Individual Fairness in Dynamic Financial Networks

Zixing Song Yueen Ma Irwin King
Department of Computer Science and Engineering
The Chinese University of Hong Kong
Sha Tin, N.T., Hong Kong SAR
{zxsong, yema21, king}@cse.cuhk.edu.hk

Abstract

In the financial world, a transaction graph is commonly used for modeling the ever-changing payee-payor relationships. Every online transaction corresponds to a directed edge from the paying party to the receiving party in this graph. Even though the superior learning capability of Graph Neural Networks (GNNs) has led to several successful financial applications like fraud detection and anti-money laundering, most of these existing works do not have fairness considerations. Apparently, the lack of fairness guarantees during the GNN-based decision-making process would cause increasingly serious societal concerns from both buyers and sellers. Furthermore, the time-varying property of the financial networks makes the fairness requirements more challenging, since current fairness measures on graph learning tasks and fairness-aware GNN models are all designed for static graphs only. In this work, we present a new generic definition of individual fairness for dynamic graphs and propose a regularization-based method to debias the GNN model in the temporal setting. We perform some preliminary experimental evaluations on two real-world datasets and demonstrate the potential efficacy of the proposed methods.

1 Introduction

The financial system is one of the most complex systems with massive interacting parties, which are also constantly updated. Therefore, graphs are commonly constructed in order to model the dynamic relational data in the financial domain, such as dynamic loan networks [1], Bitcoin trust graphs [8] and supply chain graphs [14]. Almost all of these financial networks can be abstracted into a dynamic directed graph where the nodes represent financial institutions or clients, and the edges represent the directional relations from one node to the other, like the buyer-seller relationship, the borrower-lender relationship, and the trustor-trustee relationship. The dynamic or temporal nature of this financial network makes the downstream learning tasks more challenging, as it is continuously evolving with transactions (edges) being created or canceled every second, merchants (nodes) appearing or disappearing every day, and clients' information (node features) updating any time. To gain interesting insights from these graphs, Graph Neural Networks (GNNs) are currently the mainstream learning paradigms for various tasks [7]. Among them, the family of temporal GNNs [10, 11, 15], a time-dependent variant of GNN models, is the natural fit for the ever-changing financial networks with myriad promising applications [13], including fraudster detection at the node level, future transactions at the edge level, and money laundering detection at the sub-graph level.

However, high-stakes decisions under strict laws and regulations are often involved in the financial system, and thus, fairness concerns beyond the utility of temporal GNNs are in urgent demand. For instance, if a bank makes loan approval predictions based on the transaction record network among applicants, the decision must be independent of the sensitive attribute of the applicant, like gender and race. Unfortunately, for one thing, current popular temporal GNNs fail to satisfy reasonable notions

NeurIPS 2022 New Frontiers in Graph Learning Workshop (NeurIPS GLFrontiers 2022).

of fairness due to the inherent algorithmic bias in GNNs. For another, most fairness-enhanced static GNNs cannot be directly generalized in the dynamic setting due to the overwhelming computational cost and the time-varying fairness-related information loss. Consequently, how to measure fairness for temporal graph learning tasks and how to equip dynamic GNNs with fairness considerations are still open research questions [3].

In this work, we aim to present a principled study of individual fairness for temporal GNNs on financial transaction graphs. Compared with group fairness, individual fairness quantifies potential bias or discrimination at a finer granularity and follows the general principle that any two similar individuals should be treated with similar outcomes. Two fundamental research questions would be raised naturally when we extend individual fairness notions to temporal GNNs. First, how can we quantitatively measure the overall bias of the predictions for temporal graph learning tasks? Second, how can we alleviate such bias by developing the fairness-aware temporal GNNs model accordingly? For the first question, we introduce a new definition of fairness in retrospect, which requires the predictions to be individually fair relative to the past at the time of the decision based on the Lipschitz property. This proposed fairness measure can easily enable us to quantify the overall algorithmic bias in the predicted results for the graph-based learning tasks under the temporal setting. For the second question, we propose a novel individual fairness regularization term that can be integrated into almost all temporal GNNs without much effort. As far as we are concerned, we are the first to investigate the individual fairness measures and methods for dynamic GNNs. We also present some preliminary experiments to demonstrate the potential applications in real-world financial networks.

