

Understanding Fairness–Privacy Interactions in Graph Neural Networks: A Literature Review, Taxonomy, and Open Challenges

Jingyan Zhang, Irina Tal

Dublin City University
School of Computing
Dublin, Ireland
jingyan.zhang29@mail.dcu.ie, irina.tal@dcu.ie

Abstract

Graph Neural Networks (GNNs) are increasingly applied to socially consequential tasks, such as disease surveillance on contact graphs, and credit scoring and fraud detection on financial transaction networks. In these settings, biased or unfair predictions can restrict access to opportunities and resources, while privacy leakage can expose sensitive information, leading to substantial individual and societal harms. Consequently, fairness and privacy have become central concerns for trustworthy GNN deployment. Fairness-aware GNNs seek to reduce biased predictions across sensitive entities, whereas privacy-preserving GNNs aim to protect private graph information. However, these objectives are often interdependent: fairness interventions can affect privacy outcomes, while privacy mechanisms may impose additional constraints on fairness. This paper systematically reviews 19 studies to investigate the interplay between fairness and privacy in GNNs. We propose a two-dimensional orthogonal taxonomy that organizes existing work according to fairness objective and privacy target, and use it to analyze dominant research patterns as well as the synergies and trade-offs between fairness and privacy. To the best of our knowledge, this is the first systematic review to provide a structured analysis of fairness–privacy interactions in GNNs and to identify open challenges and future research directions for trustworthy graph learning systems.

1 Introduction

Graph-structured data are pervasive in real-world scenarios, ranging from social networks (Yin et al. 2026), transportation systems (Ali et al. 2025; Jin et al. 2024), to financial transaction networks (Cheng et al. 2025). Graph Neural Networks (GNNs) have become a widely adopted learning paradigm for such data because they can learn representations by combining node attributes with relational structure, enabling models to capture dependencies, similarities, and influence patterns that are difficult to represent with independent feature-based learning. As a result, GNNs are increasingly applied to critical and sensitive domains involving complex relationships, such as disease surveillance on contact or mobility graphs (Xue et al. 2022), credit scoring and fraud detection on financial transaction networks (Cheng et al. 2025; Feng et al. 2022), and predictive policing on

criminal records (Wang et al. 2022). These tasks are socially consequential because their predictions can influence vital access to healthcare resources, financial opportunities, public support, or institutional decisions.

At the same time, the deployment of GNNs in such domains raises significant fairness and privacy concerns. Fairness-aware GNNs seek to reduce biased predictions across sensitive entities, while privacy-preserving GNNs aim to protect private graph information. In terms of fairness, the message-passing mechanism of GNNs can propagate historical biases and amplify disparities through neighborhood aggregation, particularly when the graph structure is correlated with sensitive attributes such as gender or income (Chen et al. 2024; Zhang et al. 2024a). From a privacy perspective, learned embeddings and model outputs can leak sensitive information, including private attributes and membership information (Zhang et al. 2024a; Liu et al. 2025; Zhang and Tal 2025). These concerns are especially intertwined in GNNs, as message passing enables both bias and private information to propagate and intensify through the graph structure (Dai et al. 2024). Therefore, understanding how fairness and privacy interact in GNNs is essential to develop socially responsible graph learning systems.

Although fairness-aware and privacy-preserving GNNs have both received increasing attention, their interaction remains insufficiently understood. Fairness interventions may require access to sensitive attributes or group-level statistics, potentially increasing privacy risks. In contrast, privacy-preserving mechanisms may alter graph structure, obscure node attributes, or introduce noise into model training, thereby affecting fairness outcomes. While existing surveys have provided valuable overviews of fairness-aware and privacy-preserving GNNs separately (Dai et al. 2024; Chen et al. 2024; Duricic et al. 2023; Guo et al. 2025; Ge et al. 2024; Zhang et al. 2024a,b; Liu et al. 2025; Zhang and Tal 2024), these topics are often examined independently or discussed primarily in relation to utility trade-offs. As a result, the literature lacks a structured understanding of how fairness and privacy objectives interact, conflict, or constrain each other in graph learning.

Motivated by the concerns discussed above, we conduct a comprehensive literature review of fairness–privacy interactions in GNNs. The main contributions of this review are as follows:

- We provide the first focused review of fairness–privacy interactions in GNNs, covering 19 studies that explicitly examine fairness, privacy, and their relationships.
- We propose a two-dimensional orthogonal taxonomy that organizes the existing body of work according to fairness objectives and privacy targets.
- We conduct a cross-dimensional analysis to identify dominant research patterns, underexplored intersections, and fairness–privacy trade-offs in the current GNN literature.
- We highlight open challenges and future research directions for developing socially responsible, fair, and privacy-preserving GNN systems.

The structure of this paper is organized as follows. Section 2 reviews the relevant background on existing fair-aware and privacy-preserving GNNs. Section 3 describes the methodology employed in conducting this systematic review. In Section 4, we propose a two-dimensional orthogonal taxonomy that organizes existing work according to fairness objectives and privacy targets, while Section 5 analyzes dominant research patterns as well as synergies and trade-offs between fairness and privacy. Section 6 discusses the identified open challenges and proposes future research directions. Finally, Section 7 concludes the paper with a summary of our findings and their implications for future development.

2 Background

2.1 Graph Neural Networks and Graph Learning

Graph Neural Networks (GNNs) have become the dominant framework for representation learning on graph-structured data. GNNs learn node, edge, or graph representations through iterative message passing and neighborhood aggregation. Graph Convolutional Networks (GCNs) (Kipf and Welling 2016), for instance, draw on ideas from spectral graph theory and approximate graph convolutions by aggregating information from local neighborhoods. Subsequent models further improved the expressiveness and flexibility of this aggregation process. GraphSAGE (Hamilton, Ying, and Leskovec 2017) introduced an inductive learning framework that samples and aggregates features from local neighborhoods, enabling generalization to previously unseen nodes. Graph Attention Networks (GATs) (Veličković et al. 2017) incorporated attention mechanisms to assign adaptive importance weights to neighboring nodes during aggregation. Graph Transformer models, such as Graphormer (Ying et al. 2021) and GraphGPS (Rampásek et al. 2022), combine graph structural encodings with global attention mechanisms, enabling richer modeling of long-range interactions beyond local neighborhoods.

