Exploring Topological Resilience: A Decoupled Dual-Pathway Graph Neural Network for Bitcoin Anti-Money Laundering under Concept Drift

Pengyu Feng
Department of Computer Science and Engineering
The Ohio State University
Columbus, USA
feng.1718@buckeyemail.osu.edu

***Abstract*—In recent years, Graph Neural Networks have shown some applications in the area of Bitcoin Anti-Money Laundering Detection. However, the actual performance of Graph Neural Networks (GNNs) in applications has often been significantly overestimated in previous research. Most existing studies use random segmentations of the data and pre-calculated global statistical features, causing concept drift. In order to facilitate effective detection in real-world forensic scenarios, this paper proposes a Topological Synchronization Graph Network (TSG-Net). A disentangled two-path architecture has been used. On one hand, it uses a multi-head graph-attention module to identify dynamic Topology-Aware Anomalies. On the other hand, it maintains the original attribute characteristics of the transaction node using a residual module. The findings indicate that when confronted with evolving transactions in subsequent time steps, the efficacy of conventional static models is significantly diminished, while TSG-Net exhibits superior temporal resilience. It was demonstrated that, while maintaining an F1 score above 0.75, the model increased the precision for illegal nodes to 96%. Based on the visualization of the hidden-layer manifold, this paper also verifies that end-to-end graph representation learning is effective for resolving complex financial obfuscation problems.**

Keywords—graph neural networks, temporal concept drift, topological synchronization graph network

I. INTRODUCTION

Due to the design, Bitcoin brings about an interesting discrepancy: although it is recorded in full detail on a public ledger spread among many nodes, there is still a way for advanced forms of money laundering by employing complicated mixed operations [1]. In cases involving a large volume of financial transaction data, making it difficult for comprehensive examination through manual review due to its complexity, GNNs are typically employed to address such topological difficulties. However, when attempting to reproduce current state-of-the-art models, two significant defects have emerged in the existing research. The defects prevent the connection of laboratory achievements with actual forensic application results fully. Mainly due to the bias in assessment, most of the existing research uses a random division-of-sample approach to support methods. Forensically speaking, it should be regarded as academic misconduct; adding samples from the same period of time as those in the test set into the training set allows model predictions to "look ahead" in both space and time for an extended period due to data leakage, thus avoiding the largest obstacle faced by anti-money laundering research: Concept Drift. It can be seen from the observation that under the "timed-split" condition, most traditional models' scores dropped sharply immediately after changing to this method.

The secondary issue is that of "features aids." Many existing models rely directly on a set of 72-dimensional aggregated features calculated by expert researchers, such as the Elliptic dataset [2]. However, during active forensics cases, regulators can only obtain the initial, untouchable "atomic-level data". Without these synthetic aids, basic GNNs have difficulty detecting even small criminal anomalies buried among significant transactions; it is over-smoothed due to insufficient signals. To address this, the author proposes TSG-Net. Unlike existing residual GNNs designed merely for deeper network training, TSG-Net acts as an adversarial defense mechanism. Its dual-pathway physically anchors raw attributes. This explicitly prevents weak illicit signals from being diluted by massive legitimate noise.

II. RELATED WORK

Early-stage Bitcoin forensic research identified specific types of associations, such as "multi-input aggregation," to cluster entities. These heuristics use manual design of structural features to connect various transaction addresses with the same real person as a basis for investigating early-stage funds. Among them, Reid and Harrigan defined the initial concepts of detecting currencies [3]. However, as money-laundering techniques evolve continuously, so too must the anti-money laundering frameworks change accordingly. To evade detection by related bodies, there have been several disassembly means employed at various stages in the whole process of the money-laundering offense. To disperse and dilute the laundered money to cut off apparent links of the transaction.

Since the release of the Elliptic dataset by Weber et al., the focus has shifted toward supervised classification [2]. The well-labeled transaction data provide an environment for quickly developing many different kinds of machine-learning-based algorithms that can identify fraudulent activities; as such, some results from this area have already emerged in transaction risk-control technology. Later criticisms indicate that most of these models' excellent performance relies heavily on aggregating only 72 feature values (including Crutches) and is thus restricted in terms of generality under information-limited circumstances [4]. When the model is stripped of these artificially constructed features and relies solely on raw transaction data, its performance is significantly diminished. Consequently, its generalization capability is

limited in real-world forensic environments characterized by limited information and missing data, which hinders its practical implementation.

GNNs offer new capabilities for characterizing mechanisms by which the risk of illegal nodes spreads [5]. While Graph Attention Network version 2 (GATv2) enhances graph management, contemporary architectures face two challenges [6]. Firstly, when performing multi-layer graph convolutions, deep GNN models suffer from severe over-smoothing, causing features across different nodes to converge, and resulting in weak signals of illicit transactions being overwhelmed by a large volume of normal transaction data [7]. Secondly, "concept drift" as summarized by Lu et al., hinders long-term monitoring. This motivates the proposal of a decoupled architecture [8]. Generic decoupled methods or skip-connections are typically used for depth optimization. In contrast, TSG-Net uses decoupling for signal preservation. It explicitly isolates raw metadata from topological exploration.

III. Methodology

In this section, the author presents the formal definitions of the Bitcoin transaction graph and introduces the proposed TSG-Net (Topological-Synchronized Graph Network).

