CLEAR: Complementary Tripartite Play with Bayesian Calibration for Semi-Supervised Edge Classification

Zhipeng Sun[1,†] **Fanchun Meng**[1,2,†] **Jiazhen Huang**[2] **Yongpeng Zhang**[1]
Tao Ren[1] **Yifan Wang**[3*] **Wei Ju**[4] **Xiao Luo**[5]
[1]Northeastern University, China
[2]Tsinghua University, China
[3]University of International Business and Economics, China
[4]Peking University, China
[5]University of Wisconsin–Madison, USA

Abstract

This paper studies the problem of semi-supervised edge classification, which aims to identify edge relations using both labeled and unlabeled data. This problem is highly challenging due to the inherent asymmetry of edge relations and strong prediction biases from imbalanced homophilous and heterophilous edges. Towards this end, this paper proposes a novel approach named Complementary Tripartite Play with Bayesian Calibration (CLEAR) for semi-supervised edge classification. The core of our CLEAR is to incorporate asymmetric semantic branches and a meta-teacher into a tripartite-play framework, enabling reliable guidance and optimization under label scarcity. In particular, our CLEAR introduces a topological branch and a contextual branch which extracts graph semantics, generating pseudo-labels in complementary views. To improve the reliability of pseudo-labels, we filter low-quality pseudo-labels using epistemic uncertainty, and then calibrate posterior distributions using the priors induced by neighborhood information. More importantly, we utilize a meta-teacher to guide the optimization of two branches by outputting the objective coefficients, ensuring a reliable pseudo-labeling process in a tripartite play framework. Extensive experiments on benchmark datasets validate the superiority of the proposed CLEAR in comparison to various baselines. Our code is available at https://github.com/szp4ever/CLEAR.

1 Introduction

Graph neural networks (GNNs) have emerged as the dominant paradigm for modeling complex relationships in graph-structured data [12], such as protein interactions [34, 52], social media networks [23, 29] and trust relationships [21, 22]. Extensive research has demonstrated their effectiveness in various downstream tasks, including link prediction [2, 44, 53], node classification [8, 46, 50] and graph classification [48, 45, 24]. However, edge classification, which aims to identify the relation type of observed edges, remains a critical task that requires modeling of edge-level relational semantics.

Existing edge classification approaches generally rely on graph structural information, broadly categorized into heuristic and graph representation learning methods. Heuristic methods [1] typically assign edge labels based on predefined topological rules or neighborhood similarities. For graph representation learning approaches, shallow embedding techniques [3, 11, 39] capture structural proximity through random walk or deep auto-encoding methods, while GNN-based methods [20, 35, 37] generate powerful edge embeddings by recursively aggregating structural patterns from local

*Correspondence to: Yifan Wang (yifanwang@uibe.edu.cn)
†Equal contribution.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

subgraphs. Despite their progress, these structure-based methods are highly dependent on sufficient labeled data. In real-world scenarios, obtaining extensive pairwise annotations is time-consuming and costly. However, existing methods suffer from significant performance degradation under label-sparse scenarios, due to their inability to resolve the inherent structural ambiguities within complex graphs [40]. For instance, in trust networks, edges with different semantics can exhibit highly similar structural contexts, which amplifies structural ambiguity under limited supervision. This limitation motivates the exploration of semi-supervised edge classification methods [41].

While semi-supervised learning has been extensively studied for graph-based tasks [32], existing methods are largely structure-driven, relying on neighborhood aggregation and label propagation over graph topology to learn representations [18] or generate pseudo-labels [26] for unlabeled samples in node classification tasks. Recently, co-training based methods provide another semi-supervised paradigm by multiple view-specific predictors and leverage their confident predictions to supervise each other[10]. Despite their significant progress, such paradigms are not directly suited for edge classification, where the key challenge lies in learning effective edge representations under limited supervision. In particular, two critical issues remain underexplored: ❶ ***Neglecting the inherent asymmetry across different views.*** Existing co-training methods typically assume symmetric interactions across views, while the graph often contains imbalanced homophilous and heterophilous edges whose structural and semantic patterns are captured with inherently unequal fidelity across views. ❷ ***Limited and biased supervision under edge classification scenario.*** In typical semi-supervised edge classification, the labeled edges are often extremely scarce and the generated pseudo-labels often exhibit noise and bias, which further misguides training and amplify incorrect predictions, thereby limiting discriminative edge representation learning. As shown in Figure 1, these challenges indicate that effective edge classification under limited supervision requires more than exploiting unlabeled edges with conventional pseudo-labeling. Effective solutions should preserve the asymmetric reliability of different edge views, while calibrating pseudo-label supervision to avoid amplifying local class bias and overconfident errors.

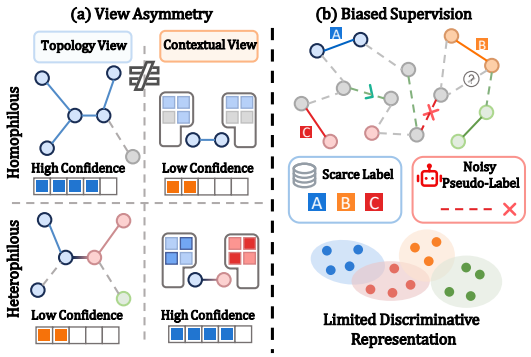


Figure 1: Motivation of our CLEAR. (a) Different views provide asymmetric reliability across homophilous and heterophilous edges. (b) Scarce labels and noisy pseudo-labels introduce biased supervision, limiting discriminative edge representation learning.

Towards this end, in this paper, we propose a Complementary Tripartite Play with Bayesian Calibration framework (CLEAR) for Semi-Supervised Edge Classification. CLEAR formulates edge learning as a complementary tripartite play among two student branches and a meta teacher. To preserve view asymmetry, the topological branch reconstructs graph topology and masked edge attributes, while the contextual branch predicts k-hop edge-neighborhood context, enabling the two predictors to retain distinct yet complementary evidence. To avoid biased pseudo-label supervision, each branch performs uncertainty-aware filtering with mutual information and uses the filtered evidence to build a neighborhood-aware Bayesian prior, which converts branch predictions into calibrated posterior pseudo-labels. The meta teacher further governs the tripartite interaction by dynamically assessing the reliability of each branch and the transferability of unlabeled edges, adaptively determining the direction and strength of cross-view supervision during training. Through validation-driven meta learning, the teacher proposes a selective cross-view supervision strategy, and the two branches follow it by learning from each other's calibrated posteriors.

This paper's contributions are summarized as follows: ❶ ***New Perspective:*** We highlight the inherent asymmetry interaction across co-training views and the limited biasd supervision for semi-supervised edge classification. ❷ ***New Methodology:*** We propose CLEAR, a tripartite play framework that retains distinct yet complementary evidence across two student branches, formalizes Bayesian pseudo-label calibration within each branch and leverages a meta-teacher to regulate their co-training dynamics. ❸ ***Extensive Experiments:*** Extensive experiments are conducted on diverse benchmark datasets, where the results demonstrate the superiority of our proposed CLEAR.

2 Related Work

Edge Classification. Edge classification [1] aims to assign semantic labels to relationships between nodes. Early approaches primarily rely on structural features, such as neighborhood overlap or path-based similarity, which are limited in capturing high-order dependencies. To improve representation capacity, embedding-based approaches have been proposed, where methods like AttrE2vec [3] aggregate features via random walks, Edge2vec [39] leverages autoencoder architectures, and TER+AER [40] encodes local topology for edge representation learning. More recently, GNNs adopt message passing to learn node embeddings and construct edge features for classification, but remain largely node-centric. Methods such as TopoEdge [4] introduce topology-aware mechanisms, combining topology entropy with reweighting and wedge-mixup to enhance edge representation learning. In semi-supervised settings, existing methods [28] typically employ pseudo-labeling or self-training to utilize unlabeled data, yet still rely heavily on supervisory signals. Consequently, learning discriminative edge representations under limited supervision remains a challenging problem.

Semi-Supervised Learning on Graphs. Semi-supervised learning on graphs has been extensively studied, with early methods mainly exploiting graph structure to propagate supervision signals for node and graph classification [27, 43, 19]. More recent approaches focus on better utilizing unlabeled data through pseudo-labeling [14, 25], self-training [42, 55], and consistency regularization [33, 49]. For example, InfoGNN [26] improves pseudo-labeling by selecting informative unlabeled nodes rather than relying solely on prediction confidence, while consistency-based methods further enhance robustness under limited labels. Despite their effectiveness, these methods are predominantly developed for node-level tasks, where supervision is propagated over node representations. Their assumptions, however, do not directly generalize to edge classification [51], where the model must distinguish structurally similar but semantically different edges under scarce supervision. Therefore, adapting graph semi-supervised learning to edge-level prediction remains a non-trivial problem.