2.2 Fairness-Aware GNNs

Unlike conventional IID learning settings, fairness issues in graphs arise not only from biased node attributes or labels, but also from graph topology, homophily, neighborhood aggregation, and the propagation of information through message passing (Mehrabi et al. 2021). As a result, sensitive

attributes such as gender, race, age, nationality, or socioeconomic status may influence learned representations even when such attributes are not explicitly used as input features. Recent surveys therefore characterize fairness-aware GNNs as a distinct research area concerned with identifying, measuring, and mitigating unfair outcomes at the graph, neighborhood, embedding, and prediction levels (Dai et al. 2024; Duricic et al. 2023). Recent studies address fairness directly within GNN models. FairGNN (Dai and Wang 2021) is a representative in-processing method that learns fair node representations under limited sensitive attribute annotations. It combines sensitive attribute estimation, adversarial debiasing, and GNN-based prediction to reduce discrimination while maintaining classification utility. Other approaches consider fairness from a counterfactual (Guo et al. 2025) or stability-oriented perspective. These methods aim to ensure that learned embeddings and predictions are less sensitive to perturbations of protected attributes or their graph-induced proxies.

2.3 Privacy-Preserving GNNs

The message-passing mechanism introduces distinctive privacy risks: node attributes, links, labels, and even membership information may be inferred from trained models, embeddings, gradients, or prediction outputs. Recent surveys categorize privacy-preserving GNN methods according to both the attack surface and the defense mechanism, including approaches based on differential privacy (DP) (Abadi et al. 2016), federated learning (FL) (McMahan et al. 2017), homomorphic encryption (HE) (Lou et al. 2020), and adversarial defense (Zhang and Tal 2025).

Among these strategies, DP is one of the most widely adopted mechanisms because it provides formal privacy guarantees by bounding the influence of any individual data record on the released output (Sajadmanesh and Gatica-Perez 2021; Zhang et al. 2021). In the context of GNNs, DP can be applied at different stages of the learning pipeline, including perturbing node features, graph structure, gradients, labels, embeddings, or model parameters. These methods aim to reduce the risk of membership inference, attribute inference, and model inversion attacks, although they often introduce a trade-off between privacy strength and predictive utility (Dai et al. 2024).

FL (Rocco di Torrepadula et al. 2025; Zhang et al. 2025; McMahan et al. 2017; Kairouz et al. 2021) offers a complementary privacy-preserving paradigm by enabling multiple clients to collaboratively train GNN models without directly sharing raw graph data. This is particularly useful when graph data are distributed across users, institutions, or platforms. However, federated GNNs remain vulnerable to information leakage through gradients, embeddings, or exchanged neighborhood representations (Dai et al. 2024). Therefore, FL is often combined with DP, secure aggregation, or cryptographic mechanisms to strengthen privacy during collaborative training (Zhang and Tal 2024).

3 Methodology

Our structured approach to conducting a literature survey in this paper draws inspiration from (Kitchenham 2004) and

(Pai, McCulloch, and Colford 2004), adhering to the broadly recognized processes for survey research. Our methodology unfolds across three phases: 1) survey planning, 2) conducting systematic review and reporting results, and 3) exploring research challenges existing in the field.

The subsequent portions of this section are dedicated to elaborating on the survey planning phase, including our research questions, and delineating the procedure for identifying relevant studies. This structured approach ensures a comprehensive systematic review of the existing literature, allowing for a thorough understanding of fairness-privacy trade-offs in GNNs, and identifying gaps and challenges for future exploration.

3.1 Research Questions

For this survey, we have identified several critical research questions to guide our exploration into the interplay of fairness and privacy in GNNs:

- **Q1:** How do fairness and privacy objectives interact in graph neural networks?
- **Q2:** How can existing fairness-privacy GNN studies be categorized according to fairness notion and privacy target?
- **Q3:** What trade-offs emerge among fairness, privacy, and utility across different GNN learning settings?
- **Q4:** How do privacy-preserving mechanisms, such as differential privacy, data perturbation, and federated learning, affect fairness outcomes in GNNs?
- **Q5:** And conversely, how do fairness interventions influence privacy?
- **Q6:** What limitations and future directions can be identified from the current literature?

Addressing these questions will facilitate a structured understanding of fairness-privacy interactions in GNNs, highlighting both open challenges and future research directions for trustworthy graph learning development.

3.2 Search Strategy

The initial phase of the study involves the development of a carefully constructed search query, centering on the theme “Fairness, Privacy, and Their Relationships in GNNs.” This search query encompasses three parallel elements: GNNs, Fairness, and Privacy. The research database employed for this study was the Web of Science. The search was conducted within the metadata fields (author, title, source, and abstract) and among specific publishers including IEEE, Springer Nature, Assoc Computer Machinery, Elsevier, Mdpi, Wiley, Sage, and Taylor & Francis. The paper collection was restricted to the period up to May of 2026.

The search string is as follows:

```
TS =(fairness OR fair) AND
TS = (privacy OR private) AND
TS = (graph OR GNN OR "graph
neural network")
```

The initial search returned 121 records. These records were first screened to remove duplicates and studies that were clearly outside the scope of the review. Following this preliminary screening, 25 papers were retained for closer examination.

In the subsequent phase, we additionally applied backward snowballing to the retained papers by examining their reference lists. This process identified 4 additional candidate studies. The candidate set was then assessed against the inclusion and exclusion criteria described in the following. Therefore, a set of 19 studies was established for systematic review within the defined scope of our research.

Inclusion Criteria A study was included if it met all the following criteria:

- It focuses on GNNs, graph representation learning, or a related graph-learning setting.
- Fairness is explicitly considered as an objective, constraint, evaluation criterion, or analytical concern.
- Privacy is explicitly considered as an objective, threat, protection mechanism, data-access constraint, or analytical concern.
- It either explicitly examines fairness-privacy interactions or provides sufficient evidence to analyze how one objective is affected by the other, even if this relationship is not explicitly discussed in the original study.

Exclusion Criteria A study was excluded if it met any of the following conditions:

- It considers fairness without substantively addressing privacy.
- It considers privacy without substantively addressing fairness.
- It mentions both fairness and privacy but provides no meaningful basis or implications for analyzing their interactions.
- It falls outside the scope of GNNs or related graph-learning methods.

4 Proposed Taxonomy

We propose a two-dimensional taxonomy for organizing fairness-privacy trade-offs in GNNs. The taxonomy is derived from a focused review of 19 papers that explicitly investigate fairness, privacy, and their potential interaction in GNN-based learning. Rather than treating fairness, privacy, and model design as a single flat classification, we organize the literature along two orthogonal dimensions: fairness objective, and privacy target. The fairness dimension distinguishes between group fairness, which aims to reduce performance disparities across sensitive groups; individual fairness, which seeks to ensure that similar nodes or users receive similar outcomes; cross-client fairness, which treats federated clients as fairness subjects and aims to reduce disparities in client-level performance, utility, or benefit; and client-aware fairness, which leverages client-level signals to guide fair federated aggregation. The privacy dimension captures the type of information that methods aim to protect, including node attribute privacy, link privacy, and local

data privacy. This taxonomy provides a structured lens for comparing existing methods according to what fairness notion they optimize, what privacy risk they address, and the interaction patterns underlying these objectives.