2 Methodology

2.1 Problem Formulation

A static graph can be denoted as $G = (V, E)$ with the node set $V = \{v_1, \ldots, v_n\}$ and the edge set $E \subseteq V \times V$. The nodes could be associated with features $X = \{\mathbf{x}_v \mid v \in V\}$ and the edges could also have features $F = \{\mathbf{f}_{uv} \mid (u, v) \in E\}$. The dynamic or temporal graph under the discrete setting can be viewed as a sequence of static graphs $\mathcal{G} = \{G_t\}_{t=1}^{T}$, where each graph has a unique timestamp t. In each snapshot $G_t = (V_t, E_t)$, the note set V_t ($|V_t| = n_t$), the edge set E_t and their characteristics could vary with t. This way of modeling can handle the dynamics of financial networks in the real world. Current dynamic graph representation learning aims to learn latent representations $\mathbf{h}_v^t \in \mathbb{R}^d$ for each node at time step $t \in \{1, \ldots, T\}$, such that $\mathbf{h}_v^t$ can encode both the neighborhood information centered at v and its temporal evolutionary behaviors of the graph [11].

In this work, we push this goal further and restrict the representation $\mathbf{h}_v^t$ to be *individually fair* with respect to the ground-truth results. For example, if the learning task is the node classification, we want to predict the label of a node u at time t based on the previous snapshots $\{G_\tau\}_{\tau=1}^{t-1}$. We use $\mathbf{Y}_t \in \{0, 1\}^{n_t \times c}$ and $\hat{\mathbf{Y}}_t \in \mathbb{R}^{n_t \times c}$ to denote the ground truth and the predicted class membership matrix, respectively. Then we hope that $\hat{\mathbf{Y}}_t[u] \in \mathbb{R}^c$ (the u-th row of $\hat{\mathbf{Y}}_t$) and $\hat{\mathbf{Y}}_t[v]$ are similar as long as node u and node v are similar on the temporal graph.

2.2 Individual Fairness Measures for Dynamic Graphs

In this section, we propose a novel metric to quantitatively measure the level of overall bias in the predicted results on the dynamic graphs. This new metric can motivate the design of our proposed method, which potentially enhances the individual fairness of dynamic GNNs.

First, let us revisit the definition of the individual fairness measure for static graphs, following the Lipschitz condition [4].

Definition 2.1 ((D_1, D_2)-Lipschitz property)**.** Given a learning model f, denote $f(v)$ as the predicted outcome of node v on a static graph G. Then f satisfies (D_1, D_2)-Lipschitz property if

$$D_1(f(u), f(v)) \leq L D_2(u, v), \quad \forall u, v \in V.$$

It is also worth noting that we assume there is an oracle similarity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ to measure the pairwise node similarity according to domain knowledge or human judgment [12]. Existing works promote individual fairness on static graphs through the defined (D_1, D_2)-Lipschitz property. The basic intuition is that individual nodes with higher similarity should be constrained to have a shorter distance in the embedding space. Therefore, if we set $D_1(f(u), f(v)) = \|\hat{\mathbf{Y}}[u] - \hat{\mathbf{Y}}[v]\|_2$, $D_2(u, v) = \frac{1}{\mathbf{S}[u,v]}$ and $L = \epsilon > 0$ (as a fairness tolerance constant), we immediately obtain the following condition Eq. (1)

$$\|\hat{\mathbf{Y}}[u] - \hat{\mathbf{Y}}[v]\|_2^2 \leq \frac{\epsilon}{\mathbf{S}[u, v]} \quad \forall u, v \in V. \tag{1}$$

The difference between the predicted results of a node pair (u, v) is upper bounded by the RHS of Eq. (1). The more similar the node u and the node v are, the tighter the upper bound in Eq. (2) is, and the smaller the difference between $\hat{\mathbf{Y}}[u]$ and $\hat{\mathbf{Y}}[v]$ will be, achieving the individual fairness goal from the perspective of Lipschitz constraint in Definition 2.1. By summing the LHS of Eq. (1) over all node pairs, we can get the overall bias for the static graph in terms of individual fairness (IF-S) as Definition 2.2 that is widely employed in the current literature [6, 2, 12]. The smaller the IF-S, the more individually fair the predicted results.

Definition 2.2 (Individual fairness on static graphs)**.** Given a static graph $G = (V, E)$ with $|V| = n$, the predicted membership matrix $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times c}$ and the oracle pairwise similarity matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$, Eq. (2) measures the overall bias regarding the individual fairness.

$$\text{IF-S} = \frac{1}{2} \sum_{u \in V} \sum_{v \in V} \|\hat{\mathbf{Y}}[u] - \hat{\mathbf{Y}}[v]\|_2^2 \mathbf{S}[u, v] = \text{Tr}(\hat{\mathbf{Y}}^\intercal \mathbf{L_S} \hat{\mathbf{Y}}). \tag{2}$$

Here, $\mathbf{L_S}$ is the Laplacian matrix of $\mathbf{S}$.

