Network-based Active Learning for Identifying Illicit Actors in Financial Transaction Networks

Extended Abstract

Amro Alabsi Aljundi
University of Virginia
Charlottesville, United States
aaljundi@virginia.edu

Abhijin Adiga
University of Virginia
Charlottesville, United States
abhijin@gmail.com

Christopher Barrett
University of Virginia
Charlottesville, United States
clb5xe@virginia.edu

Margaret J. Foster
University of Virginia
Charlottesville, United States
avj6tm@virginia.edu

Brian D. Klahn
University of Virginia
Charlottesville, United States
briandk@virginia.edu

Dustin Machi
University of Virginia
Charlottesville, United States
dm8qs@virginia.edu

Achla Marathe
University of Virginia
Charlottesville, United States
achla@virginia.edu

Philip B.K. Potter
University of Virginia
Charlottesville, United States
phil@virginia.edu

Aaron Schroeder
University of Virginia
Charlottesville, United States
ads7fg@virginia.edu

Samarth Swarup
University of Virginia
Charlottesville, United States
swarup@virginia.edu

Anil Vullikanti
University of Virginia
Charlottesville, United States
vsakumar@virginia.edu

Mandy L Wilson
University of Virginia
Charlottesville, United States
alw4ey@virginia.edu

Madhav Marathe
University of Virginia
Charlottesville, United States
marathe@virginia.edu

ABSTRACT

Identifying illicit transactions within financial networks is an important area of research. Available datasets are highly imbalanced, making the design of machine learning methods challenging. Active learning, which carefully chooses data points for annotation, has been shown to improve performance for such problems. Here, we design a new approach, C^{2}AL, for detecting illicit nodes in financial networks, which incorporates network correlations more explicitly. Our approach builds on prior work on active learning on networks, specifically, collective classification (CC), which uses predicted labels of neighboring nodes to improve classification. We extend this approach by incorporating the information from both underlying models of collective classification, as well as their contrastive information, into the active learning sample selection procedure. We show that C^{2}AL significantly improves sample efficiency, requiring up to 48% fewer labeled samples than prior methods to achieve comparable detection performance across six financial network datasets.

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/GOCK1684

KEYWORDS

Active Learning; Collective classification; Cryptocurrency; Financial networks

ACM Reference Format:
Amro Alabsi Aljundi, Abhijin Adiga, Christopher Barrett, Margaret J. Foster, Brian D. Klahn, Dustin Machi, Achla Marathe, Philip B.K. Potter, Aaron Schroeder, Samarth Swarup, Anil Vullikanti, Mandy L Wilson, and Madhav Marathe. 2026. Network-based Active Learning for Identifying Illicit Actors in Financial Transaction Networks: Extended Abstract. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/GOCK1684

1 INTRODUCTION

Illicit financial activities pose a significant burden for the world economy. Detecting and stopping them are important challenges for financial and law enforcement agencies [9]. Today's complex financial systems and cryptocurrencies such as Bitcoin, which are inherently decentralized multi-agent systems [4, 6], make detection even harder. While supervised methods perform well when enough labeled data is available [8], generating ground truth labels is expensive, requiring expert analysts to investigate complex transaction patterns [13].

Active learning (AL) is a natural approach to reduce annotation costs [12], and has been applied to financial networks [7, 10, 11].

However, prior AL methods for financial networks use network structure only implicitly, as part of features used by the learner. More advanced techniques that use network structure explicitly have been developed for other domains [1, 3], but have not been applied to financial networks. The most relevant is the work of Bilgic et al. [3], who use collective classification, a technique that infers labels using neighbor information, and select nodes based on disagreement between the collective classifier, a content-only classifier, and network cluster majority classes.

Our contributions. (1) We design C^{2}AL, a new AL approach that integrates uncertainty scoring into collective classification-based active learning [3]. (2) We evaluate C^{2}AL on six financial network datasets: two real Bitcoin datasets (Elliptic++ [8]) and four synthetic datasets from AMLworld [2], an agent-based anti-money-laundering simulator. We show that C^{2}AL achieves up to 48% reduction in the number of labeled samples needed to reach target detection performance compared to baselines.