4.1 Fairness Objectives

Group Fairness Recent studies on group fairness in graph neural networks converge on the observation that graph topology and message passing can amplify demographic disparities, yet they address this issue at different stages of the GNN lifecycle. One study (Dai and Wang 2023) proposes FairGNN and NT-FairGNN for fair node classification under limited and private sensitive-attribute information, using a GNN-based sensitive attribute estimator, adversarial debiasing, and fairness regularization to reduce Statistical Parity and Equal Opportunity gaps while preserving predictive accuracy. In contrast, MAPPING (Song and Palanisamy 2024) adopts a model-agnostic pre-processing perspective: it jointly debiases node features and graph topology through masking, reconstruction, fair message passing, pruning, distance-covariance-based constraints, and adversarial training, thereby extending group fairness to multiple sensitive attributes and reducing sensitive-attribute inference risks. The study of FROG (Chen et al. 2025) works on a complementary setting, fair graph unlearning, and shows that edge/node deletion requests may increase sensitive homophily and degrade dyadic fairness. To address this, FROG jointly performs fair edge augmentation, sparse structure pruning, and model updating so that forgetting can be achieved without exacerbating group disparity. ReCUR (Yang and Li 2026) removes the influence of selected user-item interactions through contrastive unlearning and influence estimation, and then applies fairness-aware re-ranking with group-specific promotion weights to compensate for long-tail or under-represented user/item segments.

These methods share a common motivation—mitigating bias induced or amplified by graph structure—but differ substantially in their intervention points, with FairGNN operating in-processing, MAPPING operating as pre-processing, and FROG operating during unlearning; they also differ in their fairness definitions, privacy assumptions, and task settings, covering node classification, multiple-sensitive-attribute debiasing, and link-level fair unlearning, respectively.

It is worth noting that group fairness is frequently adopted in federated settings. Two studies (Agrawal et al. 2024; Wang et al. 2024) treat group fairness primarily as a local data fairness constraint rather than proposing FL-specific fairness notions. Specifically, both define group fairness in terms of disparities in average recommendation performance across user groups and incorporate fairness objectives into model training, thereby jointly optimizing recommendation utility and inter-group performance parity in federated GNN recommendation systems.

Furthermore, several studies (Tan et al. 2026; Cui et al. 2023; Wang, Yan, and Jin 2026; Lo et al. 2025) employ group fairness metrics to support cross-client fairness and client-aware fairness objectives in FL-specific applications.

These approaches will be discussed in detail in later sections.

Individual Fairness Unlike group fairness, individual fairness in graph learning refers to the expectation that similar nodes should receive similar predictions, thereby emphasizing node-level fairness. The reviewed studies consistently indicate that the main challenge of individual fairness lies in defining and measuring similarity between nodes.

Zhang, Yuan, and Pan 2024 constructs a Jaccard similarity matrix based on graph structure rather than node attributes, thereby reducing the influence of sensitive attributes on fairness optimization. In contrast, Wang et al. 2025a, 2026, 2023 adopt ranking-based individual fairness instead of similarity-matrix-based fairness. Ranking-based methods focus only on the top-k most similar nodes according to similarity rankings, rather than considering all possible node pairs. This approach reduces computational cost while concentrating fairness optimization on the most relevant node relationships. In addition, Wang et al. 2025a utilizes group fairness to assist individual fairness calculations.

Notably, Wang et al. 2026 theoretically demonstrates that distance-based individual fairness may compromise privacy. Enforcing similar nodes to remain close in the embedding space can unintentionally cluster nodes sharing the same sensitive attributes, thereby facilitating attribute inference attacks. Ranking-based fairness mitigates this issue by preserving relative similarity orderings without enforcing absolute proximity in the embedding space. Furthermore, PCFGR (Wang et al. 2025b) utilizes counterfactual fairness. Using a top-k ranking approach, the method constructs contextual subgraphs for central nodes and performs counterfactual augmentation by flipping the central node's sensitive attribute and randomly perturbing neighboring sensitive attributes.

Overall, existing work on individual fairness is shifting from similarity matrices based on absolute distances to ranking-based methods that capture relative distances, thereby enabling more locally focused similarity representations and reducing the risk of link leakage.

Cross-Client Fairness Cross-client fairness, also referred to as client-level fairness, collaborative fairness, or participation fairness, is an important fairness concern in federated settings. It generally refers to the goal of reducing performance disparities across participating clients and ensuring that the benefits and costs of collaborative model training are distributed equitably. Their common strategy is to mitigate client-level performance disparity by explicitly modeling inter-client heterogeneity, either through fairness regularization, personalized aggregation, or adaptive reweighting of client updates.

Recent studies on fairness-aware federated graph learning reveal that cross-client fairness in GNNs is primarily challenged by graph-specific heterogeneity, including non-IID topology, imbalanced subgraph overlap, unequal structural informativeness, and heterogeneous client contributions. Although all existing work aims to mitigate unfairness under federated or decentralized graph learning, they differ substantially in how cross-client fairness is defined and oper-

ated.

FairGFL (Zhou et al. 2026) attributes unfairness to imbalanced overlapping subgraphs distributed across clients, estimates node/link overlap ratios via overlap-aware weighted aggregation, and apply a composite-loss regularizer to improve the fairness–utility trade-off. Tan et al. 2026 target macro-level client-graph restructuring, where clients are represented as nodes in a macro graph built from privacy-preserving structural statistics, and underperforming clients are upweighted through graph-constrained aggregation with macro embeddings. FedGraph-Fair (Khan 2026) further extends this client-graph perspective by combining personalization, dynamic top-(k) client communication graphs, and primal–dual optimization to control worst-case client loss and group-risk parity, thereby emphasizing tail-client protection rather than only average client consistency. Pan et al. 2024 reinterpret fairness as contribution-aware fairness among self-interested graph federation agents, allocating gradient quality and payoff according to agent value estimated from gradient alignment and graph diversity, while using motif prototypes to preserve the accuracy–fairness trade-off under graph heterogeneity. In addition, Zhou et al. 2025 studies a domain-specific form of cross-agent fairness in topology-aware federated reinforcement learning, where GNN encoders model vehicle charging networks, and fairness is achieved through reward shaping and fairness-aware aggregation, evaluated by Jain’s fairness index for equitable scheduling and resource allocation.

These works show that vanilla federated averaging (FedAvg) style aggregation is insufficient for federated fair-aware GNNs because graph topology makes client disadvantage structurally persistent. Therefore, effective cross-client fairness in federated GNNs requires explicitly modeling client relations or client disadvantage, while maintaining utility and privacy through specific mechanisms.

