employ the mean-squared error (MSE) between their features as the soft-label distillation loss for the student model, which is defined as:

$$\mathcal{L}^{\text{str}}_{\text{soft}} = ||\widetilde{\mathbf{H}}^{(T)}_{\text{DF}} - \widetilde{\mathbf{H}}^{(T)}_{\text{SF}}||_F^2, \mathcal{L}^{\text{att}}_{\text{soft}} = ||\widetilde{\mathbf{H}}^{(T)}_{\text{DF}} - \widetilde{\mathbf{H}}^{(T)}_{\text{AF}}||_F^2 \tag{16}$$

where $||\cdot||_F^2$ denotes squared Frobenius norm.

**Student Twofold Fairness Optimization.** The overall objective of the student model combines the hard-label loss for the downstream task with two soft-label distillation terms, formulated as:

$$\mathcal{L}_{\text{student}} = \alpha\mathcal{L}_{\text{hard}} + \beta\mathcal{L}^{\text{str}}_{\text{soft}} + (1-\beta)\mathcal{L}^{\text{att}}_{\text{soft}}, \tag{17}$$

where $\alpha$ and $\beta$ denote loss weights. The pseudo code of FairTCD is presented in Appendix D.

### 3.4 Time Complexity

FairTCD performs two primary operations. The first operation computes pairwise contrastive losses over all $n$ nodes in a snapshot, incurring $O(n^2)$ time. The second operation performs a graph convolution on the adjacency matrix to propagate and linearly transform node features. In graphs where the total number of edges grows linearly with the number of nodes, each graph convolution layer has a cost proportional to the number of edges and therefore runs in $O(n)$ time. Therefore, the total time complexity is $O(n^2 + n)$, which reduces to $O(n^2)$, omitting lower-order terms. In addition, the scalability of FairTCD is analyzed in Appendix E.

