BRIDGING EFFICIENCY, STABILITY, AND FAIRNESS: SELF-SUPERVISED GNN-TO-MLP KNOWLEDGE DISTILLATION FOR FINANCIAL NETWORKS

Vipul Kumar Singh [1*], **Jyotismita Barman** [1*], **Sandeep Kumar** [1,2,3] **& Jayadeva** [1,2]
[1]Department of Electrical Engineering, [2]Yardi School of Artificial Intelligence,
[3]Bharti School of Telecommunication Technology and Management
Indian Institute of Technology, Delhi
{eez218227,eez218197,ksandeep,jayadeva}@iitd.ac.in

ABSTRACT

In recent times, financial data has increasingly been modeled as network structures, enabling the capture of both individual attributes and complex relationships between financial entities. Graph Neural Networks (GNNs) have become dominant tools for analyzing such data. However, GNNs face two major challenges in critical financial applications: 1) high computational costs during inference and 2) biased predictions that can disproportionately affect underrepresented groups, which can undermine the reliability of financial decision-making. To address these issues, we propose a novel self-supervised knowledge distillation framework, transferring knowledge from GNNs to Multi-Layer Perceptrons (MLPs). This framework reduces computational costs while improving model fairness, stability, and robustness in financial contexts. Specifically, we introduce feature augmentation by adding random noise and generating counterfactual versions of the input data. Extensive experiments on real-world financial datasets show that our approach surpasses existing GNN-to-MLP distillation methods, achieving the optimal balance between utility, stability, and fairness.

1 INTRODUCTION

The financial sector is a complex system where accurate decision-making is essential for managing risks, shaping investment strategies, and ensuring regulatory compliance. Financial networks such as those representing loan transactions, credit defaults, and interbank lending capture intricate and nonlinear relationships between borrowers, lenders, and institutions. These networks provide critical insights into systemic risks and financial forecasting (Acemoglu et al., 2015; Cheng et al., 2020; Bardoscia et al., 2021). Traditional machine learning models are widely used but struggle to leverage the structural and relational information present in these networks (Xu et al., 2021; Wang et al., 2019). Graph Neural Networks (GNNs) have emerged as a powerful approach for analyzing network-structured data (Kipf & Welling, 2016; Hamilton et al., 2017). By aggregating information from connected nodes, they effectively capture complex dependencies, leading to their adoption in biomedicine, materials science, and geoscience (Li et al., 2022; Zhang et al., 2021; Reiser et al., 2022; Zhou et al., 2021; Zhao et al., 2024). Their success has extended to financial applications, including money laundering detection, credit default prediction, and fraud detection (Cheng et al., 2023; Wang et al., 2023; Liu et al., 2021).

Despite their advantages, GNNs face challenges in financial applications that require high reliability and fast decision-making. Their reliance on neighborhood aggregation increases computational complexity and limits deployment in time-sensitive environments (Liu et al., 2022). Concerns regarding prediction instability and bias further hinder their adoption (Agarwal et al., 2021; Song et al., 2022; Singh et al., 2024; Dai et al., 2024). Multi-Layer Perceptrons (MLPs) remain the preferred choice for real-world deployment due to their efficiency and low latency (Zhou et al., 2024), but they struggle to capture the relational patterns essential for analyzing financial networks. To address

[*]These authors contributed equally to this work

this limitation, recent research has explored distilling knowledge from GNNs into MLPs(Yang et al., 2024; Liang et al., 2023; Zhang et al., 2022; Tian et al., 2022; Wu et al., 2023). Existing distillation methods face two key challenges. First, they struggle to ensure both stability and fairness in the distilled model. Second, they rely on labeled data, which may be restricted in financial applications due to regulatory constraints. Regulations such as GDPR may allow access to labels during the initial training of large models but prohibit their use during distillation. This poses a fundamental challenge in designing a distillation framework that effectively transfers knowledge from an unstable and biased teacher while ensuring the student model remains both fair and stable, all without explicit label supervision.