2 APPROACH

We represent financial datasets as directed, labeled networks $G = (V, E)$ where nodes represent entities (e.g., wallets, accounts) with features $\mathbf{x}_v$, and a subset of nodes have known labels $\ell(v) \in \{0, 1\}$ (non-illicit/illicit). Let $\mathcal{L} \subseteq V$ denote the set of labeled nodes and $\mathcal{U} = V \setminus \mathcal{L}$ the unlabeled nodes. Given $\mathcal{L}$ and an annotation budget B, our goal is to choose query nodes $Q \subset \mathcal{U}$, $|Q| \leq B$, so that an effective classifier can be trained using labels from $\mathcal{L} \cup Q$.

Collective classification. C^{2}AL uses two classifiers learned at each AL iteration. The *content-only* classifier (CO) is trained on labeled node features and predicts labels for all unlabeled nodes. The *collective classifier* (CC) augments each node's features with aggregated 1-hop neighborhood label information (specifically, the proportion of neighbors predicted as illicit and non-illicit (using both in- and out-neighbors)) and retrains on these enriched features. We use XGBoost [5] as the backbone model for both CO and CC, as it outperforms random forests, SVMs, and GNN architectures (GCN, GAT, GIN, PNA) on our datasets.

Sample selection. At each AL iteration, after learning CO and CC, C^{2}AL selects b nodes for annotation through a two-step process. First, it identifies the *disagreement set* D: nodes where CO and CC predict opposite classes. For each node v in D, C^{2}AL computes an aggregated uncertainty score $\gamma(v) = \frac{1}{2} \sum_{M \in \{CO, CC\}} H_M(v)$, where $H_M(v) = -\sum_{y \in \{0,1\}} \Pr_M(y \mid \mathbf{x}_v) \log \Pr_M(y \mid \mathbf{x}_v)$ is the entropy of model M's prediction for node v and $\Pr_M(y \mid \mathbf{x}_v)$ is the probability that model M assigns label y to sample v. Nodes are then sampled from D with probability proportional to $\gamma(v)$ (and, if budget remains, sampling continues from agreement set D^c using a similar distribution.) This differs from Bilgic et al. [3], who use disagreement as the sole scoring mechanism, and include the majority class of a node's network cluster as a third source of disagreement alongside CO and CC. In contrast, C^{2}AL uses clustering only for *diversity sampling*: selecting nodes from different network communities to maximize structural coverage, rather than as a disagreement signal. We find that using cluster majority class as a disagreement source degrades performance on highly imbalanced financial networks, since illicit nodes rarely form majority-illicit clusters.

Table 1: Labeled samples needed to reach ≥75% of target F1 (supervised XGBoost+CC). Lower is better. No method achieved the target for `LI-Medium`.

	Ell++T	Ell++W	HI-Medium	HI-Small	LI-Small
AL	640	6,820	1,420	1,180	OOB
AlfNet	1,280	8,020	1,060	780	OOB
C^{2}AL	**600**	**3,840**	**920***	**740***	**16,400**

OOB = Out of annotation budget. *Without clustering.

3 EXPERIMENTAL RESULTS

Setup. We evaluate on six datasets with varying scales and illicit node rates (0.7–2.2%). Those include two Bitcoin networks (the transaction and wallet networks `Ell++T` and `Ell++W` [8, 14]), and four AMLworld [2] synthetic networks (`HI-Small`, `LI-Small`, `HI-Medium`, `LI-Medium`). As features, we use a mix of inherent node features (e.g. balance of a transfer), aggregated edge features (e.g. total transactions), and structural features (e.g. centralities and subgraph counts.) We compare against two baselines: `AL` (standard uncertainty sampling with XGBoost) and `AlfNet` (the CC-based AL method of Bilgic et al. [3]). All experiments use a 90/10 train-test split, averaged over 5 random seeds.

Sample efficiency. Table 1 shows the number of labeled samples required to achieve ≥75% of the F1 score of a fully-supervised XGBoost+CC model. C^{2}AL outperforms both baselines across all datasets. On the challenging `LI-Small` dataset (0.8% illicit rate), C^{2}AL is the only method to reach the target within the annotation budget. These improvements are consistent across different thresholds, with mean relative improvements of 27%, 31%, and 25% at 60%, 75%, and 90% of target F1 respectively. Note that, at the 75% F1 threshold, C^{2}AL achieves precision of 69–89% across datasets, meaning the majority of flagged nodes are truly illicit.