Client-Aware Fairness We identify a distinct fairness-related objective from the reviewed literature that can be characterized as client-aware fairness. The fairness-aware aggregation strategies of this objective uses client-level information to improve local and global group fairness in federated GNN optimization. Specifically, these methods incorporate fairness into federated optimization by reweighting local updates or representations during aggregation according to client-level fairness indicators, graph structural bias, contribution estimates, or accuracy–fairness trade-off measures.

Client-aware fairness differs from cross-client fairness in two key aspects. First, its fairness target is not being fair to the clients themselves, but the group fairness, either within local client models, in the global model, or both. Second, federated aggregation is used as a mechanism to suppress the propagation of biased local updates and improve group fairness in the learned model, rather than to equalize performance, utility, or loss across clients.

F2GNN (Cui et al. 2023) introduces a dual-stage framework in which clients optimize local GNNs using fairness penalties, while the server aggregates local models with weights derived from fairness metrics and graph balance

scores, thereby preventing biased clients from disproportionately influencing the global model.

FairWAG (Lo et al. 2025) extends this approach through more sophisticated aggregation strategies. Specifically, it assigns adaptive client-level weights based on structural bias and Shapley-value estimates of contributions to both accuracy and fairness. In addition, it introduces neuron-level weighting to reduce sensitivity to protected attributes during server aggregation.

Similarly, FaVGNN (Wang, Yan, and Jin 2026) computes server aggregation weights using client-level comprehensive metrics. However, unlike F2GNN and FairWAG, FaVGNN adopts a vertical federated learning setting, where clients possess disjoint feature spaces for the same entities. Its primary objective is to address incomplete sensitive attributes through local completion-driven adversarial fusion and vertically weighted aggregation.

Introducing client-aware fairness as a separate classification clarifies an intermediate category between conventional group fairness and cross-client fairness. These methods are more federated-specific than ordinary centralized fairness regularization, since the fairness intervention occurs through client update selection, weighting, or aggregation under privacy constraints.

This distinction helps explain how federated learning changes the design space of fair GNNs: privacy restrictions prevent direct access to raw graph data and sensitive attributes, so fairness must often be optimized indirectly through aggregated client-level statistics, local fairness indicators, or privacy-preserving representations. Therefore, fairness-aware aggregation provides a useful conceptual lens for understanding how fairness objectives are adapted to federated and privacy-constrained graph learning settings.

4.2 Privacy Objectives

Node Level Across the reviewed studies, node-level privacy in GNNs is treated as a coupled fairness–privacy–utility problem, but the protected node information and the privacy mechanisms differ substantially. A common observation is that message passing can encode sensitive node attributes into features, topology-derived signals, and learned embeddings, thereby enabling attribute inference, membership inference, or residual-data attacks even when sensitive attributes are missing or partially disclosed. NT-FairGNN (Dai and Wang 2023) addresses privacy mainly at the sensitive-attribute level, which exploits a sensitive-attribute estimator to compensate for limited observed attributes, while the privacy-preserving variants perturb sensitive attributes with LDP and use noise-tolerant learning to recover useful fairness signals for adversarial or counterfactual debiasing. LPF-IFGNN (Wang et al. 2023) shifts the protected object from sensitive attributes to node features and labels, applying LDP directly on local node data and then using multi-hop aggregation to denoise perturbed features before enforcing ranking-based individual fairness. MAPPING (Song and Palanisamy 2024) and FPGNN (Wang et al. 2026) instead focus more explicitly on preventing sensitive information leakage from learned representations. MAPPING (Song and Palanisamy 2024)

performs model-agnostic pre-processing by jointly debiasing node features and topology with dCov constraints and adversarial training, targeting multiple sensitive-attribute inference but without formal DP guarantees, whereas FPGNN performs in-processing representation sanitization, using differentiable ranking-based individual fairness together with adversarial removal of sensitive information from embeddings, and further handles LDP-protected sensitive attributes via forward correction. F2PGNN (Agrawal et al. 2024) considers a federated recommendation setting in which node-level user privacy is protected by keeping raw user data local and adding LDP to model updates and group statistics, thereby differing from centralized fair-GNN methods that directly operate on the full graph. Finally, FROG (Chen et al. 2025) and ReCUR (Yang and Li 2026) formulate privacy as graph unlearning rather than attribute hiding: FROG removes node/edge influence while rewiring the graph to preserve fairness and avoid embedding distortion caused by deletion-induced homophily shifts, whereas ReCUR uses influence estimation and contrastive push-pull objectives to erase user-item interactions or nodes from recommender embeddings, validating forgetting through membership-inference audits and ranking-based leakage analysis. Thus, while all eight papers aim to reduce leakage of node-level information under fairness constraints, they instantiate privacy at different layers, including raw node features, sensitive attributes, learned embeddings, federated updates, graph topology, and memorized interactions.

Edge Level Edge-level privacy is treated as the protection or removal of relational information, but the threat models and mechanisms differ. The study of PPFR (Zhang, Yuan, and Pan 2024) explicitly examines the interaction between node-level individual fairness and edge-existence leakage, showing theoretically and empirically that improving individual fairness can increase link-stealing risk. Their PPFR framework mitigates this tension through fairness-aware loss reweighting and privacy-aware graph-structure perturbation in a post-training fine-tuning stage, thereby targeting hidden-edge inference while controlling utility loss. While GRIF-PPGNN (Wang et al. 2025a) extends this line by protecting edge privacy through differentially private perturbed aggregations, so that downstream training can proceed without directly exposing graph structure. In contrast, FROG (Chen et al. 2025) and ReCUR (Yang and Li 2026) frame edge privacy mainly through graph unlearning: FROG removes requested edges or nodes so that the updated model behaves as if trained on the retained graph, and further rewires the graph by deleting redundant edges and augmenting fairness-preserving ones to avoid disparate impact induced by edge deletion. ReCUR specializes this unlearning perspective for bipartite recommendation graphs, where user-item interactions are edges; it combines influence estimation with contrastive push-pull unlearning to erase forgotten interactions while preserving retained interaction structure, and validates privacy empirically through membership-inference audits showing near-retrain behavior. Thus, PPFR (Zhang, Yuan, and Pan 2024) and GRIF-PPGNN (Wang et al. 2025a) primarily defend

against edge-existence inference from GNN outputs or aggregations, whereas FROG (Chen et al. 2025) and ReCUR (Yang and Li 2026) focus on erasing the influence of specified edges after deletion requests.