| Dataset | | LightGCN | FairGNN | FairSIN | FairGKD | DegFairGNN | GRADE | LibraGNN | FTGNN-M | FairTCD |
|---|---|---|---|---|---|---|---|---|---|---|
| ML-1M | HR@20 ↑ | 32.41 ±0.74 | 31.26 ±0.43 | 29.31 ±3.12 | 30.55 ±0.06 | 31.72 ±0.68 | 29.66 ±1.71 | 28.83 ±2.53 | 28.41 ±2.16 | 31.10 ±1.65 |
| | NDCG@20 ↑ | 5.17 ±0.07 | 4.43 ±0.09 | 4.06 ±0.56 | 4.39 ±0.08 | 5.11 ±0.06 | 3.76 ±0.30 | 4.24 ±0.43 | 4.31 ±0.42 | 4.97 ±0.23 |
| | $\text{DEO}_{\text{att}}$ ↓ | 4.59 ±0.37 | 2.37 ±0.08 | 2.79 ±0.20 | 3.79 ±0.17 | 2.11 ±0.30 | 3.06 ±0.84 | 2.89 ±0.15 | 2.80 ±0.35 | **0.13 ±0.08** |
| | $\text{DEO}_{\text{str}}$ ↓ | 8.85 ±0.58 | 3.61 ±0.21 | 4.36 ±0.30 | 5.32 ±0.22 | 2.22 ±0.28 | 3.35 ±0.52 | 4.03 ±0.13 | 3.06 ±0.12 | **0.75 ±0.10** |
| | $\text{DSP}_{\text{att}}$ ↓ | 2.31 ±0.20 | 1.80 ±0.09 | 1.83 ±0.06 | 1.84 ±0.10 | 1.95 ±0.09 | 1.37 ±0.38 | 1.79 ±0.04 | 1.97 ±0.06 | **0.30 ±0.04** |
| | $\text{DSP}_{\text{str}}$ ↓ | 5.54 ±0.48 | 1.82 ±0.02 | 1.63 ±0.07 | 1.19 ±0.12 | 1.91 ±0.07 | 1.05 ±0.53 | 1.62 ±0.04 | 1.74 ±0.01 | **0.88 ±0.11** |
| | HV ↑ | 80.26 ±1.39 | 90.71 ±0.17 | 89.76 ±0.42 | 88.34 ±0.04 | 92.05 ±0.44 | 91.33 ±0.68 | 90.03 ±0.21 | 90.74 ±0.42 | **98.82 ±0.11** |
| ELEC | HR@20 ↑ | 24.32 ±0.25 | 22.75 ±1.02 | 19.86 ±2.44 | 21.45 ±0.16 | 23.65 ±0.01 | 23.31 ±1.93 | 23.51 ±0.27 | 23.65 ±0.27 | 23.85 ±0.75 |
| | NDCG@20 ↑ | 11.01 ±0.08 | 9.97 ±0.24 | 9.49 ±0.96 | 9.52 ±0.06 | 10.86 ±0.01 | 10.04 ±0.40 | 10.57 ±0.16 | 10.69 ±0.24 | 10.89 ±0.44 |
| | $\text{DEO}_{\text{att}}$ ↓ | 5.08 ±0.09 | 4.29 ±0.86 | 3.02 ±1.44 | 3.63 ±2.63 | 1.69 ±0.18 | 2.38 ±0.12 | 3.11 ±0.51 | 5.00 ±1.71 | **1.24 ±0.33** |
| | $\text{DEO}_{\text{str}}$ ↓ | 10.19 ±0.31 | 0.54 ±0.10 | 3.13 ±2.22 | 3.42 ±3.08 | **0.43 ±0.02** | 1.79 ±0.07 | 1.42 ±0.25 | 5.52 ±0.84 | 0.50 ±0.28 |
| | $\text{DSP}_{\text{att}}$ ↓ | 3.21 ±0.17 | 5.97 ±0.98 | 3.16 ±1.63 | 0.39 ±0.37 | 7.96 ±2.25 | 2.66 ±0.16 | 2.30 ±0.36 | 5.73 ±1.75 | **0.24 ±0.05** |
| | $\text{DSP}_{\text{str}}$ ↓ | 7.86 ±0.46 | 0.38 ±0.11 | 3.04 ±2.25 | 0.15 ±0.01 | 0.97 ±0.12 | **0.12 ±0.01** | 0.28 ±0.15 | 2.69 ±0.96 | 0.28 ±0.08 |
| | HV ↑ | 76.03 ±0.85 | 89.16 ±1.80 | 88.24 ±4.40 | 92.48 ±0.05 | 89.21 ±2.49 | 93.20 ±0.01 | 93.05 ±0.66 | 82.33 ±2.09 | **97.75 ±0.54** |
| CLOTH | HR@20 ↑ | 23.10 ±0.10 | 21.94 ±0.39 | 18.29 ±0.32 | 16.20 ±0.63 | 21.90 ±0.14 | 21.73 ±0.40 | 21.66 ±0.51 | 22.16 ±0.37 | 21.82 ±0.71 |
| | NDCG@20 ↑ | 6.61 ±0.02 | 5.53 ±0.09 | 5.61 ±0.05 | 4.39 ±0.19 | 6.26 ±0.01 | 6.24 ±0.09 | 6.29 ±0.08 | 6.29 ±0.06 | 6.39 ±0.15 |
| | $\text{DEO}_{\text{att}}$ ↓ | 9.07 ±0.36 | 8.93 ±0.57 | 8.85 ±0.74 | 1.75 ±0.39 | 4.50 ±0.42 | 5.79 ±0.46 | 2.55 ±0.10 | 4.99 ±0.28 | **1.46 ±0.16** |
| | $\text{DEO}_{\text{str}}$ ↓ | 17.19 ±0.61 | 4.67 ±0.03 | 10.57 ±0.97 | 11.25 ±1.39 | 6.22 ±0.87 | 10.15 ±0.70 | 4.67 ±0.17 | 8.74 ±0.38 | **2.85 ±0.23** |
| | $\text{DSP}_{\text{att}}$ ↓ | 6.00 ±0.22 | 7.03 ±1.05 | 6.80 ±1.12 | 1.97 ±0.54 | 4.80 ±0.29 | 2.38 ±0.29 | 1.10 ±0.04 | 2.22 ±0.16 | **0.59 ±0.07** |
| | $\text{DSP}_{\text{str}}$ ↓ | 13.65 ±0.30 | 8.78 ±0.35 | 8.64 ±0.31 | 4.56 ±0.22 | 6.11 ±0.51 | 5.12 ±0.47 | 2.41 ±0.10 | 4.91 ±0.30 | **1.51 ±0.12** |
| | HV ↑ | 61.12 ±0.92 | 69.14 ±0.08 | 69.39 ±0.52 | 81.58 ±1.95 | 80.05 ±1.38 | 78.41 ±1.51 | 89.65 ±0.38 | 80.60 ±0.95 | **93.72 ±0.56** |