To address these challenges, we propose a self-supervised distillation framework that improves stability and fairness. Our approach learns representations by maximizing consistency between embeddings obtained from different data augmentations. To enhance stability and fairness, we introduce augmentations based on adversarial noise and counterfactual feature perturbations. To the best of our knowledge, this is the first work to investigate fairness and stability in GNN-to-MLP distillation. We validate our approach through extensive experiments on real-world financial networks, demonstrating its effectiveness and potential for deployment in practical financial applications.

2 Problem Formulation

Notations. Consider a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a node set $\mathcal{V}$ and an edge set $\mathcal{E}$. Each node is associated with a feature matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, where $n = |\mathcal{V}|$ denotes the total number of nodes, and d is the feature dimension. Additionally, each node $i \in \mathcal{V}$ has a binary sensitive attribute $s_i \in \{a, b\}$. The vector of sensitive attributes for all nodes is denoted as $\mathbf{S} = \{s_i\}_{i=1}^{n}$. The adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ encodes the graph structure, where $\mathbf{A}_{uv} = 1$ if nodes u and v are connected by an edge, and $\mathbf{A}_{uv} = 0$ otherwise. The graph Laplacian is defined as $\mathbf{\Phi} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is a diagonal degree matrix with entries $\mathbf{D}_{vv} = |\mathcal{N}(v)|$, and $\mathcal{N}(v)$ denotes the neighbors of node v.
Problem Setup. We consider the task of node classification, where each node belongs to one of C classes, represented by a label matrix $\mathbf{Y} \in \mathbb{R}^{n \times C}$. A subset of nodes $\mathcal{V}^L$ has known labels, with corresponding features $\mathbf{X}^L$ and labels $\mathbf{Y}^L$. The goal is to train a model that can generalize to the unlabeled nodes $\mathcal{V}^U$, predicting their class assignments $\mathbf{Y}^U$ based on their features $\mathbf{X}^U$ and the graph structure.
GNN-to-MLP Distillation. Let $\mathbf{Z}^{\text{GNN}} \in \mathbb{R}^{n \times C}$ and $\mathbf{Z}^{\text{MLP}} \in \mathbb{R}^{n \times C}$ denote the output embeddings of the teacher GNN and the student MLP, respectively. Traditional knowledge distillation from GNN to MLP involves aligning the predictions of the student with those of the teacher while also utilizing available labeled data. The training objective consists of two components: a supervised loss on labeled nodes and a soft-label alignment loss based on KL divergence. The overall loss function is given by:

$$\mathcal{L}_{\text{KD}} = \alpha \sum_{i \in \mathcal{V}^L} \mathcal{L}_{\text{CE}} \left(\sigma(\mathbf{z}_i^{\text{MLP}}), \mathbf{y}_i \right) + (1 - \alpha) \sum_{i \in \mathcal{V}} \mathcal{D}_{\text{KL}} \left(\sigma(\mathbf{z}_i^{\text{MLP}}/\tau) \parallel \sigma(\mathbf{z}_i^{\text{GNN}}/\tau) \right), \tag{1}$$

where $\alpha \in (0, 1)$ controls the trade-off between the supervised and distillation losses. The function $\sigma(\cdot)$ denotes the softmax activation, τ is the distillation temperature coefficient, and $\mathcal{L}_{\text{CE}}(\cdot)$ and $\mathcal{D}_{\text{KL}}(\cdot)$ correspond to the cross-entropy and KL-divergence loss functions, respectively.

Counterfactual Fairness. Counterfactual fairness is a notion used to quantify algorithmic bias based on the idea that treatment towards an individual remains the same if their sensitive attribute is changed (Kusner et al., 2017; Agarwal et al., 2021).

Definition 1 (Counterfactual fairness.) *(Kusner et al., 2017) A predictor $\hat{\mathbf{Y}}$ is said to be counterfactually fair if, for any given context $\mathbf{X} = \mathbf{x}$ and sensitive attribute $\mathbf{S} = a$, the following condition holds:*

$$P\left(\hat{\mathbf{Y}}_{\mathbf{S} \leftarrow a} = y \mid \mathbf{X} = \mathbf{x}, \mathbf{S} = a\right) = P\left(\hat{\mathbf{Y}}_{\mathbf{S} \leftarrow b} = y \mid \mathbf{X} = \mathbf{x}, \mathbf{S} = a\right) \tag{2}$$

for all possible outcomes $y \in \mathbf{Y}$, where $\hat{\mathbf{Y}}_{\mathbf{S} \leftarrow a}$ represents the prediction obtained from the counterfactual scenario where $\mathbf{S}$ is set to a.