3 The Proposed CLEAR

Preliminary. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \boldsymbol{F}, \boldsymbol{B})$ denote an undirected attributed graph, where $\mathcal{V} = \{v_i\}_{i=1}^{n}$ represents the node set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ contains all observed edges. Node attributes are encoded in $\boldsymbol{F} \in \mathbb{R}^{n \times d}$, while edge attributes are stored in $\boldsymbol{B} \in \mathbb{R}^{m \times d_e}$, with $m = |\mathcal{E}|$. The graph connectivity is represented by the adjacency matrix $\boldsymbol{A} \in \{0,1\}^{n \times n}$, where $\boldsymbol{A}_{ij}$ equals 1 if nodes v_i and v_j are connected, and 0 otherwise. For semi-supervised edge classification with limited annotations, let $\mathcal{E}^L$ and $\mathcal{E}^U$ represent the labeled and unlabeled edge subsets, respectively, satisfying $\mathcal{E}^L \cup \mathcal{E}^U = \mathcal{E}$ and $|\mathcal{E}^L| \ll |\mathcal{E}^U|$. Each labeled edge $e_{ij} \in \mathcal{E}^L$ is assigned a label vector $\boldsymbol{y}_{ij} \in \{0,1\}^C$, where C denotes the number of edge classes. Based on the graph topology, node and edge attributes, and the limited labeled samples, we aim to learn an edge encoder f_θ that maps each edge e_{ij} to a latent representation $\boldsymbol{h}_{ij}$. A classifier $g_\phi(\cdot)$ is subsequently trained on these representations to infer the edge label distribution $\hat{\boldsymbol{y}}_{ij} = g_\phi(\boldsymbol{h}_{ij})$ for unlabeled or test edges.

3.1 Framework Overview

This work focuses on semi-supervised edge classification, aiming to learn reliable edge representations from scarce labeled edges. To avoid reinforcing biased supervision under structural ambiguity, we propose CLEAR, a complementary tripartite play framework that coordinates three roles: dual branches that provide complementary representations, a meta teacher that governs their mutual learning, and a Bayesian calibrator that converts uncertain predictions into topology-aware posterior pseudo-labels. As illustrated in Figure 2, CLEAR employs dual branches to encode each edge e_{ij} from complementary views. The topological branch learns $\boldsymbol{h}_{ij}^s$ through topology reconstruction, while the contextual branch learns $\boldsymbol{h}_{ij}^c$ via context prediction. Guided by the meta teacher, the dual branches are dynamically updated according to the reliability of their predictions. Meanwhile, k-hop edge-neighborhood statistics are incorporated as neighborhood-aware priors for Bayesian calibration, producing calibrated posterior. Through this closed tripartite interaction, CLEAR progressively improves edge representations while suppressing the accumulation of unreliable supervision.

3.2 Asymmetric Semantic Branches for Complementary Predictors

To capture inherent view asymmetry, we leverage unlabeled edges to learn the complementary representations. Following the taxonomy of graph self-supervised learning [47], CLEAR instantiates

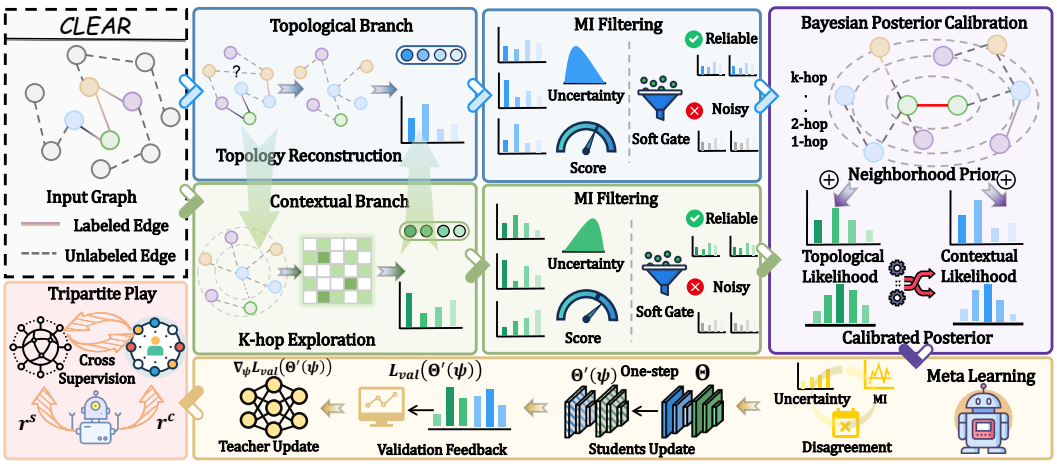


Figure 2: Framework overview of our CLEAR. Two asymmetric student branches learn complementary topological and contextual edge predictors from unlabeled edges. Each branch filters uncertain predictions with mutual information and constructs a neighborhood-aware Bayesian prior to obtain calibrated posterior pseudo-labels. A meta teacher then regulates the complementary tripartite play by assigning branch trust weights and edge-wise transfer strength for cross-view supervision.

these views as a topological branch and a contextual branch. Both branches read the same attributed graph, but they are optimized to retain different evidence for an edge.

Topological Branch for Structural Reconstruction. The topological branch learns edge representations by reconstructing the observed graph topology and masked edge attributes. Given the structural node representations, it computes the edge-existence score for a sampled node pair $(u, v) \in \Omega^+ \cup \Omega^-$ as $\hat{a}_{uv} = d_A(\boldsymbol{H}^s_u, \boldsymbol{H}^s_v)$, where $a_{uv} = 1$ if $(u, v) \in \mathcal{E}$ and $a_{uv} = 0$ otherwise. Meanwhile, for a masked observed edge $e_{ij} \in \mathcal{M}_E \subseteq \mathcal{E}$, it predicts the masked edge attribute as $\hat{\boldsymbol{b}}_{ij} = d_B(\boldsymbol{h}^s_{ij})$, where $\boldsymbol{h}^s_{ij}$ denotes its structural edge representation. Hence, the structural objective is:

$$\mathcal{L}_{\text{gae}} = -\sum_{(u,v)\in\Omega^+\cup\Omega^-} [a_{uv} \log \hat{a}_{uv} + (1 - a_{uv}) \log(1 - \hat{a}_{uv})] + \mu_e \sum_{e_{ij}\in\mathcal{M}_E} \left\| \boldsymbol{b}_{ij} - \hat{\boldsymbol{b}}_{ij} \right\|_2^2, \quad (1)$$

where the binary cross-entropy term over $\Omega^+ \cup \Omega^-$ reconstructs graph incidence, and the masked attribute regression term over $\mathcal{M}_E$, weighted by μ_e, denoises edge attributes.

Contextual Branch for K-hop Exploration. The contextual branch learns edge representations by predicting the neighborhood context of the target edge. For each edge e_{ij}, we construct a deterministic descriptor of its k-hop edge neighborhood:

$$\boldsymbol{r}_{ij} = \varrho\left(\{\boldsymbol{b}_{ab}, \boldsymbol{f}_a, \boldsymbol{f}_b, d(e_{ij}, e_{ab}) \mid e_{ab} \in \mathcal{N}_k(e_{ij})\}\right), \quad (2)$$

where $\varrho(\cdot)$ summarizes distance-bucketed edge attributes and endpoint features around e_{ij}. The prediction $\hat{\boldsymbol{r}}_{ij} = d_c(\boldsymbol{h}^c_{ij})$ is decoded from the contextual representation and estimates this neighborhood descriptor. The contextual objective is:

$$\mathcal{L}_{\text{ctx}} = \sum_{e_{ij}\in\mathcal{E}} \ell_{\text{ctx}}(\boldsymbol{r}_{ij}, \hat{\boldsymbol{r}}_{ij}), \quad (3)$$

where $\mathcal{L}_{\text{ctx}}$ encourages $\boldsymbol{h}^c_{ij}$ to preserve local edge-neighborhood semantics and ℓ_{ctx} is the cross-entropy loss in the discrete setting and the squared-error loss in the continuous setting.

3.3 Bayesian Posterior Calibration with Uncertainty-aware Filtering

To mitigate the confirmation bias caused by noisy pseudo-labels, we propose a Bayesian posterior calibration strategy. Given the complementary predictions $\boldsymbol{p}^s$ and $\boldsymbol{p}^c$, we first filter them by mutual information (MI) to retain stable pseudo-label evidence. We then use the filtered evidence to adapt a k-hop Bayesian prior around each target edge and incorporate this prior into a Bayesian calibration process for more robust posterior estimation.

Uncertainty-aware Filtering with Mutual Information. Confidence alone is an unreliable criterion for pseudo-labeling, so we estimate epistemic uncertainty with mutual information [9, 26]. For branch $* \in \{s, c\}$ and unlabeled edge e_{ij}, We obtain predictions $\{\boldsymbol{p}_{ij}^{*,t}\}_{t=1}^{K}$ by executing K stochastic forward passes under dropout. We further approximate the predictive distribution by their mean $\bar{\boldsymbol{p}}_{ij}^{*}$ and compute the MI score $\mathcal{I}_{ij}^{*}$ as:

$$\bar{\boldsymbol{p}}_{ij}^{*} = \frac{1}{K}\sum_{t=1}^{K} \boldsymbol{p}_{ij}^{*,t}, \qquad \mathcal{I}_{ij}^{*} = H(\bar{\boldsymbol{p}}_{ij}^{*}) - \frac{1}{K}\sum_{t=1}^{K} H(\boldsymbol{p}_{ij}^{*,t}), \tag{4}$$

where $H(\cdot)$ denotes predictive entropy. We further normalize $\mathcal{I}_{ij}^{*}$ by its maximum value $\log C$, as $\tilde{\mathcal{I}}_{ij}^{*}$. Based on $\tilde{\mathcal{I}}_{ij}^{*}$, we use a soft gate to produce a reliability weight for each branch prediction as:

$$m_{ij}^{*} = \sigma\left((\kappa - \tilde{\mathcal{I}}_{ij}^{*})/\gamma\right), \tag{5}$$

where γ controls the gate temperature and κ denotes the MI score threshold. Edges with low epistemic uncertainty receive larger m_{ij}^{*} and are allowed to contribute pseudo-label evidence. The resulting MI-filtered candidate from branch $*$ is $\bar{\boldsymbol{p}}_{ij}^{*}$ weighted by m_{ij}^{*}.