For dynamic graphs, the most straightforward extension of IF-S is to sum all IF-S over all the timestamps as $\sum_{t=1}^{T} \text{IF-S}_t$, where each IF-S_t follows Definition 2.2 at t-th snapshot. However, this metric views each snapshot independently, and thus fails to take the variation of the fairness constraint across time into account. Nodes with different timestamps but with similar contexts should also share similar outcomes. Therefore, in order to incorporate the temporal property into the fairness measurement, we first present a novel Lipschitz condition under the dynamic setting and then construct a new individual fairness metric on dynamic graphs accordingly.

Definition 2.3 ((K, D_1, D_2)-Lipschitz in retrospect property)**.** Given a learning model f, denote $f(v_t)$ as the predicted outcome of node v on a dynamic graph at time step t. Then f satisfies (K, D_1, D_2)-Lipschitz in retrospect property with respect to a monotonically non-decreasing function $L(\cdot)\colon \mathbb{R}^+ \to \mathbb{R}^+$ and a non-negative integer $K \in \mathbb{Z}^*$ if

$$D_1(f(u), f(v)) \leq L(t - t')D_2(u, v), \quad \forall u \in V_{t'}, v \in V_t, t - t' \leq K.$$

Note that in Definition 2.3, by setting $L(\cdot)$ to be a monotonically non-decreasing function, we can model the scenario where the past predictions have a diminishing impact on the future predictions depending on the amount of time passed. This insight is motivated by the assumption that nodes with similar features should achieve similar predictions compared with their counterparts in recent history, but could obtain dissimilar treatments compared with those in the distant past. For example, in the transaction network for load approval, both Alice and Bob earn almost the same amount of money annually, but Alice applies for the loan several years before Bob does, appearing in the earlier snapshots of the network. It would still be considered individually fair if the bank approves Alice's applications but rejects Bob's since the average personal income could grow over the years. Another interpretation of Definition 2.3 is that it can be considered as a generalized version of Definition 2.1. If we set $K = 0$ ($t' = t$) in Definition 2.3, then the degenerate definition becomes the same as Definition 2.1 with $L = L(0)$. Therefore, Definition 2.3 is a more strict constraint by restricting the current treatment to be individually fair relative to the recent past within the interval of K snapshots.

Similar to Eq. (1), if we still set $D_1(f(u), f(v)) = \|\hat{\mathbf{Y}}_{t'}[u] - \hat{\mathbf{Y}}_t[v]\|_2^2$, $D_2(u, v) = \frac{1}{\mathbf{S}_{t',t}[u,v]}$ and set $L(t - t') = \epsilon \exp(t - t')$ with $\epsilon > 0$, we can get the individual fairness requirement for the temporal graphs as Eq. (3).

$$\|\hat{\mathbf{Y}}_{t'}[u] - \hat{\mathbf{Y}}_t[v]\|_2^2 \leq \frac{\epsilon \exp(t - t')}{\mathbf{S}_{t',t}[u, v]} \quad \forall u \in V_{t'}, v \in V_t, t - t' \leq K. \tag{3}$$

Intuitively, Eq. (3) leads to a tighter upper bound on the inconsistency of the treatments for a pair of nodes if their similarity $\mathbf{S}_{t',t}[u, v]$ is greater and their existence falls into a shorter time interval $(t - t')$. We then propose a new **I**ndividual **F**airness metric for **D**ynamic graphs (IF-D) accordingly by summing all node pairs as Definition 2.4.

Table 1: Summary of dataset statistics.

	# Edges	# Nodes	Temporal Range	Snapshot Frequency	# Snapshots
Bitcoin-OTC	35,592	5,811	Nov 8, 2010 - Jan 24, 2016	weekly	279
Bitcoin-Alpha	24,186	3,783	Nov 7, 2010 - Jan 21, 2016	weekly	274