Table 1: Performance and fairness comparison of different methods. Bold denotes the best, and underline denotes the second best.

| Dataset | # User | # Item | # Edge | Sensitive attribute |
|---|---|---|---|---|
| ML-1M | 677 | 3135 | 35571 | Gender |
| ELEC | 4060 | 940 | 4368 | Gender |
| CLOTH | 11537 | 1017 | 19331 | Body shape |

Table 2: Statistics of experimental datasets.

# 4 Experiments

This section reports experimental results addressing five research questions (RQs). **RQ1:** Does FairTCD improve attribute and structural fairness while preserving embedding capability? **RQ2:** Is TCD more effective than existing alignment methods? **RQ3:** How does FairTCD balance the performance–fairness trade-off? **RQ4:** Does TCD effectively reconcile the dynamic conflicts? **RQ5:** How does hyperparameter selection influence performance and fairness? The source code of FairTCD is available at https://github.com/FairTCD.

## 4.1 Experiment Setting

**Datasets.** We evaluate the FairTCD on three real-world datasets: MovieLens-1M (ML-1M) [Harper and Konstan, 2015], Amazon Electronics (ELEC) [Wan *et al.*, 2020], and ModCloth (CLOTH) [Wan *et al.*, 2020]. We divide the ML-1M dataset of 677 users annotated with gender and 3135 movies from September 2001 to January 2003 into five snapshots. We split the ELEC dataset of 4060 gender labeled users and 940 products from January 2012 to January 2013 into five snapshots. Moreover, we split the ModCloth dataset of 11537 users annotated with body shape information and 1017 clothing items from September 2016 to August 2018 into eight snapshots. Dataset statistics are summarized in Table 2.

**Baseline methods.** We compare against eight recent fair graph embedding baseline methods, categorized as *graph embedding* (LightGCN [He *et al.*, 2020]), *attribute fairness graph embedding* (FairGNN [Dai and Wang, 2021], FairSAD [Zhu *et al.*, 2024b], FairSIN [Yang *et al.*, 2024], Fairgap [Chen *et al.*, 2024b], FairGKD [Zhu *et al.*, 2024a]), *structural fairness graph embedding* (DegFairGNN [Liu *et al.*, 2023], GRADE [Wang *et al.*, 2022]), *hybrid fairness graph embedding* (LibraGNN [Luo *et al.*, 2025b]), and *dynamic attribute fairness graph embedding* (FTGNN-M [Song *et al.*, 2025]). As none of the baselines jointly address attribute and structural fairness, we equip each with adversarial debiasing for twofold fairness and apply a decay factor to ensure fair comparison in dynamic graphs.

**Evaluation metrics.** We evaluate link prediction performance using HR@$k$ and NDCG@$k$ and measure structural and attribute fairness using DEO and DSP (i.e., $\text{DEO}_{\text{att}}$, $\text{DEO}_{\text{str}}$, $\text{DSP}_{\text{att}}$, and $\text{DSP}_{\text{str}}$). We further adopt HV as a unified twofold fairness metric.

**Implementation details.** The experiments were conducted on Ubuntu 22.04.5 with an Intel Xeon Platinum 8336C CPU, 125 GB RAM, and an NVIDIA GeForce RTX 3090 GPU. FairTCD was implemented in Python 3.11 using PyTorch 2.1, with ReLU activation, dropout 0.5, and Adam (learning rate 1e-3), trained for 150 epochs with 64-dimensional hidden and output layers. The results are averaged over five runs with mean and standard deviation reported.