Bayesian Calibration with Neighborhood-aware Prior. To correct the class bias in a local subgraph, we follow the Bayesian view of combining branch likelihoods with an adaptive prior [54]. Branch predictions serve as likelihoods, while k-hop edge statistics define a local prior from labeled edges and filtered unlabeled edges. For each branch $* \in \{s, c\}$, we first instantiate local evidence, each neighboring edge e provides an evidence vector $\tilde{\boldsymbol{q}}_{e}^{*}$ that represents its class direction under branch $*$, each $\tilde{\boldsymbol{q}}_{e}^{*}$ receives a branch-specific reliability mass ρ_{e}^{*} that determines its contribution to local prior:

$$\tilde{\boldsymbol{q}}_{e}^{*} = \begin{cases} \boldsymbol{y}_{e}, & e \in \mathcal{E}^{L}, \\ \bar{\boldsymbol{p}}_{e}^{*}, & e \in \mathcal{E}^{U}, \end{cases} \qquad \rho_{e}^{*} = \begin{cases} 1, & e \in \mathcal{E}^{L}, \\ m_{e}^{*}, & e \in \mathcal{E}^{U}. \end{cases} \tag{6}$$

For target edge e_{ij}, each branch forms a graph-local prior by aggregating its neighborhood evidence:

$$\pi_{ij,c}^{*} = \frac{\alpha_0 \beta_c + \sum_{e \in \mathcal{N}_k(e_{ij})} w(e_{ij}, e)\rho_{e}^{*}\tilde{q}_{e,c}^{*}}{\alpha_0 + \sum_{e \in \mathcal{N}_k(e_{ij})} w(e_{ij}, e)\rho_{e}^{*}}, \qquad * \in \{s, c\}, \tag{7}$$

where $\boldsymbol{\beta}$ is a global class prior, α_0 controls smoothing strength, and $w(e_{ij}, e)$ measures neighborhood relevance. The resulting $\boldsymbol{\pi}_{ij}^{*}$ summarizes the local class tendency seen by branch $*$ around e_{ij}. For $* \in \{s, c\}$, the calibrated posterior is computed by a normalized product between the branch likelihood and the branch-specific local prior:

$$q_{ij,c}^{*} = \frac{(\bar{p}_{ij,c}^{*})^{1/\tau_*}\pi_{ij,c}^{*}}{\sum_{c'=1}^{C}(\bar{p}_{ij,c'}^{*})^{1/\tau_*}\pi_{ij,c'}^{*}}, \qquad * \in \{s, c\}, \tag{8}$$

where τ_* controls the sharpness of the branch likelihood before it is combined with the adaptive prior.

3.4 Controllable Tripartite Play with Meta Learning

To alleviate the branch-specific bias, we introduce a meta teacher and formulate the cross-supervision between two student branches as a tripartite play process [16]. Specifically, we enable each branch to exploit complementary evidence from the other view and introduce a selective-trust meta teacher to determine the local contributions of the two branches and the transferability of calibrated target edge for cross-view supervision at the current stage. We construct the input feature for the teacher as:

$$\boldsymbol{u}_{ij} = \left[\mathcal{I}_{ij}^{s}, \mathcal{I}_{ij}^{c}, H(\boldsymbol{q}_{ij}^{s}), H(\boldsymbol{q}_{ij}^{c}), \mathrm{JS}(\boldsymbol{q}_{ij}^{s}, \boldsymbol{q}_{ij}^{c}), \max_{c} \pi_{ij,c}^{s}, \max_{c} \pi_{ij,c}^{c}, |\mathcal{N}_k(e_{ij})|\right], \tag{9}$$

where $\mathrm{JS}(\cdot, \cdot)$ measures branch disagreement. The teacher $T_\psi(\cdot)$ maps these features to:

$$(r_{ij}^{s}, r_{ij}^{c}, \lambda_{ij}^{u}) = T_\psi(\boldsymbol{u}_{ij}), \tag{10}$$

where r_{ij}^{s} and r_{ij}^{c} denotes branch trust weights, and λ_{ij}^{u} denote the edge-wise strength of unsupervised training. Let $\Theta = \{\theta_s, \theta_c\}$ denote the parameters of the two student branches and ψ denote the parameters of the teacher, the cross-view objective can be defined as:

$$\mathcal{L}_{\mathrm{cross}}(\Theta, \psi) = \sum_{e_{ij} \in \mathcal{E}^{U}} \lambda_{ij}^{u}\left[m_{ij}^{c} r_{ij}^{c} \mathrm{CE}\big(\mathrm{sg}(\boldsymbol{q}_{ij}^{c}), \boldsymbol{p}_{ij}^{s}\big) + m_{ij}^{s} r_{ij}^{s} \mathrm{CE}\big(\mathrm{sg}(\boldsymbol{q}_{ij}^{s}), \boldsymbol{p}_{ij}^{c}\big)\right], \tag{11}$$

where $\mathrm{sg}(\cdot)$ denotes stop-gradient. Note that here m^*_{ij} and $\boldsymbol{q}^*_{ij}$ play distinct roles, where m^*_{ij} decides whether branch $*$ is eligible to teach while $\boldsymbol{q}^*_{ij}$ provides calibrated label target.We meta-train the teacher based on the assumption that improved student calibration leads to better validation performance. During each training iteration, the students first take a one-step update under the teacher-weighted training loss:

$$\Theta'(\psi) = \Theta - \eta \nabla_\Theta \left[\mathcal{L}_{\mathrm{sup}}(\Theta) + \eta_1 \mathcal{L}_{\mathrm{gae}}(\Theta) + \eta_2 \mathcal{L}_{\mathrm{ctx}}(\Theta) + \eta_3 \mathcal{L}_{\mathrm{cross}}(\Theta, \psi)\right]. \tag{12}$$

where $\mathcal{L}_{\mathrm{sup}}$ denotes the supervised cross-entropy loss computed on labeled edges in $\mathcal{E}^L$ for both branch predictions. The teacher is then updated to reduce validation loss on held-out labeled edges:

$$\psi \leftarrow \psi - \eta_\psi \nabla_\psi \mathcal{L}_{\mathrm{val}}\big(\Theta'(\psi)\big). \tag{13}$$

In practice, this gives the teacher a Stackelberg-style role [38, 15]: the teacher proposes a pseudo-labeling strategy under the fixed MI reliability filter, the student branches follow it during training, and the teacher receives feedback from validation performance.

3.5 Theoretical Guarantees

In this section, we theoretically demonstrate that the label calibration turns branch predictions into a regularized posterior estimate, and the resulting calibrated target makes cross-view co-training a bounded-error surrogate of oracle supervision. Let $\eta_{ij} \in \Delta^{C-1}$ denote the oracle (or ground-truth) label distribution of edge e_{ij}, with $\eta_{ij,c} = P(y_{ij} = c \mid e_{ij}, G)$, and let $c^\star_{ij} := \arg\max_c \eta_{ij,c}$ be the Bayes-optimal class. We use the notation $\bar{p}^*_{ij}$, m^*_{ij}, π^*_{ij}, q^*_{ij}, and p^*_{ij} as defined above.

Assumption 1 (Corrective local prior)**.** *There exists $\delta > 0$ such that every component of $\bar{p}^*_{ij}$, π^*_{ij}, and p^*_{ij} is lower bounded by δ, which can be ensured by smoothing or clipping. The temperature parameter satisfies $\tau_* > 0$ for each branch $* \in \{s, c\}$. Moreover, for a noisy edge satisfying $\arg\max_c \bar{p}^*_{ij,c} \neq c^\star_{ij}$, there exists a margin $\Delta^*_{ij} > 0$ such that, for all $c \neq c^\star_{ij}$,*

$$\log \frac{\pi^*_{ij,c^\star_{ij}}}{\pi^*_{ij,c}} \geq -\frac{1}{\tau_*} \log \frac{\bar{p}^*_{ij,c^\star_{ij}}}{\bar{p}^*_{ij,c}} + \Delta^*_{ij}. \tag{14}$$

This states that the graph-local prior constructed from labeled edges and MI-filtered unlabeled evidence is locally informative enough to compensate for the bias of the raw branch likelihood.

Proposition 1 (Calibration corrects noisy pseudo-labels)**.** *For fixed e_{ij} and $* \in \{s, c\}$, define*

$$\mathcal{J}^*_{ij}(z) := \tau_* \mathrm{KL}\big(z \,\|\, \pi^*_{ij}\big) - \big\langle z, \log \bar{p}^*_{ij} \big\rangle, \qquad z \in \Delta^{C-1}. \tag{15}$$

Then the calibrated posterior q^*_{ij} in Eq. (8) is the unique minimizer of Eq. (15), i.e., $q^*_{ij} = \arg\min_{z \in \Delta^{C-1}} \mathcal{J}^*_{ij}(z)$. And if e_{ij} is noisy, then $\arg\max_c q^*_{ij,c} = c^\star_{ij}$.