Definition 2.4 (Individual fairness on dynamic graphs). Given a dynamic graph $\mathcal{G} = \{G_t\}_{t=1}^{T}$, where $G_t = (V_t, E_t)$ with $|V_t| = n_t$, the predicted membership matrix $\hat{\mathbf{Y}}_t \in \mathbb{R}^{n_t \times c}$ at time step t and the oracle pairwise similarity matrices $\mathbf{S}_{t',t} \in \mathbb{R}^{n_{t'} \times n_t}$ denoting the similarity distance between the nodes at the t-th snapshot with nodes from any previous K snapshots with $\max(0, t-K) \leq t' \leq t$. We define the bias regarding the individual fairness on dynamic graphs at the t-th snapshot IF-D$_t$ as

$$\text{IF-D}_t = \frac{1}{2} \sum_{t'=[t-K]^+}^{t} \exp(t'-t) \sum_{u \in V_{t'}} \sum_{v \in V_t} \|\hat{\mathbf{Y}}_{t'}[u] - \hat{\mathbf{Y}}_t[v]\|_2^2 \mathbf{S}_{t',t}[u,v]. \quad (4)$$

2.3 Individual Fairness Regularization for Dynamic Graphs

To achieve the individual fairness goal on dynamic graphs, we naturally propose a simple yet effective regularization-based method that could be plugged into almost all current state-of-the-art dynamic GNN models [5, 11, 10]. Based on the proposed Definition 2.4, we can make the assumption that $|V_{t'}| = |V_t|$ ($n_{t'} = n_t$) when K is small. This corresponds to the fact that the number of clients in a transaction network will barely change in a sufficiently short time interval. Even if the number of nodes changes, we can still add some “pseudo-nodes" with random features and links during the training process to make up for the difference in terms of the number of nodes. Thus, we can construct a simplified individual fairness regularization loss for dynamic graphs based on Definition 2.4 as Eq. (5).

$$L_{\text{IF-D}_t} = \sum_{t'=[t-K]^+}^{t} \exp(t'-t)\, \text{Tr}(\hat{\mathbf{Y}}_{t'}^{\intercal} \mathbf{L}_{\mathbf{S}_{t',t}} \hat{\mathbf{Y}}_t) \quad (5)$$

Note that now $\mathbf{S}_{t',t} \in \mathbb{R}^{n_t \times n_t}$ and $\mathbf{L}_{\mathbf{S}_{t',t}} \in \mathbb{R}^{n_t \times n_t}$ (the Laplacian matrix of $\mathbf{S}_{t',t}$) are square matrices. For the sake of training efficiency, we can set $K = 3$. Then, suppose that the utility loss in the backbone dynamic GNN model is L_{util}, then the total training loss can be obtained as $L = L_{\text{util}} + \alpha \sum_{t=1}^{k} L_{\text{IF-D}_t}$. Here, k is the total number of snapshots for training, and α is the balancing factor.

3 Preliminary Experiments

In this section, we present some preliminary experiments on two real-world datasets from the financial domain to validate the efficacy of the proposed method. This is ongoing work, and we are still conducting more experiments on more temporal graph datasets with more baseline models as well.

Datasets The datasets we use are Bitcoin-OTC and Bitcoin-Alpha datasets. They contain who-trusts-whom transaction networks of people who trade on the OTC and Alpha platforms [8, 9]. The statistics of the datasets are summarized in Table 1.

Task We evaluate the proposed method over the future link prediction task (edge classification). We concatenate the embeddings of two linked nodes from the last layer and apply an MLP to obtain the class membership probability matrix. At each time step t, we can utilize the history information accumulated up time t to predict the edge in the snapshot $t+1$. For other tasks like node classification or graph classification, we plan to add them in the future to further verify the capability of the model.

Evaluation metrics We choose the mean reciprocal rank (MRR) to assess the models in terms of utility and choose the proposed IF-D to evaluate the models in terms of individual fairness. MRR score is the mean of reciprocal ranks over all nodes u. We use the last 10% of snapshots for testing. For each node u with positive edge (u, v) at $t+1$, we randomly sample 1,000 negative edges starting from node u and identify the rank of edge (u, v)'s prediction score among all other negative edges.

Table 2: Performance of the link prediction task on two financial datasets. IF-D metrics are reported in a relative manner. The minimum IF-D in each column is set to 1.

	Bitcoin-OTC		Bitcoin-Alpha	
	MRR↑	IF-D↓	MRR↑	IF-D↓
EvolveGCN-H	0.0690	1.325	0.1104	1.201
EvolveGCN-H + Ours	0.0712	1.087	0.1093	1.126
EvolveGCN-O	**0.0968**	1.264	**0.1185**	1.198
EvolveGCN-O+Ours	0.0953	**1.000**	0.1179	**1.000**

Baselines For the backbone dynamic GNN models, we choose the most popular EvolveGCN [10]. Both EvolveGCN-H and EvolveGCN-O use an RNN to dynamically update the training parameters of GNNs. The former can be implemented by using a standard GRU, while the latter requires a straightforward extension of the standard LSTM from the vector version to the matrix version.