Local Data Level Local data privacy or client-level privacy in federated GNNs is primarily achieved by data locality, i.e., keeping raw client graphs, node/edge features, sensitive attributes on local clients and communicating only model updates, embeddings, prototypes, or aggregated statistics. F2GNN (Cui et al. 2023), FPFR (Wang et al. 2024), (Pan et al. 2024), (Zhou et al. 2025) and FairWAG (Lo et al. 2025) mainly inherit the standard FL privacy assumption: clients retain local graph data and upload parameters, gradients, fairness metrics, or contribution signals, thereby avoiding direct exposure of raw graph structures and client records. In contrast, FairGFL (Zhou et al. 2026) provides a stronger local-data privacy mechanism by explicitly perturbing sampled subgraph information with local differential privacy: node features are encoded and obfuscated, and links are randomized before being used to estimate cross-client overlap ratios, which directly targets leakage from overlapping nodes and edges. Similarly, (Wang et al. 2026) constructs a client-level macro graph from privacy-preserving structural statistics and applies a Gaussian mechanism to differentially private macro embeddings, thereby protecting topology-derived client representations while enabling structure-aware aggregation. FedGraph-Fair (Khan 2026) further formalizes privacy compatibility through client-level differential privacy and secure aggregation, analyzing how DP noise affects convergence and fairness in a dynamic client graph. FaVGNN (Wang, Yan, and Jin 2026) differs from the horizontal-FL papers by considering vertical federated GNNs.

Overall, these works share the common objective of preventing direct raw-data transfer in federated graph learning, but they differ in privacy granularity: some protect raw data by protocol design through FL data locality, some protect graph structure and client relationships through perturbed structural summaries or macro embeddings, and only a subset provides formal privacy semantics via LDP, DP, or secure aggregation. Thus, the literature suggests a progression from implicit privacy preservation by decentralized training toward explicit protection of graph topology, overlap information, sensitive attributes, and client-level statistics.

4.3 Taxonomy-Based Mapping of the Papers

After defining the two taxonomy dimensions, we map each reviewed paper to the corresponding categories according to its primary fairness objective and privacy target, as shown in Table 1. Specifically, we identify whether a study focuses on group fairness, individual fairness, cross-client fairness, or client-aware fairness; and whether it protects node privacy, edge privacy, or local data privacy. When a study addresses multiple categories, we assign multiple labels where appropriate and use the primary technical contribution for summary-level analysis. The resulting mapping serves two purposes: it makes the classification of the literature transparent and it highlights dominant patterns and underex-

explored intersections in current research that will be discussed in next section.

5 Taxonomy-Guided Literature Discussion

5.1 Group Fairness Under Node Attribute Privacy Constraints

Group fairness under node attribute privacy constraints exposes a fundamental tension in fair graph learning: sensitive attributes are often necessary for estimating and correcting demographic disparities, yet their collection and use may itself increase node-level privacy risk. Dai and Wang 2023 formalize this issue by showing that fair GNN training with limited sensitive attributes is vulnerable to privacy leakage when ground-truth sensitive attributes are directly used; NT-FairGNN addresses this by perturbing sensitive attributes under local differential privacy and then learning fairness constraints from noisy sensitive estimates, thereby treating privacy as a constraint on fairness supervision rather than as a post-hoc protection mechanism.

Song and Palanisamy 2024 further broaden this perspective by arguing that sensitive information is not only contained in explicit node attributes but also recoverable from non-sensitive features and graph topology; their MAPPING framework therefore jointly debiases features and topology with distance-covariance constraints and adversarial training, aiming to improve group fairness while reducing multiple sensitive-attribute inference leakage.

In contrast, FROG (Chen et al. 2025) highlights a complementary privacy-driven setting—graph unlearning—where removing user-requested nodes or edges can satisfy privacy-oriented deletion requirements but may unintentionally alter homophily and worsen dyadic group fairness; FROG mitigates this effect through fair graph rewiring and model updating. Taken together, these works suggest that the fairness–privacy relationship in GNNs is not a simple zero-sum trade-off: privacy mechanisms such as local differential privacy may degrade the quality of sensitive supervision, while fairness interventions may either expose or suppress sensitive information depending on whether they account for topology-induced leakage. A key implication is that trustworthy GNNs should jointly optimize utility, group fairness, and node-level privacy, explicitly modeling sensitive attributes as both fairness-critical signals and privacy-sensitive information embedded in node features, graph structure, and learned representations.

5.2 Individual Fairness Under Link Privacy Constraints

Individual fairness under edge-level privacy constraints reveals an inherent tension in fairness-privacy GNN design: enforcing similar predictions for similar nodes reduces individual-level discrimination, but it can also amplify the distinguishability exploited by link-stealing attacks. Zhang, Yuan, and Pan 2024 theoretically and empirically show that improving Lipschitz-based individual fairness tends to decrease prediction distances among structurally similar or neighboring nodes, thereby increasing the separability between connected and unconnected node-pair

distributions and raising edge leakage risk; they address this via PPFR, which combines fairness-aware loss reweighting with privacy-aware graph perturbation to jointly reduce bias and edge inference risk with limited accuracy degradation.

Wang et al. 2025a extend this perspective to ranking-based individual fairness, observing that fairness promotion can similarly make posterior representations of linked or similar nodes more alike, exacerbating edge-stealing attacks, while also introducing group disparity in fairness gains. Their GRIF-PPGNN framework therefore adds differentially private noise to graph aggregations, avoids direct graph-structure querying during inference, and optimizes both ranking-based individual fairness and equality of fairness promotion across sensitive groups. Taken together, these studies suggest that individual fairness in GNNs should not be optimized as an isolated objective; rather, fair GNNs must be designed under explicit edge-privacy constraints, where the optimization objective jointly controls prediction smoothness, structural leakage, subgroup-level fairness disparity, and utility.

5.3 Fairness Degradation Induced by Differential Privacy Noise

Differential privacy introduces a non-negligible source of fairness degradation in graph neural networks because the injected noise perturbs precisely the signals used to define or enforce fairness. Existing studies show that the degradation mechanism depends on the privatized object: LDP-protected sensitive attributes corrupt the supervision required by group- or counterfactual-fair training, LDP-protected node features distort the similarity matrix used for ranking-based individual fairness, and edge-DP perturbations alter the graph topology that governs both message passing and fairness regularization. For example, NT-FairGNN (Dai and Wang 2023) and PCFGR (Wang et al. 2025b) demonstrate that randomized-response noise on sensitive attributes weakens fairness constraints unless noise-aware estimation or forward correction is introduced, while LPF-IFGNN (Wang et al. 2023) shows that feature-level LDP can impair individual-fairness estimation but can be partially mitigated through neighborhood aggregation under homophily. In contrast, PPFR (Zhang, Yuan, and Pan 2024) and GRIF-PPGNN (Wang et al. 2025a) highlight a structural privacy–fairness tension: enforcing individual fairness may increase edge leakage risk, whereas DP-based edge perturbation can reduce privacy risk at the cost of degraded fairness or predictive accuracy. Overall, these works suggest that DP noise should not be treated as a fairness-neutral privacy layer in GNNs; rather, fair privacy-preserving GNNs require explicit noise-aware fairness objectives, denoising mechanisms, or robust estimators to prevent privacy protection from systematically degrading fairness.