Proposition 1 shows that the calibration of CLEAR is the closed-form solution of a KL-regularized posterior estimation problem. Moreover, when the constructed graph-local prior is locally corrective, this posterior explicitly converts stable topological evidence into a corrected pseudo-label. Next, we connect this calibrated posterior form to the cross-view training objective.

Proposition 2 (Cross-view co-training is a bounded oracle surrogate)**.** *For fixed teacher parameters ψ, define the oracle cross-view objective as*

$$\mathcal{L}_{\mathrm{oracle}}(\Theta, \psi) := \sum_{e_{ij} \in E^U} \lambda^u_{ij} \left[m^c_{ij} r^c_{ij} \mathrm{CE}(\eta_{ij}, p^s_{ij}) + m^s_{ij} r^s_{ij} \mathrm{CE}(\eta_{ij}, p^c_{ij})\right]. \tag{16}$$

If $r^s_{ij}, r^c_{ij}, \lambda^u_{ij} \geq 0$, then the cross-view objective in Eq. (11) satisfies

$$\begin{aligned} &|\mathcal{L}_{\mathrm{cross}}(\Theta, \psi) - \mathcal{L}_{\mathrm{oracle}}(\Theta, \psi)| \\ &\leq \log(1/\delta) \sum_{e_{ij} \in E^U} \lambda^u_{ij} \left[m^c_{ij} r^c_{ij} \|q^c_{ij} - \eta_{ij}\|_1 + m^s_{ij} r^s_{ij} \|q^s_{ij} - \eta_{ij}\|_1\right]. \end{aligned} \tag{17}$$

Proposition 2 shows that the deviation between CLEAR's cross-view loss and the oracle objective is controlled by $\|q^*_{ij} - \eta_{ij}\|_1$. Thus, when Bayesian calibration improves the posterior estimate q^*_{ij}, the cross-view loss becomes a closer surrogate to oracle supervision. The meta teacher further regulates this surrogate by assigning r^s_{ij}, r^c_{ij}, and λ^u_{ij} according to validation feedback in Eqs. (12)-(13), thereby selecting when and how strongly calibrated targets should be transferred.

Table 1: Balanced Accuracy and Macro-F1 results for binary edge classification across four real-world benchmarks. For each dataset, the overall top score is underlined, whereas the leading result within each backbone is highlighted in **bold**.

	Bitcoin-alpha		Epinions		HSPPI		Reddit	
	B_acc	Macro_F1	B_acc	Macro_F1	B_acc	Macro_F1	B_acc	Macro_F1
ExtWF	0.425±0.021	0.414±0.010	0.492±0.007	0.448±0.003	0.482±0.001	0.412±0.005	0.429±0.006	0.486±0.003
TER+AER-G	0.684±0.025	0.625±0.013	0.563±0.032	0.533±0.047	0.500±0.002	0.432±0.002	0.561±0.014	0.563±0.022
TER+AER-P	0.688±0.021	0.611±0.017	0.570±0.031	0.517±0.042	0.500±0.001	0.432±0.003	0.549±0.008	0.549±0.005
GCN	0.526±0.006	0.533±0.010	0.571±0.003	0.580±0.004	0.464±0.003	0.467±0.004	0.499±0.006	0.502±0.005
+ TopoEdge	0.590±0.011	0.587±0.007	0.590±0.019	0.603±0.014	0.500±0.006	0.438±0.002	0.524±0.006	0.549±0.008
+ CLEAR	**0.704±0.008**	**0.595±0.004**	**0.618±0.010**	**0.608±0.009**	**0.504±0.002**	**0.501±0.001**	**0.598±0.004**	**0.571±0.012**
GAT	0.528±0.014	0.452±0.109	0.593±0.002	0.560±0.010	0.464±0.003	0.467±0.005	0.503±0.018	0.531±0.005
+ TopoEdge	0.628±0.012	0.658±0.010	0.611±0.018	0.625±0.022	0.500±0.001	0.432±0.000	0.543±0.014	0.548±0.002
+ CLEAR	**0.725±0.008**	**0.664±0.005**	**0.621±0.021**	**0.628±0.025**	**0.521±0.004**	**0.498±0.003**	**0.551±0.016**	**0.553±0.020**
Chebnet	0.533±0.003	0.511±0.048	0.586±0.003	0.592±0.008	0.516±0.004	0.428±0.001	0.527±0.008	0.543±0.006
+ TopoEdge	0.563±0.016	0.585±0.014	0.621±0.008	**0.635±0.004**	0.500±0.001	0.432±0.000	0.573±0.015	0.558±0.010
+ CLEAR	**0.712±0.003**	**0.625±0.002**	**0.625±0.011**	0.632±0.003	**0.525±0.003**	**0.510±0.006**	**0.586±0.008**	**0.564±0.009**

4 Experiments

4.1 Experimental Setting

Datasets. We evaluate CLEAR's perforamnce on six real-world benchmark graph datasets from diverse domains. *Bitcoin-alpha* [7] represents a trust network of cryptocurrency in which edges denote user ratings, with the majority class covering over 90% of edges. *Epinions* [13] is a consumer trust graph with edge features derived from user review embeddings, where trust relationships account for 85.4% of edges. *HSPPI* [34] contains protein-protein interactions and uses protein sequence features to distinguish indirect functional associations from direct physical interactions, with the latter comprising 23.9% of samples. *Reddit* [21] models inter-community hyperlinks with text-property features to classify sentiment, where neutral and positive links dominate over negative ones. *NID* [30] is an IoT network security dataset for classifying traffic flows into nine categories, dominated by injection attacks at 62.0%. *MAG* [31] is a scholarly citation/co-authorship network for classifying edges into ten fields of study, with the largest class accounting for 58.6% of edges.

Baselines. We compare CLEAR with representative baselines from different categories. Specifically, we include ExtWF [1], a heuristic method based on label dominance among similar edges, and TER+AER-G/P [40], which combines local topological encoding with edge attributes. We also evaluate standard GNN backbones, including GCN [20], GAT [37], and ChebNet [5]. To further examine the compatibility with different backbones, we integrate both TopoEdge [4] and our CLEAR into these GNNs, where TopoEdge estimates edge difficulty via label-propagation-based entropy, and then applies topological reweighting and wedge-based mixup to enhance hard-edge learning.

Implementation Details. All experiments are conducted with PyTorch and PyTorch Geometric with Adam as the optimizer on a NVIDIA A100 GPU. For data partitioning, the entire edge set is divided into training, validation, and test subsets following a 2:4:4 ratio. To simulate a limited-label scenario, only 1% of the edges are assigned labels for supervised training, while the other edges in the training subset are retained as unlabeled data. The hidden representation dimension is fixed at 256 across all backbone architectures to maintain consistency in model configurations.We set the early-stopping patience to 50 epochs as a regularization measure against overfitting. To ensure an unbiased and reproducible evaluation, we configure all baseline methods strictly according to the hyperparameter guidelines reported in their original literature. Furthermore, we measure classification quality via Balanced Accuracy and Macro-F1, presenting both the mean scores and standard deviations aggregated over five differently seeded runs.

4.2 Result and Analyses

The empirical outcomes for binary and multi-class edge classification tasks are respectively detailed in Table 1 and Table 2. The results demonstrate the effectiveness of CLEAR from the following aspects.1) *Model-agnostic generality.* CLEAR consistently improves the performance of all three GNN backbones across different datasets and metrics, showing that it is not tailored to a specific encoder architecture. This demonstrates that CLEAR can serve as a general enhancement framework for different GNN backbones in semi-supervised edge classification. 2) *Significant performance gains.* Compared with vanilla GNN backbones and strong edge classification baselines, CLEAR achieves superior results in most dataset and metric combinations. For example, on the binary

Table 2: Results of various approaches on three real-world multi-class edge classification datasets measured by Balanced Accuracy and Macro-F1, with corresonding standard deviation.

	Bitcoin-alpha		NID		MAG	
	B_acc	**Macro_F1**	**B_acc**	**Macro_F1**	**B_acc**	**Macro_F1**
ExtWF	0.223±0.002	0.216±0.007	0.397±0.004	0.384±0.006	0.211±0.008	0.209±0.002
TER+AER-G	0.254±0.021	0.268±0.023	0.547±0.028	0.463±0.018	0.298±0.012	0.305±0.009
TER+AER-P	0.271±0.019	0.298±0.016	0.564±0.022	0.473±0.030	0.309±0.011	0.311±0.009
GCN	0.248±0.013	0.240±0.006	0.468±0.021	0.448±0.018	0.289±0.002	0.309±0.005
+ TopoEdge	0.315±0.007	0.312±0.009	0.594±0.013	0.482±0.006	0.336±0.008	0.318±0.004
+ CLEAR	**0.359±0.020**	**0.322±0.009**	**0.596±0.017**	**0.494±0.013**	**0.363±0.009**	**0.321±0.007**
GAT	0.281±0.017	0.278±0.021	0.492±0.009	0.461±0.007	0.284±0.004	0.297±0.007
+ TopoEdge	0.328±0.013	0.345±0.008	0.474±0.024	0.449±0.018	0.334±0.008	0.340±0.006
+ CLEAR	**0.335±0.008**	**0.350±0.005**	**0.510±0.027**	**0.467±0.020**	**0.367±0.011**	**0.351± 0.017**
ChebNet	0.291±0.005	0.307±0.007	0.518±0.007	0.497±0.010	0.300±0.001	0.313±0.003
+ TopoEdge	0.304±0.018	0.322±0.013	0.621±0.008	0.518±0.008	0.341±0.027	**0.347±0.013**
+ CLEAR	**0.330±0.014**	**0.338±0.009**	**0.674±0.015**	**0.527±0.014**	**0.374±0.011**	0.317±0.008