## 4.2 RQ1: Performance and Fairness Comparison

Table 1 reports the performance and twofold fairness results on three dynamic graph benchmarks. Overall, FairTCD maintains competitive performance while substantially reducing both attribute and structural bias across all datasets. Compared with the performance-oriented baseline LightGCN, FairTCD achieves a comparable performance while improving the HV by 23.12%-53.33%. In contrast, existing

| ELEC | SCA | CCA | MOS | TCD |
|---|---|---|---|---|
| HR@20 ↑ | 23.18 ±1.17 | 23.65 ±0.62 | 23.95 ±0.43 | 23.85 ±0.75 |
| $DSP_{att}$ ↓ | 3.66 ±1.97 | 4.60 ±2.69 | 5.51 ±1.73 | 0.24 ±0.05 |
| $DSP_{str}$ ↓ | 2.63 ±1.66 | 1.98 ±0.86 | 2.65 ±1.09 | 0.28 ±0.08 |
| HV ↑ | 93.80 ±2.45 | 93.52 ±2.92 | 91.98 ±2.11 | 99.48 ±0.09 |
| **CLOTH** | SCA | CCA | MOS | TCD |
| HR@20 ↑ | 21.67 ±0.46 | 21.86 ±0.43 | 21.76 ±0.31 | 21.82 ±0.71 |
| $DSP_{att}$ ↓ | 1.93 ±0.08 | 2.40 ±0.13 | 3.98 ±0.12 | 0.59 ±0.07 |
| $DSP_{str}$ ↓ | 4.82 ±0.11 | 5.88 ±0.36 | 9.04 ±0.15 | 1.51 ±0.12 |
| HV ↑ | 93.34 ±0.18 | 91.86 ±0.46 | 87.34 ±0.24 | 97.91 ±0.13 |

Table 3: Ablation study of FairTCD.

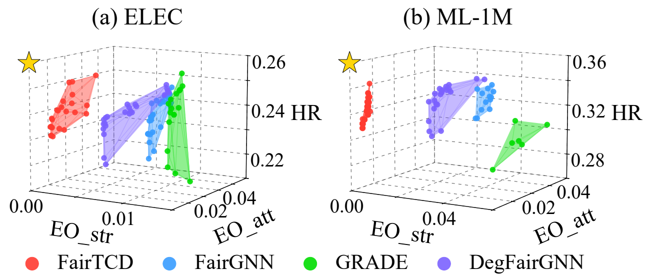


Figure 3: Pareto front comparison. Star denotes the optimal point in objective space.

fairness methods that do not explicitly reconcile the conflict between two fairness tend to improve one fairness dimension at the expense of the other, resulting in lower HV.

Although LibraGNN incorporates a decay factor module to enable temporal modeling, its static preference control remains limited in coordinating the coupling between dynamically evolving dual fairness objectives, resulting in a 4.53%–9.76% lower HV compared with FairTCD. Moreover, although FTGNN-M is equipped with an additional structural fairness module, it fails to reconcile the dynamic coupling between attribute and structural fairness objectives over time, causing gradient components induced by different fairness constraints to conflict across snapshots and consequently amplify biases along both dimensions. In summary, by explicitly modeling and reconciling snapshot-wise dynamic conflicts through the TCD mechanism, FairTCD achieves the highest HV scores, demonstrating its advantage in jointly balancing the dynamic twofold fairness objectives.

## 4.3 RQ2: Ablation study for FairTCD

Table 3 investigates the contribution of the TCD module by replacing it with three alternative designs, namely simple cosine alignment (SCA), canonical correlation analysis (CCA), and an ordinary multi-objective scalarization with learned weights (MOS). When TCD is substituted by SCA or CCA, the model retains comparable performance but reduced HV (computed based on $\mathrm{DSP}_{\mathtt{att}}$ and $\mathrm{DSP}_{\mathtt{str}}$) by 2.19%–6.58%, indicating that static or projection-based alignment is insufficient to reconcile the complex and temporal shifting conflicts. In contrast, TCD explicitly treats the structural and attribute fairness representations of the same node as positive pairs and enforces snapshot-wise alignment through contrastive objectives. This design preserves node-level consistency between the two fairness teachers and suppresses conflicting gradient components induced by competing fairness objectives as the graph evolves over time. By extending contrastive distillation to the temporal dimension, TCD further captures the non-stationary nature of fairness conflicts, enabling adaptation to shifting coupling patterns between structural and attribute biases. Furthermore, removing TCD and adopting a simple MOS to directly distill dual teachers leads to unstable fairness trade-offs, which validates the limitations of scalar-based multi-objective optimization under conflicting fairness objectives. The additional ablation results on the ML-1M dataset are presented in Appendix F.