Ablation. We find that uncertainty score aggregation from both CO and CC has the highest impact on performance. Disagreement filtering provides an additive benefit, particularly in early stages of AL. Cluster-based diversity sampling has a dataset-dependent effect: it helps on some networks but slightly hurts on others. Using undirected (both in- and out-neighbor) aggregation substantially outperforms directed variants, confirming that illicit actors operate as both senders and receivers in financial networks.

4 CONCLUSION

We presented C^{2}AL, an active learning approach for detecting illicit nodes in financial networks that incorporates network structure via collective classification combined with uncertainty scoring. Across six diverse financial network datasets, C^{2}AL achieves up to 48% reduction in the number of labeled samples needed to reach target detection performance. Key lessons from our study: (1) Cluster majority class, while used as a disagreement signal in prior work, degrades performance on highly imbalanced financial networks where illicit nodes rarely dominate any cluster. (2) Undirected neighborhood aggregation is critical; using only in- or out-neighbors misses important signals, as illicit actors operate on both sides of transactions. Future work includes extending C^{2}AL to handle noisy oracles and incorporating domain-specific patterns of illicit behavior into the selection strategy.

ACKNOWLEDGMENTS

This work is partially supported by University of Virginia Strategic Investment Fund Awards: SIF160, SIF186, the Contagion Science P2PE fund SIF176A. This research is also based on work supported in part by the Office of the Director of National Intelligence (ODNI) via Contract No. 2024-24070100001. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, or the US Government.

REFERENCES

[1] Charu C Aggarwal, Xiangnan Kong, Quanquan Gu, Jiawei Han, and Philip S Yu. 2014. Active learning: A survey. In *Data classification*. Chapman and Hall/CRC, 599–634.

[2] Erik Altman, Jovan Blanuša, Luc Von Niederhäusern, Béni Egressy, Andreea Anghel, and Kubilay Atasu. 2024. Realistic synthetic financial transactions for anti-money laundering models. *Advances in Neural Information Processing Systems* 36 (2024).

[3] Mustafa Bilgic, Lilyana Mihalkova, and Lise Getoor. 2010. Active learning for networked data. In *Proceedings of the 27th international conference on machine learning (ICML-10)*.

[4] Davide Calvaresi, Alevtina Dubovitskaya, Jean Paul Calbimonte, Kuldar Taveter, and Michael Schumacher. 2018. Multi-agent systems and blockchain: Results from a systematic literature review. In *International conference on practical applications of agents and multi-agent systems*. Springer, 110–126.

[5] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

[6] Luisanna Cocco, Roberto Tonelli, and Michele Marchesi. 2019. An agent-based artificial market model for studying the bitcoin trading. *IEEE Access* 7 (2019), 42908–42920.

[7] Leandro L Cunha, Miguel A Brito, Domingos F Oliveira, and Ana P Martins. 2023. Active Learning in the Detection of Anomalies in Cryptocurrency Transactions. *Machine Learning and Knowledge Extraction* 5, 4 (2023), 1717–1745.

[8] Youssef Elmougy and Ling Liu. 2023. Demystifying fraudulent transactions and illicit nodes in the bitcoin network for financial forensics. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 3979–3990.

[9] Rasmus Ingemann Tuffveson Jensen and Alexandros Iosifidis. 2023. Fighting money laundering with statistics and machine learning. *Ieee Access* 11 (2023), 8889–8903.

[10] Danilo Labanca, Luca Primerano, Marcus Markland-Montgomery, Mario Polino, Michele Carminati, and Stefano Zanero. 2022. Amaretto: An active learning framework for money laundering detection. *IEEE Access* 10 (2022), 41720–41739.

[11] Joana Lorenz, Maria Inês Silva, David Aparício, João Tiago Ascensão, and Pedro Bizarro. 2020. Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. In *Proceedings of the first ACM international conference on AI in finance*. 1–8.

[12] Burr Settles. 2009. Active learning literature survey. (2009).

[13] Michele Starnini, Charalampos E Tsourakakis, Maryam Zamanipour, André Panisson, Walter Allasia, Marco Fornasiero, Laura Li Puma, Valeria Ricci, Silvia Ronchiadin, Angela Ugrinoska, et al. 2021. Smurf-based anti-money laundering in time-evolving transaction networks. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part IV 21*. Springer, 171–186.

[14] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. 2019. Anti-money laundering in bitcoin: Experimenting with graph convolutional networks for financial forensics. *arXiv preprint arXiv:1908.02591* (2019).