Counterfactual fairness can be seen as a specific form of individual fairness (Kusner et al., 2017).

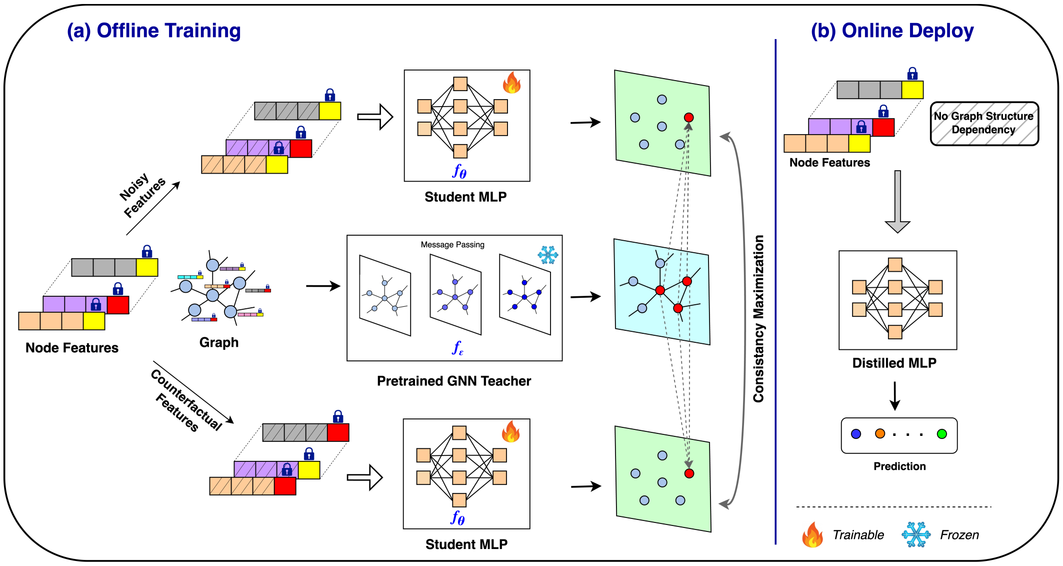

Figure 1: Overview of BEST-FAIR workflow: (a) The MLP is trained by distilling knowledge from a pre-trained GNN. (b) During inference, the MLP operates independently, utilizing only node features for predictions.

Group Fairness. The concept of group fairness in algorithmic bias aims to mitigate disparities in treatment across different demographic groups defined by sensitive attributes (Dwork et al., 2012).

Definition 2 (Statistical Parity.) *(Dwork et al., 2012) Statistical parity ensures that a classifier assigns positive predictions at equal rates across groups defined by a sensitive attribute. A binary predictor* $\hat{\mathbf{Y}}$ *satisfies statistical parity if*

$$P(\hat{\mathbf{Y}} = 1 | \mathbf{S} = a) = P(\hat{\mathbf{Y}} = 1 | \mathbf{S} = b). \tag{3}$$

Definition 3 (Equal Opportunity.) *(Hardt et al., 2016) Equal opportunity evaluates the fairness of a binary predictor* $\hat{\mathbf{Y}}$ *by ensuring that the true positive rate is equal across groups defined by a sensitive attribute* $\mathbf{S}$*, conditioned on the true label* $\mathbf{Y} = 1$*. Specifically, the equal opportunity fairness criterion holds if:*

$$P(\hat{\mathbf{Y}} = 1 | \mathbf{S} = a, \mathbf{Y} = 1) = P(\hat{\mathbf{Y}} = 1 | \mathbf{S} = b, \mathbf{Y} = 1). \tag{4}$$

Building on the foundational concepts discussed above, we formally define the research problem as follows:

Problem 1 *Given a graph dataset* $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ *with sensitive attributes* $\mathbf{S}$ *and a pre-trained GNN, the objective is to distill knowledge to an MLP that simultaneously: (1) maximizes predictive utility, (2) enhances counterfactual fairness, and (3) maximize group fairness.*

3 Methodology

The framework for **B**ridging **E**fficiency, **ST**ability, and **FAIR**ness (BEST-FAIR) is shown in Figure 1. Given input features $\mathbf{X}$ and sensitive attributes $\mathbf{S}$, we create two views: a noisy view and a counterfactual view. These views pass through an MLP encoder to produce $\mathbf{Z}_N^{\text{MLP}}$ and $\mathbf{Z}_C^{\text{MLP}}$, respectively. The MLP embeddings are supervised by the teacher GNN embeddings $\mathbf{Z}^{\text{GNN}}$. We also maximize the consistency between the MLP embeddings.

Table 1: Dataset statistics used to evaluate `BEST-FAIR`.

Dataset	# Nodes	# Features	# Edges in A	Sensitive Attribute	Task
`Income`	14,821	14	100,483	race	Above/below $50K Income
`Credit`	30,000	13	304,754	age	Default/no default Payment

Table 2: Performance comparison of different methods on the `Income` and `Credit` datasets. The best-performing values for each metric are highlighted in red, while the second-best values are shown in blue.

Dataset	Method	AUROC (↑)	CF (↓)	Stability (↓)	Δ_{SP} (↓)	Δ_{EO} (↓)
Income	GCN	$78.93_{\pm 0.23}$	$9.28_{\pm 1.68}$	$33.25_{\pm 1.15}$	$28.37_{\pm 1.21}$	$29.39_{\pm 2.42}$
	GLNN	$83.10_{\pm 0.13}$	$18.01_{\pm 2.66}$	$37.27_{\pm 0.51}$	$37.48_{\pm 1.91}$	$49.28_{\pm 4.20}$
	NOSMOG	$83.19_{\pm 0.26}$	$10.80_{\pm 1.01}$	$32.95_{\pm 1.99}$	$31.64_{\pm 2.37}$	$38.53_{\pm 3.98}$
	KRD	$83.34_{\pm 0.46}$	$21.62_{\pm 2.09}$	$36.18_{\pm 1.27}$	$30.68_{\pm 2.92}$	$55.84_{\pm 4.04}$
	`BEST-FAIR`	$84.05_{\pm 0.23}$	$3.20_{\pm 1.34}$	$33.15_{\pm 4.47}$	$15.87_{\pm 2.21}$	$25.02_{\pm 3.92}$
Credit	GCN	$70.21_{\pm 0.15}$	$19.90_{\pm 11.21}$	$25.75_{\pm 4.47}$	$12.53_{\pm 0.65}$	$11.81_{\pm 0.93}$
	GLNN	$72.98_{\pm 0.11}$	$2.96_{\pm 0.66}$	$32.29_{\pm 0.27}$	$14.03_{\pm 0.61}$	$11.39_{\pm 0.63}$
	NOSMOG	$73.60_{\pm 0.11}$	$3.04_{\pm 0.65}$	$28.67_{\pm 2.06}$	$16.19_{\pm 1.35}$	$14.91_{\pm 1.57}$
	KRD	$73.43_{\pm 0.17}$	$6.01_{\pm 2.30}$	$41.18_{\pm 3.36}$	$16.88_{\pm 2.70}$	$15.66_{\pm 3.08}$
	`BEST-FAIR`	$73.28_{\pm 4.42}$	$1.34_{\pm 1.15}$	$31.84_{\pm 4.14}$	$11.63_{\pm 3.03}$	$9.81_{\pm 2.94}$