## 4.4 RQ3: Pareto Front Trade-off Analysis

Figure 3 presents the Pareto fronts obtained by FairTCD, FairGNN, GRADE, and DegFairGNN on the ELEC and ML-1M datasets, where each Pareto solution is selected from 2000 randomly sampled hyperparameter configurations and simultaneously considers HR and two fairness objectives. A solution is Pareto-dominant if it achieves higher recommendation performance while incurring lower attribute and structural fairness gaps. As shown in Figure 3, FairTCD consistently achieves a Pareto front that dominates those of the other three methods, indicating a more favorable performance–fairness trade-off across a densely explored hyperparameter space. In contrast, although baseline methods are additionally equipped with debiasing modules for the other fairness dimension, they still lack explicit mechanisms to reconcile competing fairness objectives, leading to constrained and suboptimal trade-off regions even under extensive hyperparameter sampling. By dynamically reconciling multi-objective conflicts, FairTCD expands the feasible solution space and attains Pareto superior solutions.

## 4.5 RQ4: Dual-Teacher Knowledge Alignment

To validate the effectiveness of the TCD in reconciling dynamic conflicts, Figure 4(a) first presents the L2-distance distributions between the student model and the two fairness teacher models when removing the TCD. The results indicate that the L2-distance between the student model embeddings and the teacher model embeddings consistently remains at a high value, exhibiting significant fluctuations across different snapshots. It indicates that the student model fails to distill twofold fairness knowledge from attribute and structural teachers and the dynamic conflicts consistently persist across different snapshots. By introducing TCD, as shown in Figure 4(b), the L2-distance becomes smaller and more concentrated across snapshots. This is because representations produced by the two fairness teachers for the same node are explicitly treated as positive pairs and aligned at each snapshot, which continuously suppresses opposing gradient components as the graph evolves. These results demonstrate that TCD effectively reconciles dynamically shifting conflicts between attribute and structural fairness.

## 4.6 RQ5: Hyperparameter Sensitivity Analysis

We conduct a hyperparameter analysis on the ELEC dataset by varying the number of snapshots and the temporal decay

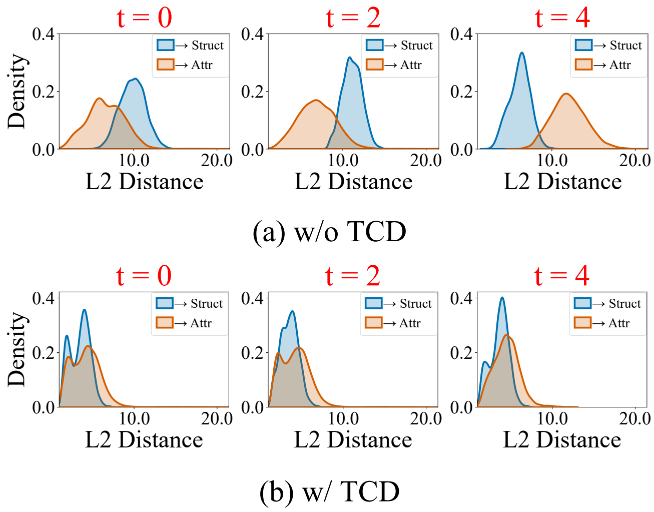


Figure 4: Visualization of L2-distance distribution between student model and dual-teacher model over snapshot.

factor, as shown in Figure 5, with performance evaluated using HR and NDCG at $k = 30$, and attribute and structural fairness measured by $\mathrm{DEO}_{\mathtt{att}}$ and $\mathrm{DEO}_{\mathtt{str}}$. **(1) Number of snapshots:** Figure 5(a) shows the performance peaks at four snapshots, reflecting a balance between capturing short-term dynamics and avoiding data sparsity. As the number of snapshots increases, both performance and fairness slightly decline because low-degree nodes experience sparse interactions while high-degree nodes remain relatively active. **(2) Decay factor:** Figure 5(b) indicates that larger decay factors, which emphasize long-term history, lead to mismatches with current preferences and bias accumulation, resulting in degraded performance and fairness. Additional hyperparameter analysis on the ML-1M dataset is presented in Appendix C.