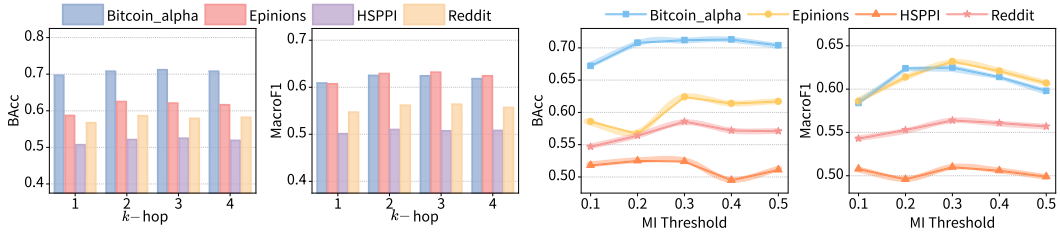


Figure 3: Hyperparameters Analysis on four binary classification datasets under ChebNet backbone.

classification task, CLEAR improves the BAcc of GAT from 0.528 to 0.725 and further outperforms GAT+TopoEdge by 0.097, indicating its strong capability in distinguishing ambiguous edges under label-scarce scenarios. 3) *Robustness to class imbalance.* The consistent gains in both BAcc and Macro-F1 suggest that CLEAR can better handle imbalanced edge distributions. Since Macro-F1 is more sensitive to minority classes, the improvements on this metric further indicate that CLEAR alleviates majority-class dominance and produces more balanced predictions. Overall, the superiority of CLEAR can be attributed to its coordinated design of complementary representation learning and reliable pseudo-label refinement. The dual branches capture asymmetric structural and contextual evidence, Bayesian calibration converts uncertain predictions into topology-aware soft pseudo-labels, and the meta teacher adaptively regulates cross-view supervision. These components jointly reduce noisy supervision and confirmation bias, leading to more discriminative edge representations.

4.3 Ablation Studies

As shown in Table 3, using ChebNet as the backbone, we evaluate the individual contributions of key components in CLEAR through extensive ablation experiments across four binary classification datasets. 1) CLEAR w/o Topological Branch: Removing the topological branch leads to severe degradation performance, showing that topology reconstruction provides useful structural evidence for edge classification, especially when labeled edges are scarce. 2) CLEAR w/o Contextual Branch: Without the contextual branch, performance consistently declines, indicating that neighborhood-semantic information complements structural evidence. 3) CLEAR w/o Bayesian Calibration: Without Bayesian calibration, performance drops noticeably, indicating that topology-aware posterior pseudo-labels are beneficial for correcting biased predictions and providing more reliable supervision on unlabeled edges. 4) CLEAR w/o Controllable Tripartite Play: Removing controllable tripartite play weakens the model significantly, showing that cross supervision between two views, coupled with meta learning, helps exploit reliable unlabeled edges and confirms the benefit of our tripartite play.

4.4 Sensitivity Analysis

Hyperparameter Sensitivity. Figure 3 shows the effectiveness of two key hyperparameters, which are the MI threshold κ and the neighborhood range k used for constructing the local topological prior. For MI threshold κ, the model achieves the best performance around $\kappa = 0.3$, indicating that a moderate threshold better balances pseudo-label reliability and coverage. A smaller threshold filters out too many unlabeled edges, while a larger one may introduce uncertain predictions and noisy

Table 3: Ablation studies on four benchmarks, 'w/o' denotes moving certain component of CLEAR.

Methods	Bitcoin-alpha		Epinions		HSPPI		Reddit	
	B_acc	Macro_F1	B_acc	Macro_F1	B_acc	Macro_F1	B_acc	Macro_F1
w/o Topological branch	0.623	0.548	0.586	0.607	0.501	0.438	0.527	0.543
w/o Contextual Branch	0.634	0.568	0.594	0.611	0.482	0.446	0.545	0.548
w/o Bayesian Calibration	0.694	0.595	0.608	0.619	0.514	0.479	0.579	0.562
w/o Controllable Tripartite Play	0.606	0.541	0.589	0.601	0.472	0.430	0.534	0.541
CLEAR	**0.712**	**0.625**	**0.625**	**0.632**	**0.525**	**0.510**	**0.586**	**0.564**

pseudo-label evidence. For the neighborhood range k, the best result is obtained at $k = 2$, while $k = 3$ yields comparable performance. This suggests that a two-hop neighborhood provides sufficient local topological evidence for Bayesian calibration. Since larger neighborhoods bring limited gains but increase computational cost, we set $k = 2$ as the default choice.

Label Rates Sensitivity. We further evaluate the sensitivity of CLEAR to label rates by varying the proportion of labeled edges from 1% to 10%. As shown in Figure 4, both BAcc and Macro-F1 consistently improve as more labeled edges are available, with the most notable gains observed from 1% to 5%. For instance, on Bitcoin-alpha, BAcc increases from 0.712 to 0.844, while Macro-F1 rises from 0.625 to 0.806. When the label rate increases beyond 5%, CLEAR shows marginal improvement, indicating that CLEAR can effectively exploit limited supervision and gradually reaches stable performance with moderate supervision. Notably, CLEAR still achieves competitive results under the low label rate of 1%, demonstrating its robustness in label-scarce scenarios. Overall, the results further highlight the consistent performance of our method when the availability of labeled data varies.

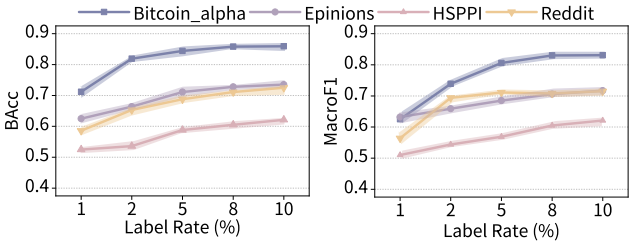
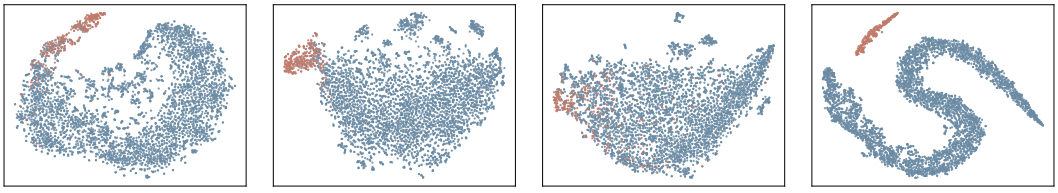


Figure 4: Label sensitivity analyses on four real-world binary edge classification datasets using ChebNet as the backbone.



(a) GAT (b) ChebNet (c) TopoEdge (d) CLEAR

Figure 5: t-SNE projections of edge embeddings learned by GAT, ChebNet, TopoEdge, and CLEAR on Bitcoin-alpha. Different colors represent different ground-truth edge classes.

4.5 Visualizations

We further conduct t-SNE [36] visualization on Bitcoin-Alpha to compare the learned edge embeddings. As shown in Figure 5, GAT and ChebNet produce noticable mixed representations, where different edge classes are difficult to separate. TopoEdge yields a slightly clearer distribution, but noticeable class overlap still remains. By contrast, CLEAR achieves stronger intra-class cohesion and clearer inter-class separation. This suggests that the proposed asymmetric dual-branch learning, Bayesian posterior calibration, and meta-guided cross-view supervision together refine edge representations, leading to more discriminative embeddings under label-scarce settings.

5 Conclusion

In this paper, we present CLEAR, a complementary tripartite play framework for semi-supervised edge classification. CLEAR learns asymmetric topological and contextual predictors from unlabeled

edges, filters uncertain branch predictions with mutual information, and converts them into posterior pseudo-labels through neighborhood-aware Bayesian calibration. A meta teacher further regulates cross-view supervision by estimating branch trust and edge-wise transfer strength from uncertainty, disagreement, and local-prior features. Experiments on various real-world datasets show consistent gains across GCN, GAT, and ChebNet backbones, suggesting that `CLEAR` provides an effective solution for semi-supervised edge classification under scarce labeled supervision.

Acknowledgements

Tao Ren is supported by the Fundamental Research Funds for the Central Universities (N25GFZ011). Yifan Wang is supported by the Fundamental Research Funds for the Central Universities in UIBE (Grant No. 23QN02), the Humanities and Social Sciences Research Fund of the Ministry of Education of China under Grant 25YJCZH275 and Open Research Fund of The State Key Laboratory of Multimodal Artificial Intelligence Systems.

Competing Interests. The authors declare no competing interests.

References

[1] Charu Aggarwal, Gewen He, and Peixiang Zhao. Edge classification in networks. In *2016 IEEE 32nd international conference on data engineering (ICDE)*, pages 1038–1049. IEEE, 2016.

[2] Baole Ai, Zhou Qin, Wenting Shen, and Yong Li. Structure enhanced graph neural networks for link prediction. *arXiv preprint arXiv:2201.05293*, 2022.

[3] Piotr Bielak, Tomasz Kajdanowicz, and Nitesh V Chawla. Attre2vec: Unsupervised attributed edge representation learning. *Information Sciences*, 592:82–96, 2022.