Optimization Problem Based on the above discussion, we define a unified optimization objective for training the `BEST-FAIR` framework:

$$\mathcal{L}_{\texttt{BEST-FAIR}} = \alpha \mathop{\mathbb{E}}_{i \in \mathcal{V}} \left\| \mathbf{z}_{N_i}^{\text{MLP}} - \mathbf{z}_{C_i}^{\text{MLP}} \right\| + \beta \mathop{\mathbb{E}}_{k \in \{N,C\}} \left\| \left(\mathbf{Z}_k^{\text{MLP}}\right)^T \left(\mathbf{Z}_k^{\text{MLP}}\right) - \mathbf{I} \right\|_F^2$$
$$+\gamma \mathop{\mathbb{E}}_{i \in \mathcal{V}} \mathop{\mathbb{E}}_{j \in \mathcal{N}(i) \cup i} \left[\mathcal{D}_{\text{KL}} \left(\sigma(\mathbf{z}_{N_i}^{\text{MLP}}/\tau) \,\|\, \sigma(\mathbf{z}_j^{\text{GNN}}/\tau) \right) + \mathcal{D}_{\text{KL}} \left(\sigma(\mathbf{z}_{C_i}^{\text{MLP}}/\tau) \,\|\, \sigma(\mathbf{z}_j^{\text{GNN}}/\tau) \right) \right] \tag{5}$$

where α, β, and γ are hyperparameters that control the trade-off between the different optimization objectives. The first term enforce consistency between the embeddings of noisy and counterfactual views, ensuring robustness. To enhance representation diversity we introduce an orthogonality objective defined as the second term. Lastly, the distillation loss leverages neighborhood information to guide the MLP-encoded node embeddings. Instead of prioritizing a single objective at the cost of others, our approach explicitly models utility, fairness, and stability as a Pareto trade-off, ensuring that improvements in one metric are made with a careful balance, preventing unintended sacrifices in the others.

4 Experiments

In this section, we conduct a comprehensive evaluation of `BEST-FAIR` on two real-world financial network datasets for node classification. The detailed statistics of both the `Income` and `Credit` datasets are described in Table 1.

Baselines. We benchmark `BEST-FAIR` against several state-of-the-art baselines, including GNN Teacher, GLNN Zhang et al. (2022), NOSMOG Tian et al. (2022), and KRD Wu et al. (2023). All GNN-to-MLP distillation methods are trained without leveraging any labels.

Evaluation Metrics. We evaluate the predictive performance of all approaches on node classification using the Area Under the Receiver Operating Characteristic Curve (AUROC). To assess Counterfactual Fairness (CF), we compute the percentage of test nodes whose predicted labels change when the sensitive attribute is flipped. For stability evaluation, we measure the percentage of test nodes whose predicted labels change under random perturbations to the input features. For group fairness evaluation we consider the gap in statistical parity and equal opportunity.

Table 3: Ablation study of different BEST-FAIR variants on the Income dataset. The best-performing values for each metric are highlighted in red, while the second-best values are shown in blue.

Variations	AUROC (↑)	CF (↓)	Stability (↓)	Δ_{SP} (↓)	Δ_{EO} (↓)
BEST-FAIR	$84.05_{\pm 0.23}$	$3.20_{\pm 1.34}$	$33.15_{\pm 4.47}$	$15.87_{\pm 2.21}$	$25.02_{\pm 3.92}$
Remove α	$83.44_{\pm 0.63}$	$6.88_{\pm 1.44}$	$35.76_{\pm 4.30}$	$18.35_{\pm 5.29}$	$19.64_{\pm 0.34}$
Remove β	$84.21_{\pm 0.37}$	$3.98_{\pm 0.38}$	$35.81_{\pm 1.96}$	$18.58_{\pm 3.48}$	$27.83_{\pm 3.93}$
Remove γ	$59.23_{\pm 13.30}$	$4.51_{\pm 2.91}$	$26.76_{\pm 14.08}$	$5.57_{\pm 6.36}$	$5.48_{\pm 4.41}$
w/o CF features	$83.02_{\pm 0.88}$	$19.28_{\pm 2.14}$	$34.19_{\pm 1.12}$	$30.08_{\pm 5.27}$	$48.51_{\pm 3.00}$

4.1 Results

Balancing Utility, Stability, and Fairness. Table 2 demonstrates that BEST-FAIR achieves the most favorable trade-off between utility and fairness across both datasets. On the Income dataset, it achieves the best performance in terms of AUROC and fairness. BEST-FAIR improves utility by **1.03%** and reduces counterfactual (un)-fairness by **70.37%** with only a **0.60%** deterioration in stability compared to NOSMOG.