# 5 Related Work

## 5.1 Attribute Fairness Graph Embedding

Traditional attribute fairness methods mainly mitigate sensitive attribute bias through pre-processing and in-processing strategies. Pre-processing approaches rebalance random walks to ensure fair exposure, such as Fairwalk, which modifies Node2Vec transition probabilities to balance visitation across sensitive groups [Rahman *et al.*, 2019; Grover and Leskovec, 2016]. In-processing methods typically adopt adversarial debiasing to remove sensitive information from representations [Luo *et al.*, 2025a; Li *et al.*, 2024a; Zhang *et al.*, 2024]. For instance, FairGNN couples a GCN encoder with an adversarial classifier to eliminate gender information while handling missing sensitive labels [Dai and Wang, 2021]. In addition, contrastive learning-based approaches have been proposed to disentangle sensitive and non-sensitive factors [Zhang *et al.*, 2025; Ling *et al.*, 2023]. FDGNN employs disentangled contrastive learning to suppress sensitive information in node representations [Zhang *et al.*, 2025], while FairGKD and FairSIN leverage knowledge distillation and fairness-facilitating neighbor features, respectively, to further mitigate attribute bias [Zhu *et al.*, 2024a; Yang *et al.*, 2024]. In the attribute fairness dynamic graph embedding, FTGNN-M [Song *et al.*, 2025] monitored attribute fairness in real time, identifying sensitive embedding channels responsible for bias, and adaptively adjusting the corresponding model parameters.

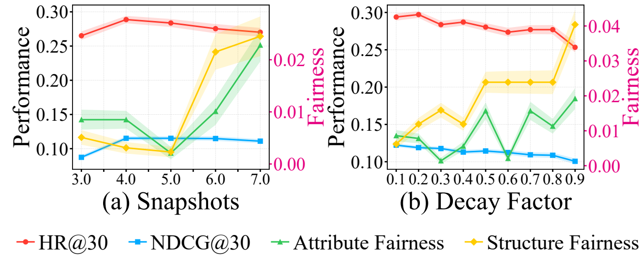


Figure 5: Sensitivity of number of snapshots and decay factor.

## 5.2 Structural Fairness Graph Embedding

Structural fairness in node embeddings has recently received considerable attention due to biases between high- and low-degree nodes. RawlsGCN [Kang *et al.*, 2022] employs Rawlsian pre- and in-processing to balance predictive accuracy across degrees without altering model architecture. GRADE [Wang *et al.*, 2022] applies graph contrastive learning to mitigate representation bias between nodes of varying degree. DegFairGNN [Liu *et al.*, 2023] introduces a learnable debiasing function to refine low- and high-degree neighborhoods and eliminate degree bias. FairDGE [Li *et al.*, 2024b] extends fairness to dynamic graphs by applying dual debiasing for fairness over time. In addition, LibraGNN [Luo *et al.*, 2025b] jointly considers hybrid fairness via multi-teacher distillation and Pareto optimization, but it relies on static fairness teachers and preference control, making it unable to handle the temporal evolution of coupling relationship of twofold fairness objectives in dynamic graphs, thereby cannot achieving the twofold fairness in dynamic graph settings. ***However, prior studies only mitigate either attribute or structural bias in isolation, failing to consider the coupling relationship between them and ensure twofold fairness in dynamic graphs.***

# 6 Conclusion

This study addressed the key problem of twofold fairness in dynamic graph embedding. To decouple the relationship between twofold fairness objectives, we proposed dual-teacher fairness model to learn each fairness feature and separate the optimization processes. To reconcile the dynamic conflicts, a TCD was introduced to induce consistency between the two fairness representations while preserving diversity and temporal semantics. Finally, a student model optimized downstream performance and aligned with both teachers, achieving twofold fairness. Experiments on three benchmarks showed FairTCD improves the twofold fairness while keeping embedding performance. In future work, we will explore reinforcement learning to adaptively learn distillation weights, enabling the model to adaptively adjust to the dynamic conflicts between fairness objectives.

## Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No.62472220, No.62502204), the Natural Science Fund of Jiangsu Province (Grant No. BK20251381), the Open Project Program of State Key Laboratory of Novel Software Technology, Nanjing University (KFKT2025B14), the Fundamental Research Funds for the Central Universities (NS2025041), Collaborative Innovation Center of Novel Software Technology and Industrialization. This work was partially supported by High Performance Computing Platform of Nanjing University of Aeronautics and Astronautics.

## Contribution Statement

Yuxuan Gu and Yicong Li made equal contributions to this work.

## References

[Anand and Maurya, 2024] Vineeta Anand and Ashish Kumar Maurya. A survey on recommender systems using graph neural network. *ACM Trans. Inf. Syst.*, 43(1), November 2024.

[Barros *et al.*, 2021] Claudio D. T. Barros, Matheus R. F. Mendonça, Alex B. Vieira, and Artur Ziviani. A survey on embedding dynamic graphs. *ACM Comput. Surv.*, 55(1), November 2021.

[Chen *et al.*, 2024a] April Chen, Ryan A. Rossi, Namyong Park, Puja Trivedi, Yu Wang, Tong Yu, Sungchul Kim, Franck Dernoncourt, and Nesreen K. Ahmed. Fairness-aware graph neural networks: A survey. *ACM Trans. Knowl. Discov. Data*, 18(6), April 2024.

[Chen *et al.*, 2024b] Wei Chen, Yiqing Wu, Zhao Zhang, Fuzhen Zhuang, Zhongshi He, Ruobing Xie, and Feng Xia. Fairgap: Fairness-aware recommendation via generating counterfactual graph. *ACM Trans. Inf. Syst.*, 42(4), February 2024.

[Dai and Wang, 2021] 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*, WSDM '21, page 680–688, New York, NY, USA, 2021. Association for Computing Machinery.

[Dong *et al.*, 2023] Yushun Dong, Jing Ma, Song Wang, Chen Chen, and Jundong Li. Fairness in graph mining: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(10):10583–10602, 2023.

[Grover and Leskovec, 2016] Aditya Grover and Jure Leskovec. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 855–864, New York, NY, USA, 2016. Association for Computing Machinery.

[Guerreiro *et al.*, 2021] Andreia P. Guerreiro, Carlos M. Fonseca, and Luís Paquete. The hypervolume indicator: Computational problems and algorithms. *ACM Comput. Surv.*, 54(6), July 2021.

[Harper and Konstan, 2015] F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Trans. Interact. Intell. Syst.*, 5(4), December 2015.

[He *et al.*, 2020] Xiangnan He, Kuan Deng, Xiang Wang, Yan Li, YongDong Zhang, and Meng Wang. Lightgcn: Simplifying and powering graph convolution network for recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '20, page 639–648, New York, NY, USA, 2020. Association for Computing Machinery.

[Kang *et al.*, 2022] Jian Kang, Yan Zhu, Yinglong Xia, Jiebo Luo, and Hanghang Tong. Rawlsgcn: Towards rawlsian difference principle on graph convolutional network. In *Proceedings of the ACM Web Conference 2022*, WWW '22, page 1214–1225, New York, NY, USA, 2022. Association for Computing Machinery.

[Khajehnejad *et al.*, 2020] Moein Khajehnejad, Ahmad Asgharian Rezaei, Mahmoudreza Babaei, Jessica Hoffmann, Mahdi Jalili, and Adrian Weller. Adversarial graph embeddings for fair influence maximization over social networks, 2020.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

[Li *et al.*, 2024a] Chengyu Li, Debo Cheng, Guixian Zhang, and Shichao Zhang. Contrastive learning for fair graph representations via counterfactual graph augmentation. *Knowledge-Based Systems*, 305:112635, 2024.

[Li *et al.*, 2024b] Yicong Li, Yu Yang, Jiannong Cao, Shuaiqi Liu, Haoran Tang, and Guandong Xu. Toward structure fairness in dynamic graph embedding: A trend-aware dual debiasing approach. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, page 1701–1712, New York, NY, USA, 2024. Association for Computing Machinery.