[4] Xueqi Cheng, Yu Wang, Yunchao Liu, Yuying Zhao, Charu C Aggarwal, and Tyler Derr. Edge classification on graphs: new directions in topological imbalance. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, pages 392–400, 2025.

[5] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.

[6] Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30, 2006.

[7] Tyler Derr, Chenxing Wang, Suhang Wang, and Jiliang Tang. Signed node relevance measurements. *arXiv preprint arXiv:1710.07236*, 2017.

[8] Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli. A fair comparison of graph neural networks for graph classification. *arXiv preprint arXiv:1912.09893*, 2019.

[9] Yarin Gal, Riashat Islam, and Zoubin Ghahramani. Deep bayesian active learning with image data. In *International conference on machine learning*, pages 1183–1192. PMLR, 2017.

[10] Maoguo Gong, Hui Zhou, A Kai Qin, Wenfeng Liu, and Zhongying Zhao. Self-paced co-training of graph neural networks for semi-supervised node classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):9234–9247, 2022.

[11] 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*, pages 855–864, 2016.

[12] Zhonghui Gu, Xiao Luo, Jiaxiao Chen, Minghua Deng, and Luhua Lai. Hierarchical graph transformer with contrastive learning for protein function prediction. *Bioinformatics*, 39(7):btad410, 2023.

[13] Ramanthan Guha, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. Propagation of trust and distrust. In *Proceedings of the 13th international conference on World Wide Web*, pages 403–412, 2004.

[14] Zhongkai Hao, Chengqiang Lu, Zhenya Huang, Hao Wang, Zheyuan Hu, Qi Liu, Enhong Chen, and Cheekong Lee. Asgn: An active semi-supervised graph neural network for molecular property prediction. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 731–752, 2020.

[15] Kjell Hausken, Jonathan W Welburn, and Jun Zhuang. A review of game theory and risk and reliability analysis in infrastructures and networks. *Reliability Engineering & System Safety*, 261:111123, 2025.

[16] Hongyang He, Xinyuan Song, Yangfan He, Zeyu Zhang, Yanshu Li, Haochen You, Lifan Sun, and Wenqiao Zhang. Trico: Triadic game-theoretic co-training for robust semi-supervised learning. *arXiv preprint arXiv:2509.21526*, 2025.

[17] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: a review. *Data mining and knowledge discovery*, 33(4):917–963, 2019.

[18] Wenhao Jiang and Yuebin Bai. Sgcl: Semi-supervised graph contrastive learning with confidence propagation algorithm for node classification. *Knowledge-Based Systems*, 301:112271, 2024.

[19] Wei Ju, Xiao Luo, Meng Qu, Yifan Wang, Chong Chen, Minghua Deng, Xian-Sheng Hua, and Ming Zhang. Tgnn: A joint semi-supervised framework for graph-level classification. *arXiv preprint arXiv:2304.11688*, 2023.

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

[21] Srijan Kumar, William L Hamilton, Jure Leskovec, and Dan Jurafsky. Community interaction and conflict on the web. In *Proceedings of the 2018 world wide web conference*, pages 933–943, 2018.

[22] Srijan Kumar, Francesca Spezzano, VS Subrahmanian, and Christos Faloutsos. Edge weight prediction in weighted signed networks. In *2016 IEEE 16th international conference on data mining (ICDM)*, pages 221–230. IEEE, 2016.

[23] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600, 2010.

[24] Haoyang Li, Xin Wang, Zeyang Zhang, Haibo Chen, Ziwei Zhang, and Wenwu Zhu. Disentangled graph self-supervised learning for out-of-distribution generalization. In *Forty-first International Conference on Machine Learning*, 2024.

[25] Jia Li, Yu Rong, Hong Cheng, Helen Meng, Wenbing Huang, and Junzhou Huang. Semi-supervised graph classification: A hierarchical graph perspective. In *The World Wide Web Conference*, pages 972–982, 2019.

[26] Yayong Li, Jie Yin, and Ling Chen. Informative pseudo-labeling for graph neural networks with few labels: Y. li et al. *Data Mining and Knowledge Discovery*, 37(1):228–254, 2023.

[27] Xiao Luo, Yusheng Zhao, Yifang Qin, Wei Ju, and Ming Zhang. Towards semi-supervised universal graph classification. *IEEE Transactions on Knowledge and Data Engineering*, 36(1):416–428, 2023.

[28] Yihong Ma, Yijun Tian, Nuno Moniz, and Nitesh V Chawla. Class-imbalanced learning on graphs: A survey. *ACM Computing Surveys*, 57(8):1–16, 2025.

[29] Babita Pandey, Praveen Kumar Bhanodia, Aditya Khamparia, and Devendra Kumar Pandey. A comprehensive survey of edge prediction in social networks: Techniques, parameters and challenges. *Expert Systems with Applications*, 124:164–181, 2019.

[30] Mohanad Sarhan, Siamak Layeghy, Nour Moustafa, and Marius Portmann. Netflow datasets for machine learning-based network intrusion detection systems. In *International Conference on Big Data Technologies and Applications*, pages 117–135. Springer, 2020.

[31] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-June Hsu, and Kuansan Wang. An overview of microsoft academic service (mas) and applications. In *Proceedings of the 24th international conference on world wide web*, pages 243–246, 2015.

[32] Zixing Song, Xiangli Yang, Zenglin Xu, and Irwin King. Graph-based semi-supervised learning: A comprehensive review. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):8174–8194, 2022.

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

[34] Damian Szklarczyk, Rebecca Kirsch, Mikaela Koutrouli, Katerina Nastou, Farrokh Mehryary, Radja Hachilif, Annika L Gable, Tao Fang, Nadezhda T Doncheva, Sampo Pyysalo, et al. The string database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic acids research*, 51(D1):D638–D646, 2023.

[35] Shanshan Tang, Bo Li, and Haijun Yu. Chebnet: efficient and stable constructions of deep neural networks with rectified power units via chebyshev approximation. *Communications in Mathematics and Statistics*, pages 1–27, 2024.

[36] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

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

[38] Heinrich Von Stackelberg, Alan T Peacock, Erich Schneider, and TW Hutchison. The theory of the market economy. *Economica*, 20(80):384, 1953.

[39] Changping Wang, Chaokun Wang, Zheng Wang, Xiaojun Ye, and Philip S Yu. Edge2vec: Edge-based social network embedding. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(4):1–24, 2020.

[40] Hewen Wang, Renchi Yang, Keke Huang, and Xiaokui Xiao. Efficient and effective edge-wise graph representation learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2326–2336, 2023.

[41] Jie Wang, Zheng Yan, Haiguang Wang, Tieyan Li, and Witold Pedrycz. A survey on trust models in heterogeneous networks. *IEEE Communications Surveys & Tutorials*, 24(4):2127–2162, 2022.

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

[43] Yifan Wang, Xiao Luo, Chong Chen, Xian-Sheng Hua, Ming Zhang, and Wei Ju. Disensemi: Semi-supervised graph classification via disentangled representation learning. *IEEE transactions on neural networks and learning systems*, 36(5):8192–8204, 2024.

[44] Yu Wang, Tong Zhao, Yuying Zhao, Yunchao Liu, Xueqi Cheng, Neil Shah, and Tyler Derr. A topological perspective on demystifying gnn-based link prediction performance. *arXiv preprint arXiv:2310.04612*, 2023.

[45] Zixiao Wang and Jicong Fan. Graph classification via reference distribution learning: Theory and practice. *Advances in Neural Information Processing Systems*, 37:137698–137740, 2024.

[46] Bang Wu, Xiangwen Yang, Shirui Pan, and Xingliang Yuan. Adapting membership inference attacks to gnn for graph classification: Approaches and implications. In *2021 IEEE International Conference on Data Mining (ICDM)*, pages 1421–1426. IEEE, 2021.

[47] Lirong Wu, Haitao Lin, Cheng Tan, Zhangyang Gao, and Stan Z Li. Self-supervised learning on graphs: Contrastive, generative, or predictive. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):4216–4235, 2021.

[48] Yulei Yang and Dongsheng Li. Nenn: Incorporate node and edge features in graph neural networks. In *Asian conference on machine learning*, pages 593–608. PMLR, 2020.

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

[50] Zeping Yu, Rui Cao, Qiyi Tang, Sen Nie, Junzhou Huang, and Shi Wu. Order matters: Semantic-aware neural networks for binary code similarity detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 1145–1152, 2020.

[51] Juwei Yue, Haikuo Li, Jiawei Sheng, Yihan Guo, Xinghua Zhang, Chuan Zhou, Tingwen Liu, and Li Guo. Graph wave networks. In *Proceedings of the ACM on Web Conference 2025*, pages 1365–1379, 2025.

[52] Min Zeng, Min Li, Zhihui Fei, Fang-Xiang Wu, Yaohang Li, and Yi Pan. A deep learning framework for identifying essential proteins based on protein-protein interaction network and gene expression data. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 583–588. IEEE, 2018.

[53] Muhan Zhang and Yixin Chen. Link prediction based on graph neural networks. *Advances in neural information processing systems*, 31, 2018.

[54] Lihua Zhou, Mao Ye, Shuaifeng Li, Nianxin Li, Xiatian Zhu, Lei Deng, Hongbin Liu, and Zhen Lei. Bayesian test-time adaptation for vision-language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 29999–30009, 2025.