5.4 Fairness Under Federated Privacy Constraints

A key observation from the reviewed fair-aware federated GNN literature is that federated learning is often adopted as a privacy-preserving infrastructure, but its implications

	Group fairness	Individual fairness	Cross-client fairness	Client-aware fairness
Node privacy	Song and Palanisamy 2024; Agrawal et al. 2024; Dai and Wang 2023; Chen et al. 2025; Yang and Li 2026	Wang et al. 2026; Wang et al. 2023; Wang et al. 2025b	—	—
Edge privacy	Wang et al. 2025a; Chen et al. 2025; Yang and Li 2026	Zhang, Yuan, and Pan 2024; Wang et al. 2025a	—	—
Local data privacy	Tan et al. 2026; Agrawal et al. 2024; Cui et al. 2023; Wang, Yan, and Jin 2026; Wang et al. 2024; Lo et al. 2025	—	Zhou et al. 2026; Tan et al. 2026; Khan 2026; Zhou et al. 2025; Pan et al. 2024	Cui et al. 2023; Wang, Yan, and Jin 2026; Lo et al. 2025

Table 1: Taxonomy: Fairness Objectives vs Privacy Targets

for fairness are not always explicitly theorized. Many fair-aware GNNs use FL mainly to avoid centralized access to user graphs, sensitive attributes, or organizational data, and then study how this distributed setting affects utility, convergence, personalization, or accuracy–fairness trade-offs. However, in much of this literature, the federated setting is treated primarily as a utility and optimization constraint, such as non-IID data, incomplete sensitive attributes, feature heterogeneity, or communication limitations, rather than as a structural constraint that changes how fairness itself must be measured and optimized.

This distinction is important when comparing client-aware fairness aggregation with cross-client fairness. Cross-client fairness is fundamentally a fairness constraint on performance distribution among clients. Its central question is whether the learned model serves different clients consistently or equitably. By contrast, it is not primarily a privacy constraint on fairness: it does not necessarily arise because fairness information is hidden by privacy restrictions, but because client-level outcomes may be uneven. Therefore, cross-client fairness addresses the equity of the federation for utility, while most fair-aware GNNs address demographic or task-specific group fairness within the prediction problem.

Client-aware fairness aggregation is a fairness mechanism induced by federated privacy constraints. Unlike ordinary group-fairness regularization, it is optimized through the aggregation of client updates, rather than through direct access to centralized data. These methods suggest that fairness-aware aggregation should be understood as a privacy-constrained fairness mechanism: because the server cannot freely inspect or pool raw graph data and sensitive attributes, fairness must be approximated, transmitted, and optimized through privacy-compatible client-level summaries, intermediate representations, or weighted updates.

Client-aware fairness does not merely use FL as a privacy-preserving training environment; it makes the aggrega-

tion process itself part of the fairness intervention. Compared with cross-client fairness, its target is still usually demographic group fairness, not equality of performance across clients. The client is not the protected subject, but the privacy-bounded carrier of local fairness information. Consequently, fairness-aware aggregation reveals a distinctive fairness–privacy interaction in federated GNNs: privacy constraints limit the observability of global fairness-relevant information, while aggregation becomes the mechanism through which local, partial, and privacy-preserving fairness signals are transformed into global fairness improvements.

6 Open Challenges and Future Directions

6.1 Toward Unified Modeling of Fairness–Privacy Interactions

Existing studies often analyze fairness and privacy interactions in isolated settings, such as group fairness under node attribute privacy or fairness degradation caused by differential privacy noise. However, the field still lacks a unified theoretical and methodological framework to characterize when fairness and privacy objectives conflict, when they are mutually reinforcing, and when one objective imposes constraints on the other. Future work should move beyond case-specific solutions and develop frameworks that describe fairness–privacy interactions across different graph structures, learning tasks, privacy threats, and fairness objectives.

6.2 Understanding Fairness Degradation Under Privacy-Preserving Mechanisms

Privacy-preserving mechanisms may degrade fairness in non-uniform ways. For example, differential privacy noise, graph perturbation, or feature masking affect groups, nodes, or clients differently depending on their structural positions, data density, or label distributions. Minority groups or low-degree nodes may be especially vulnerable because they

often have fewer reliable signals for representation learning. Future research should investigate how different privacy mechanisms reshape fairness outcomes across graph sub-populations and develop privacy-preserving methods that avoid disproportionate fairness degradation.

6.3 Fairness-Privacy-Utility Tensions in Federated GNNs

Federated GNNs introduce a distinctive fairness-privacy-utility tension. Federated mechanisms protect local graph data but reduce model utility by limiting the relational information needed for effective GNN training. At the same time, cross-client fairness objectives impose additional utility costs because they seek to reduce performance disparities across heterogeneous clients rather than optimize only average global accuracy. More importantly, privacy constraints may undermine fairness itself: client-aware fairness aggregation requires information about local losses, subgroup performance, sensitive attribute distributions, or client-level data heterogeneity, yet such information may reveal private properties of local datasets. Thus, privacy protection can constrain the measurement and optimization of fairness in federated GNNs. Future research should develop aggregation and auditing mechanisms that are both privacy-preserving and fairness-aware, enabling the server to mitigate client-level disparities without exposing sensitive local statistics.

6.4 Beyond Pairwise Trade-offs: Fairness, Privacy, and Utility

Although utility is beyond the scope of this review, the literature screening process revealed a major limitation of existing work: the lack of systematic analysis of the three-way relationship among fairness, privacy, and utility. Many studies report fairness and privacy metrics separately, while others focus primarily on privacy-utility trade-offs without examining whether utility degradation is distributed unevenly across groups or clients. Future work should evaluate fairness, privacy, and predictive performance jointly, preferably through trade-off curves, Pareto-front analysis, and sensitivity studies over privacy budgets, fairness constraints, and graph perturbation levels.