From Table 2, we can conclude that our proposed method can greatly improve individual fairness for temporal financial networks without sacrificing the utility of the backbone model too much. Surprisingly, in some datasets, our method can even improve the utility when combined with some backbone models. We can further conduct more experiments to verify the potential of our method in terms of fairness awareness on more temporal graph datasets.

4 Conclusion

In this work, we extend the investigation of individual fairness from static graphs to temporal graphs. We first revisit the current individual fairness metric for static graphs based on the (D_1, D_2)-Lipschitz property and then generalize it under the temporal setting by taking the time-aware fairness constraint into account. A new fairness requirement termed (K, D_1, D_2)-Lipschitz is presented accordingly based on the principle of individual fairness in retrospect. We also propose a novel bias measure with regard to individual fairness on dynamic graphs, together with a regularization-based fairness-aware method for dynamic GNNs. Some preliminary experimental results on two financial networks are shown to validate the promising efficacy of the proposed method. As an ongoing work, we will try to deal with the challenge of the varying number of nodes in the different snapshots, the training efficiency for large-scale real-world datasets, and the theoretical guarantee of the proposed method. We will also conduct more experiments on more temporal datasets, try other designs of the monotonically non-decreasing function $L(\cdot)$, and evaluate performance in the live-update training paradigm [15].

References

[1] Dawei Cheng, Xiaoyang Wang, Ying Zhang, and Liqing Zhang. Risk guarantee prediction in networked-loans. In *IJCAI*, pages 4483–4489. ijcai.org, 2020.

[2] Yushun Dong, Jian Kang, Hanghang Tong, and Jundong Li. Individual fairness for graph neural networks: A ranking based approach. In *KDD*, pages 300–310. ACM, 2021.

[3] Yushun Dong, Jing Ma, Chen Chen, and Jundong Li. Fairness in graph mining: A survey. *CoRR*, abs/2204.09888, 2022.

[4] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard S. Zemel. Fairness through awareness. In *ITCS*, pages 214–226. ACM, 2012.

[5] Ehsan Hajiramezanali, Arman Hasanzadeh, Krishna R. Narayanan, Nick Duffield, Mingyuan Zhou, and Xiaoning Qian. Variational graph recurrent neural networks. In *NeurIPS*, pages 10700–10710, 2019.

[6] Jian Kang, Jingrui He, Ross Maciejewski, and Hanghang Tong. Inform: Individual fairness on graph mining. In *KDD*, pages 379–389. ACM, 2020.

[7] Seyed Mehran Kazemi, Rishab Goel, Kshitij Jain, Ivan Kobyzev, Akshay Sethi, Peter Forsyth, and Pascal Poupart. Representation learning for dynamic graphs: A survey. *J. Mach. Learn. Res.*, 21:70:1–70:73, 2020.

[8] Srijan Kumar, Bryan Hooi, Disha Makhija, Mohit Kumar, Christos Faloutsos, and V. S. Subrahmanian. REV2: fraudulent user prediction in rating platforms. In *WSDM*, pages 333–341. ACM, 2018.

[9] Srijan Kumar, Francesca Spezzano, V. S. Subrahmanian, and Christos Faloutsos. Edge weight prediction in weighted signed networks. In *ICDM*, pages 221–230. IEEE Computer Society, 2016.

[10] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, Tao B. Schardl, and Charles E. Leiserson. Evolvegcn: Evolving graph convolutional networks for dynamic graphs. In *AAAI*, pages 5363–5370. AAAI Press, 2020.

[11] Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks. In *WSDM*, pages 519–527. ACM, 2020.

[12] Weihao Song, Yushun Dong, Ninghao Liu, and Jundong Li. GUIDE: group equality informed individual fairness in graph neural networks. In *KDD*, pages 1625–1634. ACM, 2022.

[13] Jianian Wang, Sheng Zhang, Yanghua Xiao, and Rui Song. A review on graph neural network methods in financial applications. *CoRR*, abs/2111.15367, 2021.

[14] Shuo Yang, Zhiqiang Zhang, Jun Zhou, Yang Wang, Wang Sun, Xingyu Zhong, Yanming Fang, Quan Yu, and Yuan Qi. Financial risk analysis for smes with graph-based supply chain mining. In *IJCAI*, pages 4661–4667. ijcai.org, 2020.

[15] Jiaxuan You, Tianyu Du, and Jure Leskovec. ROLAND: graph learning framework for dynamic graphs. In *KDD*, pages 2358–2366. ACM, 2022.