[Ling *et al.*, 2023] Hongyi Ling, Zhimeng Jiang, Youzhi Luo, Shuiwang Ji, and Na Zou. Learning fair graph representations via automated data augmentations. In *International Conference on Learning Representations (ICLR)*, 2023.

[Liu *et al.*, 2023] Zemin Liu, Trung-Kien Nguyen, and Yuan Fang. On generalized degree fairness in graph neural networks. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(4):4525–4533, Jun. 2023.

[Luo *et al.*, 2025a] Renqiang Luo, Ziqi Xu, Xikun Zhang, Qing Qing, Huafei Huang, Enyan Dai, Zhe Wang, and Bo Yang. Fairness in graph learning augmented with machine learning: A survey, 2025.

[Luo *et al.*, 2025b] Zihan Luo, Hong Huang, Jianxun Lian, Xiran Song, and Hai Jin. Towards controllable hybrid fairness in graph neural networks. In *Proceedings of the 31st*

*ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, KDD '25, page 950–961, New York, NY, USA, 2025. Association for Computing Machinery.

[Mansoury *et al.*, 2020] Masoud Mansoury, Himan Abdollahpouri, Jessie Smith, Arman Dehpanah, Mykola Pechenizkiy, and Bamshad Mobasher. Investigating potential factors associated with gender discrimination in collaborative recommender systems, 2020.

[Pham *et al.*, 2025] Phu Pham, Quang-Thinh Bui, Ngoc Thanh Nguyen, Robert Kozma, Philip S. Yu, and Bay Vo. Topological data analysis in graph neural networks: Surveys and perspectives. *IEEE Transactions on Neural Networks and Learning Systems*, 36(6):9758–9776, 2025.

[Rahman *et al.*, 2019] Tahleen Rahman, Bartlomiej Surma, Michael Backes, and Yang Zhang. Fairwalk: Towards fair graph embedding. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2019.

[Song *et al.*, 2025] Zixing Song, Muzhi Li, Yifei Zhang, Irwin King, and José Miguel Hernández-Lobato. Track and tweak: Monitoring and improving group fairness for temporal graph neural networks in real time. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*, KDD '25, page 2644–2655, New York, NY, USA, 2025. Association for Computing Machinery.

[Wan *et al.*, 2020] Mengting Wan, Jianmo Ni, Rishabh Misra, and Julian McAuley. Addressing marketing bias in product recommendations. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, WSDM '20, page 618–626, New York, NY, USA, 2020. Association for Computing Machinery.

[Wang *et al.*, 2022] Ruijia Wang, Xiao Wang, Chuan Shi, and Le Song. Uncovering the structural fairness in graph contrastive learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 32465–32473. Curran Associates, Inc., 2022.

[Yang *et al.*, 2024] Cheng Yang, Jixi Liu, Yunhe Yan, and Chuan Shi. Fairsin: Achieving fairness in graph neural networks through sensitive information neutralization. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(8):9241–9249, Mar. 2024.

[Zhang *et al.*, 2024] An Zhang, Wenchang Ma, Pengbo Wei, Leheng Sheng, and Xiang Wang. General debiasing for graph-based collaborative filtering via adversarial graph dropout. In *Proceedings of the ACM Web Conference 2024*, WWW '24, page 3864–3875, New York, NY, USA, 2024. Association for Computing Machinery.

[Zhang *et al.*, 2025] Guixian Zhang, Guan Yuan, Debo Cheng, Lin Liu, Jiuyong Li, and Shichao Zhang. Disentangled contrastive learning for fair graph representations. *Neural Networks*, 181:106781, 2025.

[Zhu *et al.*, 2024a] Yuchang Zhu, Jintang Li, Liang Chen, and Zibin Zheng. The devil is in the data: Learning fair graph neural networks via partial knowledge distillation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, WSDM '24, page 1012–1021, New York, NY, USA, 2024. Association for Computing Machinery.

[Zhu *et al.*, 2024b] Yuchang Zhu, Jintang Li, Zibin Zheng, and Liang Chen. Fair graph representation learning via sensitive attribute disentanglement. In *Proceedings of the ACM Web Conference 2024*, WWW '24, page 1182–1192, New York, NY, USA, 2024. Association for Computing Machinery.