[55] Xiaojin Zhu. *Semi-supervised learning with graphs*. Carnegie Mellon University, 2005.

A Limitations.

Although CLEAR provides an effective solution for semi-supervised edge classification, it still has several limitations. First, the Bayesian calibration assumes that k-hop edge neighborhoods contain useful local class evidence; when the graph is highly heterophilic or the neighborhood statistics are noisy, the estimated topological prior may become less reliable. Second, MI-based uncertainty filtering and meta-teacher regulation introduce additional training cost compared with standard pseudo-labeling. Future work may explore adaptive neighborhood selection and more efficient reliability estimation for scalable edge-level semi-supervised learning.

B Large Language Model Usage Statement.

During the preparation of this paper, LLMs were employed solely as writing assistants to support language refinement, including grammar checking, improving textual clarity, and exploring alternative expressions. Their use was limited to editorial suggestions, and no section of the paper was generated entirely by an LLM. The authors independently developed the research ideas and methodology, conducted the experiments, and performed all result interpretation and analysis. LLMs played no role in the scientific development or execution of this work. Accordingly, their contribution was limited to language-related assistance and does not warrant authorship.

C Broader Impact.

The proposed CLEAR framework advances semi-supervised edge classification through the integration of complementary topological and contextual learning, MI-guided pseudo-label filtering, and Bayesian calibration. Its asymmetric dual-branch architecture and meta-teacher-guided cross-view supervision enhance the reliability of edge representations, facilitating effective learning even when labeled data are scarce. The framework has potential applications in domains requiring reliable edge analysis, including social and trust networks, cybersecurity, and biological interaction networks. Its ability to reduce reliance on extensive edge annotations while mitigating the effects of noisy pseudo-labels makes it a promising approach for robust and efficient graph-based learning. Nevertheless, erroneous edge predictions may affect downstream decisions in sensitive applications, underscoring the importance of domain-specific validation and appropriate human oversight before deployment.

D Detailed Proofs of Theoretical Guarantees.

D.1 Proof of Proposition 1

Proof. By Assumption 1, all components of π^*_{ij} and $\bar{p}^*_{ij}$ are positive. Since $\mathrm{KL}(z\|\pi^*_{ij})$ is strictly convex in z, while the second term is linear in z, $\mathcal{J}^*_{ij}(z)$ is also strictly convex over the probability simplex. Therefore, it is sufficient to characterize the stationary point under the simplex constraint. Consider the Lagrangian:

$$\mathcal{L}(z,\lambda) = \tau_* \sum_{c=1}^{C} z_c \log \frac{z_c}{\pi^*_{ij,c}} - \sum_{c=1}^{C} z_c \log \bar{p}^*_{ij,c} + \lambda \left(\sum_{c=1}^{C} z_c - 1 \right). \tag{18}$$

Taking the derivative with respect to z_c and setting it to zero gives

$$\tau_*(\log z_c - \log \pi^*_{ij,c} + 1) - \log \bar{p}^*_{ij,c} + \lambda = 0. \tag{19}$$

Equivalently,

$$z_c = \exp\left(-1 - \frac{\lambda}{\tau_*}\right) \pi^*_{ij,c} (\bar{p}^*_{ij,c})^{1/\tau_*} \propto \pi^*_{ij,c} (\bar{p}^*_{ij,c})^{1/\tau_*}. \tag{20}$$

Normalizing over classes yields

$$z_c = \frac{(\bar{p}^*_{ij,c})^{1/\tau_*} \pi^*_{ij,c}}{\sum_{c'=1}^{C} (\bar{p}^*_{ij,c'})^{1/\tau_*} \pi^*_{ij,c'}} = q^*_{ij,c}, \tag{21}$$

which is exactly Eq. (8). The minimizer is unique by strict convexity.

Moreover, from Eq. (8), the log-ratio between Bayes-optimal class $c^\star_{ij}$ and any other class $c \neq c^\star_{ij}$ is

$$\log \frac{q^*_{ij,c^\star_{ij}}}{q^*_{ij,c}} = \frac{1}{\tau_*} \log \frac{\bar{p}^*_{ij,c^\star_{ij}}}{\bar{p}^*_{ij,c}} + \log \frac{\pi^*_{ij,c^\star_{ij}}}{\pi^*_{ij,c}}. \tag{22}$$

Combining Eq. (22) with Eq. (14) gives $\log(q^*_{ij,c^\star_{ij}}/q^*_{ij,c}) \geq \Delta^*_{ij} > 0$. Therefore, we have $q^*_{ij,c^\star_{ij}} > q^*_{ij,c}$ for all $c \neq c^\star_{ij}$, which implies $\arg\max_c q^*_{ij,c} = c^\star_{ij}$.

The same argument also yields a posterior confidence bound. Let $\Delta^{*,\min}_{ij} := \min_{c \neq c^\star_{ij}} \Delta^*_{ij}$. Since $q^*_{ij,c^\star_{ij}}/q^*_{ij,c} \geq \exp(\Delta^{*,\min}_{ij})$, we have $q^*_{ij,c} \leq q^*_{ij,c^\star_{ij}} \exp(-\Delta^{*,\min}_{ij})$ for all $c \neq c^\star_{ij}$. Summing over all non-optimal classes gives

$$1 - q^*_{ij,c^\star_{ij}} \leq (C-1) q^*_{ij,c^\star_{ij}} \exp(-\Delta^{*,\min}_{ij}). \tag{23}$$

Rearranging Eq. (23), we obtain

$$q^*_{ij,c^\star_{ij}} \geq \frac{1}{1 + (C-1)\exp(-\Delta^{*,\min}_{ij})}. \tag{24}$$

This shows that a larger corrective margin in the graph-local prior leads to a stronger posterior preference for the Bayes-optimal class. □

D.2 Proof of Proposition 2

Proof. We prove the first term, and the second term follows symmetrically. By the definition of cross-entropy,

$$\mathrm{CE}(q^c_{ij}, p^s_{ij}) - \mathrm{CE}(\eta_{ij}, p^s_{ij}) = -\left\langle q^c_{ij} - \eta_{ij}, \log p^s_{ij} \right\rangle. \tag{25}$$

By Hölder's inequality, we have

$$\left|\mathrm{CE}(q^c_{ij}, p^s_{ij}) - \mathrm{CE}(\eta_{ij}, p^s_{ij})\right| \leq \|q^c_{ij} - \eta_{ij}\|_1 \|\log p^s_{ij}\|_\infty. \tag{26}$$

By Assumption 1, $p^s_{ij,c} \geq \delta$ for all c, so $\|\log p^s_{ij}\|_\infty \leq \log(1/\delta)$. Therefore,

$$\left|\mathrm{CE}(q^c_{ij}, p^s_{ij}) - \mathrm{CE}(\eta_{ij}, p^s_{ij})\right| \leq \log(1/\delta) \|q^c_{ij} - \eta_{ij}\|_1. \tag{27}$$

Similarly, for the opposite cross-view direction,

$$\mathrm{CE}(q^s_{ij}, p^c_{ij}) - \mathrm{CE}(\eta_{ij}, p^c_{ij}) = -\left\langle q^s_{ij} - \eta_{ij}, \log p^c_{ij} \right\rangle, \tag{28}$$

and the same argument gives

$$\left|\mathrm{CE}(q^s_{ij}, p^c_{ij}) - \mathrm{CE}(\eta_{ij}, p^c_{ij})\right| \leq \log(1/\delta) \|q^s_{ij} - \eta_{ij}\|_1. \tag{29}$$

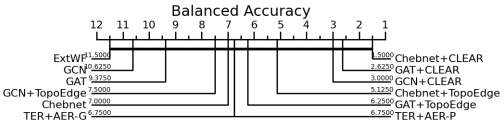
Now subtracting $\mathcal{L}_{\mathrm{oracle}}(\Theta, \psi)$ from $\mathcal{L}_{\mathrm{cross}}(\Theta, \psi)$ gives

$$\begin{aligned} &\mathcal{L}_{\mathrm{cross}}(\Theta, \psi) - \mathcal{L}_{\mathrm{oracle}}(\Theta, \psi) \\ &= \sum_{e_{ij} \in E^U} \lambda^u_{ij} \Big[m^c_{ij} r^c_{ij} \big(\mathrm{CE}(q^c_{ij}, p^s_{ij}) - \mathrm{CE}(\eta_{ij}, p^s_{ij})\big) \\ &\qquad + m^s_{ij} r^s_{ij} \big(\mathrm{CE}(q^s_{ij}, p^c_{ij}) - \mathrm{CE}(\eta_{ij}, p^c_{ij})\big) \Big]. \end{aligned} \tag{30}$$

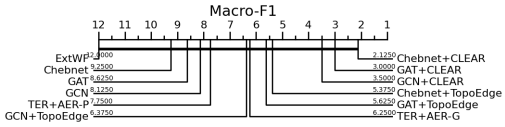
Since $r^s_{ij}, r^c_{ij}, \lambda^u_{ij} \geq 0$ and $m^s_{ij}, m^c_{ij} \geq 0$ by construction, applying the triangle inequality to Eq. (30), and then using Eq. (27) and Eq. (29), yields

$$\begin{aligned} &|\mathcal{L}_{\mathrm{cross}}(\Theta, \psi) - \mathcal{L}_{\mathrm{oracle}}(\Theta, \psi)| \\ &\leq \log(1/\delta) \sum_{e_{ij} \in E^U} \lambda^u_{ij} \left[m^c_{ij} r^c_{ij} \|q^c_{ij} - \eta_{ij}\|_1 + m^s_{ij} r^s_{ij} \|q^s_{ij} - \eta_{ij}\|_1 \right], \end{aligned} \tag{31}$$

which is Eq. (17). □



(a) Balanced Accuracy



(b) Macro-F1

Figure 6: Critical difference (CD) diagrams of different methods under BAcc and Macro-F1 metrics.