On the Credit dataset, BEST-FAIR achieves the lowest counterfactual and group (un)-fairness. While the GCN teacher model exhibits the highest stability, it does so with a **93.27%** higher bias level and a **4.37%** reduction in utility compared to BEST-FAIR. The improved stability in NOSMOG is due to its reliance on structural information from test nodes, which may not always be available. Additionally, the computation of positional encodings for large graphs is costly, limiting its applicability to real-time financial network analysis.

These results highlight that existing GNN-to-MLP methods, despite improving utility, are unreliable for sensitive applications. A holistic evaluation is crucial before deployment. This work addresses stringent reliability criteria and demonstrates that the proposed framework achieves a significantly improved balance across all metrics, making it the most suitable choice for financial applications.

Ablation Study. The objective function of BEST-FAIR is comprised of three distinct terms, each addressing a specific facet of performance. To evaluate the contribution of each term, we perform an ablation study, systematically removing one term at a time from the objective function. The results, presented in Table 3, demonstrate that each term positively influences the trade-off between fairness and utility. Specifically, omitting any term results in a noticeable degradation of the framework's ability to balance these competing objectives, underscoring the importance of all components in optimizing both fairness and predictive performance. Our results also highlight the critical role of counterfactual feature-based training in achieving superior fairness outcomes. The ablation results highlight that by removing the distillation objective ($\gamma = 0$) from the BEST-Fair objective the method fails in learning complex relationship between the samples, however that does make the model less unfair.

5 Conclusion

In this work, we extend the GNN-to-MLP distillation method to enhance its reliability for deployment in financial network analysis. Through extensive experiments, we show that our approach effectively navigates the trade-offs between conflicting objectives, achieving an optimal balance across key evaluation metrics in diverse financial networks. While our current focus is on scenarios where labels are missing, a more stringent privacy requirement would involve situations where no training data is available. We plan to extend this framework to address such challenges in our future work.

Acknowledgment

Vipul Kumar Singh is supported by Prime Minister's Research Fellowship (1402107), Jyotismita Barman is supported by Tata Consultancy Services Research Fellowship.

REFERENCES

Daron Acemoglu, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608, 2015.

Chirag Agarwal, Himabindu Lakkaraju, and Marinka Zitnik. Towards a unified framework for fair and stable graph representation learning. In *Uncertainty in Artificial Intelligence*, pp. 2114–2124. PMLR, 2021.

Marco Bardoscia, Paolo Barucca, Stefano Battiston, Fabio Caccioli, Giulio Cimini, Diego Garlaschelli, Fabio Saracco, Tiziano Squartini, and Guido Caldarelli. The physics of financial networks. *Nature Reviews Physics*, 3(7):490–507, 2021.

Dawei Cheng, Xiaoyang Wang, Ying Zhang, and Liqing Zhang. Risk guarantee prediction in networked-loans. In *IJCAI International Joint Conference on Artificial Intelligence*, 2020.

Dawei Cheng, Yujia Ye, Sheng Xiang, Zhenwei Ma, Ying Zhang, and Changjun Jiang. Anti-money laundering by group-aware deep graph learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(12):12444–12457, 2023.

Enyan Dai, Tianxiang Zhao, Huaisheng Zhu, Junjie Xu, Zhimeng Guo, Hui Liu, Jiliang Tang, and Suhang Wang. A comprehensive survey on trustworthy graph neural networks: Privacy, robustness, fairness, and explainability. *Machine Intelligence Research*, 21(6):1011–1061, 2024.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, pp. 214–226, 2012.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Moritz Hardt, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29, 2016.

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

Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. *Advances in neural information processing systems*, 30, 2017.

Michelle M Li, Kexin Huang, and Marinka Zitnik. Graph representation learning in biomedicine and healthcare. *Nature Biomedical Engineering*, 6(12):1353–1369, 2022.

Chen Liang, Simiao Zuo, Qingru Zhang, Pengcheng He, Weizhu Chen, and Tuo Zhao. Less is more: Task-aware layer-wise distillation for language model compression. In *International Conference on Machine Learning*, pp. 20852–20867. PMLR, 2023.