7 Conclusion

In this review, we analyzed the emerging literature on fairness-privacy interactions in GNNs. Rather than treating fairness-aware and privacy-preserving GNNs as independent research directions, we emphasized their interdependence. To support this analysis, we proposed a literature-derived taxonomy that organizes the existing body of studies along two orthogonal dimensions. We mapped the reviewed papers to identify dominant research patterns and underexplored intersections. The mapping shows that the existing work is concentrated around several recurring themes, including group fairness under node attribute privacy constraints, individual fairness under link privacy constraints, fairness degradation induced by differential privacy noise, and fairness under federated privacy constraints. According

to the intersections that have not been sufficiently studied, we identified potential directions of fairness-privacy interactions in GNNs. In conclusion, this review provides a foundation for future research on fair and private GNNs in high-stakes real-world applications.

Acknowledgements

This work was funded by Taighde Éireann – Research Ireland through the Research Ireland Centre for Research Training in Machine Learning (18/CRT/6183).

References

- Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H. B.; Mironov, I.; Talwar, K.; and Zhang, L. 2016. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, 308–318.
- Agrawal, N.; Sirohi, A. K.; Kumar, S.; and Jayadeva. 2024. No prejudice! fair federated graph neural networks for personalized recommendation. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, volume 38 of AAAI’24/IAAI’24/EAAI’24, 10775–10783. AAAI Press. ISBN 978-1-57735-887-9.
- Ali, A.; Naeem, H. Y.; Sharafian, A.; Qiu, L.; Wu, Z.; and Bai, X. 2025. Dynamic multi-graph spatio-temporal learning for citywide traffic flow prediction in transportation systems. *Chaos, Solitons & Fractals*, 199: 116898.
- Chen, A.; Rossi, R. A.; Park, N.; Trivedi, P.; Wang, Y.; Yu, T.; Kim, S.; Derroncourt, F.; and Ahmed, N. K. 2024. Fairness-Aware Graph Neural Networks: A Survey. *ACM Trans. Knowl. Discov. Data*, 18(6).
- Chen, Z.; Cheng, J.; Amiri, H.; Nag, K.; Lin, L.; Liu, S.; Tolomei, G.; and Sun, X. 2025. FROG: Fair Removal on Graph. In *Proceedings of the 34th ACM International Conference on Information and Knowledge Management, CIKM ’25*, 415–424. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-2040-6.
- Cheng, D.; Zou, Y.; Xiang, S.; and Jiang, C. 2025. Graph neural networks for financial fraud detection: a review. *Front. Comput. Sci.*, 19(9).
- Cui, N.; Wang, X.; Wang, W. H.; Chen, V.; and Ning, Y. 2023. Equipping Federated Graph Neural Networks with Structure-aware Group Fairness. In *2023 IEEE International Conference on Data Mining (ICDM)*, 980–985. ISSN: 2374-8486.
- Dai, E.; and Wang, S. 2021. 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*, 680–688. New York, NY, USA: Association for Computing Machinery. ISBN 9781450382977.
- Dai, E.; and Wang, S. 2023. Learning Fair Graph Neural Networks With Limited and Private Sensitive Attribute In-

- formation. *IEEE Transactions on Knowledge and Data Engineering*, 35(7): 7103–7117.
- Dai, E.; Zhao, T.; Zhu, H.; Xu, J.; Guo, Z.; Liu, H.; Tang, J.; and Wang, S. 2024. A Comprehensive Survey on Trustworthy Graph Neural Networks: Privacy, Robustness, Fairness, and Explainability. *Machine Intelligence Research*, 21(6): 1011–1061.
- Duricic, T.; Kowald, D.; Lacic, E.; and Lex, E. 2023. Beyond-accuracy: a review on diversity, serendipity, and fairness in recommender systems based on graph neural networks. *Frontiers in Big Data*, Volume 6 - 2023.
- Feng, B.; Xu, H.; Xue, W.; and Xue, B. 2022. Every Corporation Owns Its Structure: Corporate Credit Rating via Graph Neural Networks. In *Pattern Recognition and Computer Vision: 5th Chinese Conference, PRCV 2022, Shenzhen, China, November 4–7, 2022, Proceedings, Part I*, 688–699. Berlin, Heidelberg: Springer-Verlag. ISBN 978-3-031-18906-7.
- Ge, L.; Li, Y.; Li, H.; Tian, L.; and Wang, Z. 2024. A review of privacy-preserving research on federated graph neural networks. *Neurocomputing*, 600: 128166.
- Guo, Z.; Wu, Z.; Xiao, T.; Aggarwal, C.; Liu, H.; and Wang, S. 2025. Counterfactual learning on graphs: A survey. *Machine Intelligence Research*, 22(1): 17–59.
- Hamilton, W. L.; Ying, R.; and Leskovec, J. 2017. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, 1025–1035. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510860964.
- Jin, G.; Liang, Y.; Fang, Y.; Shao, Z.; Huang, J.; Zhang, J.; and Zheng, Y. 2024. Spatio-Temporal Graph Neural Networks for Predictive Learning in Urban Computing: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 36(10): 5388–5408.
- Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2021. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2): 1–210.
- Khan, K. 2026. FedGraph-Fair: Federated learning with personalization and fairness via dynamic graphs and distributionally robust optimization. *Information Sciences*, 728: 122710.
- Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.
- Kitchenham, B. 2004. Procedures for performing systematic reviews.
- Liu, R.; Xing, P.; Deng, Z.; Li, A.; Guan, C.; and Yu, H. 2025. Federated Graph Neural Networks: Overview, Techniques, and Challenges. *IEEE Transactions on Neural Networks and Learning Systems*, 36(3): 4279–4295.
- Lo, K.-C.; He, Y.; Jiang, Y.; and Parthasarathy, S. 2025. FairWAG: Fairness-aware Weighted Aggregation for Graph Learning in a Federated Setting. In *Proceedings of the 5th ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*, EAAMO ’25, 119–150. New York, NY, USA: Association for Computing Machinery. ISBN 979-8-4007-2140-3.
- Lou, Q.; Feng, B.; Charles Fox, G.; and Jiang, L. 2020. Glyph: Fast and accurately training deep neural networks on encrypted data. *Advances in neural information processing systems*, 33: 9193–9202.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.
- Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; and Galstyan, A. 2021. A Survey on Bias and Fairness in Machine Learning. *ACM Comput. Surv.*, 54(6).
- Pai, M.; McCulloch, M.; and Colford, J. 2004. Systematic Review: A Road Map Version 2.2. Systematic Reviews Group, UC Berkeley, 2002.
- Pan, C.; Xu, J.; Yu, Y.; Yang, Z.; Wu, Q.; Wang, C.; Chen, L.; and Yang, Y. 2024. Towards fair graph federated learning via incentive mechanisms. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, volume 38 of AAAI’24/IAAI’24/EAAI’24, 14499–14507. AAAI Press. ISBN 978-1-57735-887-9.
- Rampášek, L.; Galkin, M.; Dwivedi, V. P.; Luu, A. T.; Wolf, G.; and Beaini, D. 2022. Recipe for a General, Powerful, Scalable Graph Transformer. In Koyejo, S.; Mohamed, S.; Agarwal, A.; Belgrave, D.; Cho, K.; and Oh, A., eds., *Advances in Neural Information Processing Systems*, volume 35, 14501–14515. Curran Associates, Inc.
- Rocco di Torrepadula, F.; Fisichella, M.; Di Martino, S.; and Mazzocca, N. 2025. Fedflow: a personalized federated learning framework for passenger flow prediction. *Machine Learning*, 114(7): 166.
- Sajadmanesh, S.; and Gatica-Perez, D. 2021. Locally private graph neural networks. In *Proceedings of the 2021 ACM SIGSAC conference on computer and communications security*, 2130–2145.
- Song, Y.; and Palanisamy, B. 2024. MAPPING: debiasing graph neural networks for fair node classification with limited sensitive information leakage. *World Wide Web*, 27(6): 74.
- Tan, D.; Ying, Z.; Wang, Y.; Pan, Z.; Tan, Y.-A.; Zhu, T.; and Zhou, W. 2026. Enhancing fairness and privacy in federated graph neural networks via macro-level restructuring. *Knowledge-Based Systems*, 342: 115807.
- Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Wang, C.; Lin, Z.; Yang, X.; Sun, J.; Yue, M.; and Shahabi, C. 2022. HAGEN: Homophily-Aware Graph Convolutional Recurrent Network for Crime Forecasting. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(4): 4193–4200.