E Statistical Significance Analysis.

To further validate the robustness of CLEAR, we conduct statistical significance analyses using Critical Difference (CD) diagrams [6, 17] on four binary edge classification datasets in terms of both Balanced Accuracy and Macro-F1. As shown in Figure 6, lower ranks indicate better overall performance across different methods and datasets. The results show that CLEAR consistently achieves the top rankings under different GNN backbones, including GCN, GAT, and ChebNet, demonstrating its strong backbone-agnostic effectiveness. Moreover, CLEAR maintains superior performance on both BAcc and Macro-F1 across the four binary datasets, indicating that the proposed framework can stably improve discriminative edge representation learning under different graph structures.

F Datasets Details.

The statistics of each dataset are presented in Table 4, described as follows: 1) *Bitcoin-alpha* [7] is a cryptocurrency trust network in which edges represent ratings between users. We perform both binary and multi-class classification, with the majority class accounting for over 90% of edges. 2) *Epinions* [13] serves as a consumer trust graph where edge features are derived from user review embeddings, resulting in a dataset where trust relationships constitute 85.4% of the total edges. 3) *HSPPI* [34] focuses on protein-protein interactions where protein sequence features are used to distinguish between indirect functional and direct physical interactions, the latter of which represents a minority 23.9% of the data. 4) *Reddit* [21] captures sentiment in inter-community hyperlinks, utilizing text property vectors to identify post attitudes that are predominantly neutral or positive compared to a small minority of negative sentiments. 5) *NID* [30] addresses IoT network security by classifying traffic flows into nine categories, where injection attacks dominate at 62.0% while benign and other malicious flows are significantly underrepresented. 6) *MAG* [31] is a scholar-coauthorship citation network that requires categorizing edges into ten fields of study, where the most frequent class accounts for 58.6% of the distribution.

Table 4: Basic statistics of the imbalanced benchmark datasets, where L_0 and L_1 represent the two largest frequency ratios.

Datasets	Nodes	Edges	L_0	L_1	Labels
Bitcoin-alpha	3,784	24,207	90.3%	3.5%	2/4
Epinions	81,350	763,538	85.4%	14.6%	2
HSPPI	17,895	1,008,006	76.1%	23.9%	2
Reddit	67,180	901,056	90.5%	9.5%	2
NID	109,802	431,372	62.0%	14.7%	9
MAG	40,000	141,725	58.6%	14.2%	10

G Training Strategy for CLEAR.

Algorithm 1 summarizes the training strategy of CLEAR. At each iteration, the two student branches first learn complementary edge representations and produce branch-specific predictions. These predictions are filtered by MI-based uncertainty estimation and calibrated with neighborhood-aware Bayesian priors to form posterior pseudo-labels. The meta teacher then estimates the reliability of cross-view transfer and is updated through validation feedback, after which the student branches are optimized with the teacher-weighted objective. During inference, the meta teacher is discarded and the two calibrated branch posteriors are averaged for prediction.

Algorithm 1 Training procedure of CLEAR

Require: Graph $\mathcal{G}$, labeled edges $\mathcal{E}^L$, unlabeled edges $\mathcal{E}^U$, neighborhood range k, stochastic passes K
Ensure: Trained student parameters Θ and teacher parameters ψ
Initialize the topological branch, contextual branch, classifier heads, and meta teacher T_ψ
for each training iteration **do**
  Compute edge representations $\boldsymbol{h}_{ij}^s$ and $\boldsymbol{h}_{ij}^c$
  Compute branch predictions $\boldsymbol{p}_{ij}^s$ and $\boldsymbol{p}_{ij}^c$
  Compute self-supervised losses $\mathcal{L}_{\text{gae}}$ and $\mathcal{L}_{\text{ctx}}$ // Eq. (1), Eq. (3)
  for each branch $* \in \{s, c\}$ **do**
    Estimate $\bar{\boldsymbol{p}}_{ij}^*$ and normalized MI score $\tilde{\mathcal{I}}_{ij}^*$ // Eq. (4)
    Compute uncertainty gate m_{ij}^* // Eq. (5)
    Instantiate local evidence $\tilde{\boldsymbol{q}}_e^*$ and reliability mass ρ_e^* // Eq. (6)
    Build neighborhood-aware prior $\boldsymbol{\pi}_{ij}^*$ from local evidence // Eq. (7)
    Obtain calibrated posterior $\boldsymbol{q}_{ij}^*$ // Eq. (8)
  end for
  Build reliability feature $\boldsymbol{u}_{ij}$ // Eq. (9)
  Predict $(r_{ij}^s, r_{ij}^c, \lambda_{ij}^u) = T_\psi(\boldsymbol{u}_{ij})$ // Eq. (10)
  Compute cross-view objective $\mathcal{L}_{\text{cross}}$ // Eq. (11)
  Obtain virtual student parameters $\Theta'(\psi)$ through one teacher-weighted update // Eq. (12)
  Update ψ through the validation objective // Eq. (13)
  Update Θ with the teacher fixed
end for
At inference, average the two calibrated branch posteriors

H Further Experiments.

H.1 More Ablation Studies.

Table 5: Further ablation studies on the key components of CLEAR.

Methods	Bitcoin-alpha		NID		MAG	
	B_acc	Macro_F1	B_acc	Macro_F1	B_acc	Macro_F1
w/o Topological branch	0.292	0.306	0.603	0.479	0.342	0.298
w/o Contextual Branch	0.283	0.310	0.609	0.482	0.351	0.293
w/o Bayesian Calibration	0.312	0.308	0.626	0.504	0.332	0.301
w/o Controllable Tripartite Play	0.324	0.294	0.645	0.511	0.352	0.292
CLEAR	**0.330**	**0.338**	**0.674**	**0.527**	**0.374**	**0.317**

Table 5 reports additional ablation results on multi-class edge classification datasets. The full CLEAR achieves the best performance across all three datasets and both metrics, indicating that the proposed components remain effective when the label space becomes larger and class imbalance becomes more complex. Removing either the topological or contextual branch consistently degrades performance, especially on NID, where BAcc drops from 0.674 to 0.603 and 0.609, respectively. This verifies that structural reconstruction and neighborhood-context prediction provide complementary evidence for distinguishing fine-grained edge classes. Removing Bayesian calibration also leads to clear performance drops, such as a BAcc decrease from 0.374 to 0.332 on MAG, showing that the neighborhood-aware posterior is important for correcting biased pseudo-labels. In addition, removing controllable tripartite play reduces Macro-F1 on all datasets, with the largest drop on Bitcoin-alpha from 0.338 to 0.294, confirming that meta-learned cross-view regulation helps select more reliable supervision under multi-class settings.

Table 6: Comparison of different pseudo-label selection strategies in terms of Balanced Accuracy and Macro-F1.

Strategies	Bitcoin-alpha		Epinions		HSPPI		Reddit	
	B_acc	Macro_F1	B_acc	Macro_F1	B_acc	Macro_F1	B_acc	Macro_F1
Confidence Threshold	0.667	0.564	0.591	0.587	0.477	0.465	0.556	0.543
Entropy Threshold	0.643	0.556	0.588	0.564	0.472	0.471	0.561	0.558
MI Threshold	0.694	0.595	0.606	0.620	0.516	0.473	0.581	0.560
CLEAR	**0.712**	**0.625**	**0.625**	**0.632**	**0.525**	**0.510**	**0.586**	**0.564**

H.2 Effectiveness of Bayesian Calibration.

Table 6 compares different pseudo-label selection strategies on four binary edge classification datasets. Confidence and entropy thresholding provide simple ways to select pseudo-labels, but they only depend on the prediction distribution and can still retain overconfident noisy samples. MI thresholding improves the results on most datasets, indicating that epistemic uncertainty is a more reliable signal for filtering unlabeled edges. The full CLEAR achieves the best performance on all four datasets, showing that pseudo-label quality benefits from the combination of uncertainty-aware filtering, neighborhood-aware Bayesian calibration, and meta-teacher regulation. Compared with MI thresholding, CLEAR improves BAcc and Macro-F1 from $0.694/0.595$ to $0.712/0.625$ on Bitcoin-alpha, and from $0.516/0.473$ to $0.525/0.510$ on HSPPI. These results demonstrate that our strategy not only selects reliable pseudo-labels, but also calibrates their label distributions and controls their cross-view transfer more effectively than conventional threshold-based selection.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We present abstract and introduction that accurately reflect the contribution of our work.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss the limitation in Appendix Section A.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: We provide comprehensive proof in Appendix Section D.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: These can be seen in section implementation details. we also provide our code on github.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: We provide our code on github, we also provide the download link for the datasets.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We provide these details in section implementation details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We provide experiment statistical significance in Appendix Section E.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We present it in section implementation details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We fully abide the code of ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We provide details of broader impact in Appendix Sec. C.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The assets we use have been properly credited.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: We provide our code on github.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [Yes]

Justification: We provide LLM usage in Appendix Sec. B.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.