Xin Liu, Mingyu Yan, Lei Deng, Guoqi Li, Xiaochun Ye, Dongrui Fan, Shirui Pan, and Yuan Xie. Survey on graph neural network acceleration: An algorithmic perspective. In Lud De Raedt (ed.), *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pp. 5521–5529. International Joint Conferences on Artificial Intelligence Organization, 7 2022. doi: 10.24963/ijcai.2022/772. URL `https://doi.org/10.24963/ijcai.2022/772`. Survey Track.

Yang Liu, Xiang Ao, Zidi Qin, Jianfeng Chi, Jinghua Feng, Hao Yang, and Qing He. Pick and choose: a gnn-based imbalanced learning approach for fraud detection. In *Proceedings of the web conference 2021*, pp. 3168–3177, 2021.

Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, et al. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93, 2022.

Vipul Kumar Singh, Sandeep Kumar, Avadhesh Prasad, et al. A unified optimization-based framework for certifiably robust and fair graph neural networks. *IEEE Transactions on Signal Processing*, 2024.

Weihao Song, Yushun Dong, Ninghao Liu, and Jundong Li. Guide: Group equality informed individual fairness in graph neural networks. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1625–1634, 2022.

Yijun Tian, Chuxu Zhang, Zhichun Guo, Xiangliang Zhang, and Nitesh Chawla. Learning mlps on graphs: A unified view of effectiveness, robustness, and efficiency. In *The Eleventh International Conference on Learning Representations*, 2022.

Daixin Wang, Jianbin Lin, Peng Cui, Quanhui Jia, Zhen Wang, Yanming Fang, Quan Yu, Jun Zhou, Shuang Yang, and Yuan Qi. A semi-supervised graph attentive network for financial fraud detection. In *2019 IEEE international conference on data mining (ICDM)*, pp. 598–607. IEEE, 2019.

Daixin Wang, Zhiqiang Zhang, Yeyu Zhao, Kai Huang, Yulin Kang, and Jun Zhou. Financial default prediction via motif-preserving graph neural network with curriculum learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2233–2242, 2023.

Lirong Wu, Haitao Lin, Yufei Huang, and Stan Z Li. Quantifying the knowledge in gnns for reliable distillation into mlps. In *International Conference on Machine Learning*, pp. 37571–37581. PMLR, 2023.

Bingbing Xu, Huawei Shen, Bingjie Sun, Rong An, Qi Cao, and Xueqi Cheng. Towards consumer loan fraud detection: Graph neural networks with role-constrained conditional random field. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 4537–4545, 2021.

Zhendong Yang, Zhe Li, Ailing Zeng, Zexian Li, Chun Yuan, and Yu Li. Vitkd: Feature-based knowledge distillation for vision transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1379–1388, 2024.

Shichang Zhang, Yozen Liu, Yizhou Sun, and Neil Shah. Graph-less neural networks: Teaching old mlps new tricks via distillation. In *International Conference on Learning Representations*, 2022. URL `https://arxiv.org/abs/2110.08727`.

Xiao-Meng Zhang, Li Liang, Lin Liu, and Ming-Jing Tang. Graph neural networks and their current applications in bioinformatics. *Frontiers in genetics*, 12:690049, 2021.

Shan Zhao, Zhaiyu Chen, Zhitong Xiong, Yilei Shi, Sudipan Saha, and Xiao Xiang Zhu. Beyond grid data: Exploring graph neural networks for earth observation. *IEEE Geoscience and Remote Sensing Magazine*, 2024.

Fan Zhou, Rongfan Li, Goce Trajcevski, and Kunpeng Zhang. Land deformation prediction via slope-aware graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 15033–15040, 2021.

Yexu Zhou, Tobias King, Haibin Zhao, Yiran Huang, Till Riedel, and Michael Beigl. Mlp-har: Boosting performance and efficiency of har models on edge devices with purely fully connected layers. In *Proceedings of the 2024 ACM International Symposium on Wearable Computers*, pp. 133–139, 2024.