- Wang, S.; Tao, H.; Li, J.; Ji, X.; Gao, Y.; and Gong, M. 2024. Towards fair and personalized federated recommendation. *Pattern Recogn.*, 149(C).
- Wang, X.; Gu, T.; Bao, X.; Chang, L.; and Li, L. 2023. Individual fairness for local private graph neural network. *Knowledge-Based Systems*, 268: 110490.
- Wang, X.; Li, Y.; Gu, T.; Bao, X.; Chang, L.; Cai, G.; and Liu, T. 2025a. GRIF-PPGNN: Group equality informed Ranking-based Individual Fairness for Privacy-Preserving Graph Neural Network. *Neurocomputing*, 647: 130583.
- Wang, X.; Li, Y.; Gu, T.; Bao, X.; Chang, L.; Cai, G.; and Liu, T. 2026. Towards unified frameworks for fair and privacy-preserving graph neural networks. *Neural Networks*, 196: 108448.
- Wang, X.; Wang, L.; Gu, T.; and Bao, X. 2025b. Counterfactual Fairness for Graph Neural Networks with Limited and Privacy Protected Sensitive Attributes. In *Proceedings of the 16th Asian Conference on Machine Learning*, 719–734. PMLR.
- Wang, Z.; Yan, X.; and Jin, Y. 2026. Fairness-oriented vertical federated GNNs with incomplete sensitive attributes. *Information Fusion*, 127: 103871.
- Xue, J.; Yabe, T.; Tsubouchi, K.; Ma, J.; and Ukkusuri, S. 2022. Multiwave COVID-19 Prediction from Social Awareness Using Web Search and Mobility Data. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, 4279–4289. New York, NY, USA: Association for Computing Machinery. ISBN 9781450393850.
- Yang, T.-H.; and Li, C.-T. 2026. ReCUR: Bipartite Graph Contrastive Unlearning with Influence Estimation for Privacy-Preserved Recommendation. *Mach. Learn.*, 115(3).
- Yin, M.; Wan, M.; Lin, Z.; and Jiang, J. 2026. Moralization-aware identity fusion for detecting violent radicalization in social media. *Information Processing & Management*, 63(2, Part A): 104413.
- Ying, C.; Cai, T.; Luo, S.; Zheng, S.; Ke, G.; He, D.; Shen, Y.; and Liu, T.-Y. 2021. Do Transformers Really Perform Badly for Graph Representation? In Ranzato, M.; Beygelzimer, A.; Dauphin, Y.; Liang, P.; and Vaughan, J. W., eds., *Advances in Neural Information Processing Systems*, volume 34, 28877–28888. Curran Associates, Inc.
- Zhang, H.; Wu, B.; Yuan, X.; Pan, S.; Tong, H.; and Pei, J. 2024a. Trustworthy Graph Neural Networks: Aspects, Methods, and Trends. *Proceedings of the IEEE*, 112(2): 97–139.
- Zhang, H.; Yuan, X.; and Pan, S. 2024. Unraveling Privacy Risks of Individual Fairness in Graph Neural Networks. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, 1712–1725. ISSN: 2375-026X.
- Zhang, J.; and Tal, I. 2024. A Systematic Review of Contemporary Applications of Privacy-Aware Graph Neural Networks in Smart Cities. In *Proceedings of the 19th International Conference on Availability, Reliability and Security*, ARES '24. New York, NY, USA: Association for Computing Machinery. ISBN 9798400717185.
- Zhang, J.; and Tal, I. 2025. FL-AdvGNN: A Federated Privacy-Preserving Framework of Adversarial Graph Neural Networks. In Skopik, F.; Naessens, V.; and De Sutter, B., eds., *Availability, Reliability and Security*, 5–22. Cham: Springer Nature Switzerland. ISBN 978-3-032-00642-4.
- Zhang, S.; Yin, H.; Chen, T.; Huang, Z.; Cui, L.; and Zhang, X. 2021. Graph Embedding for Recommendation against Attribute Inference Attacks. In *Proceedings of the Web Conference 2021, WWW '21*, 3002–3014. New York, NY, USA: Association for Computing Machinery. ISBN 9781450383127.
- Zhang, X.; Zeng, Z.; Zhou, X.; Niyato, D.; and Shen, Z. 2025. Personalized federated knowledge graph embedding with client-wise relation graph: X. Zhang et al. *Applied Intelligence*, 55(5): 318.
- Zhang, Y.; Zhao, Y.; Li, Z.; Cheng, X.; Wang, Y.; Kotevska, O.; Yu, P. S.; and Derr, T. 2024b. A Survey on Privacy in Graph Neural Networks: Attacks, Preservation, and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 36(12): 7497–7515.
- Zhou, L.; Huo, D.; Chen, J.; Bo, B.; and Li, H. 2025. Federated reinforcement learning with constrained markov decision processes and graph neural networks for fair and grid-constrained coordination of large-scale electric vehicle charging networks. *Scientific Reports*, 15(1): 39593.
- Zhou, Z.; Yang, S.; Zhao, F.; and Ren, X. 2026. FairGFL: Privacy-Preserving Fairness-Aware Federated Learning With Overlapping Subgraphs. *IEEE Transactions on Parallel and Distributed Systems*, 37(3): 710–725.