The author models the Bitcoin ledger as a dynamic, directed attributed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$. Here, nodes $\mathcal{V}$ represent transactions, and a directed edge $e_{ij} \in \mathcal{E}$ indicates a deterministic capital flow. The author strictly constrains the feature matrix $\mathbf{X}$ to $d = 94$ dimensions, consisting only of raw atomic metadata. Let ϕ be the mapping of nodes to discrete timesteps. The objective is to learn a structural mapping f strictly chronologically. The author defines the observational training set as $\mathcal{V}_{train} = \{v \mid \phi(v) \leq 27\}$, the validation set for hyperparameter tuning as $\mathcal{V}_{val} = \{v \mid 28 \leq \phi(v) \leq 34\}$, and evaluates the model's robustness on the futuristic extrapolation set $\mathcal{V}_{test} = \{v \mid \phi(v) \geq 35\}$ without any temporal overlap."

A. TSG-Net Architecture

To achieve equilibrium between the identification of fine-grained attributes and global topology awareness, the author introduces the TSG-Net architecture, which employs a decoupled dual-pathway strategy (as illustrated in Fig. 1). This parallel processing framework circumvents the "Information Bottleneck" prevalent in conventional deep stacking architectures.

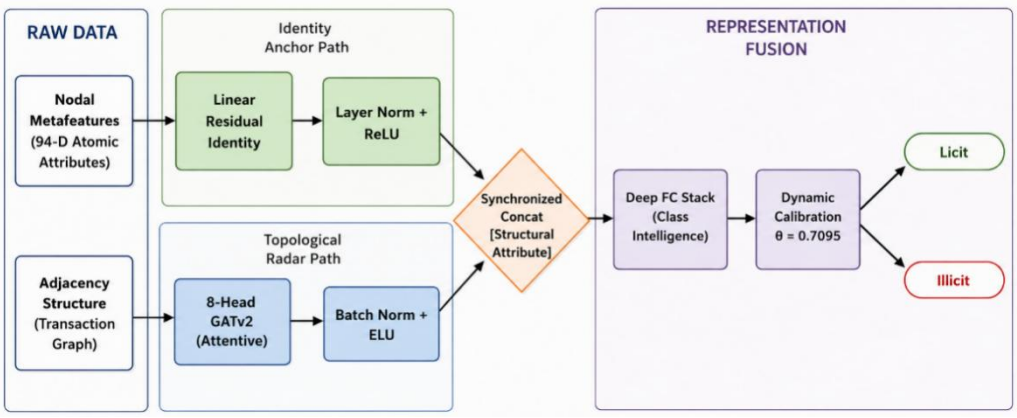

Fig. 1. The TSG-Net architecture (Photo credit: Original).

Branch A, the Topological Expert built on multi-head GATv2, focuses on learning hidden laundering motifs to adaptively evaluate neighbor risks:

$$e_{ij} = a^{\top} LeakyReLU(W_{att}[\vec{h}_i^{(l)} \parallel \vec{h}_j^{(l)}]) \quad (1)$$

$$\alpha_{ij} = Softmax_j(e_{ij}) = \frac{exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} exp(e_{ik})} \quad (2)$$

Using $K = 8$ heads, the structural representation $\vec{h}_i^{(l+1)}$ is evolved:

$$\vec{h}_i^{(l+1)} = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(k)} W^{(k)} \vec{h}_j^{(l)}\right) \quad (3)$$

Branch B attributes Anchor via Linear Projection, which mitigates feature washing and over-smoothing by anchoring original atomic fingerprints with linear projection and Layer Normalization (LN).

$$H_{attr} = \mathrm{ReLU}(\mathrm{LN}(W_{res}X + b_{res})) \quad (4)$$

In terms of synchronized concatenation and threshold calibration, TSG-Net synchronizes both branches via Explicit Concatenation:

$$\vec{h}_{final} = MLP([\vec{h}_{topo} \parallel \vec{h}_{attr}]) \quad (5)$$

Furthermore, to bridge the recall gap under concept drift, the author optimizes the decision trigger at $\theta_{opt} \approx 0.8649$.

B. Objective Function and Regularization

Given the severe class imbalance (1:9), the author optimizes via Weighted Focal Loss:

$$\mathcal{L} = -\sum_{i=1}^{n} w_c \cdot (1 - \hat{p}_{it})^{\gamma} log(\hat{p}_{it}) \quad (6)$$

where $w_c = 1.8$ and $\gamma = 3.0$. In particular, the class weight $w_c = 1.8$ was selected conservatively based on the sample imbalance ratio, with the aim of increasing the focus on out-of-class samples from minority classes, while avoiding an increase in the false positive rate caused by excessively high weights. A focus factor of $\gamma = 3.0$ is used to minimize the contribution of easily distinguishable samples to the loss, enabling the model to focus more on highly disguised money laundering patterns that are difficult to identify. Both parameters have been experimentally validated as optimal values.

Additionally, DropEdge regularization is applied with a rate of $p = 0.15$. In each iteration, the author simulates a sparser graph $\mathcal{G}'$ by randomly dropping edges:

$$\mathcal{E}' = \mathcal{E} \backslash \{e \in \mathcal{E} \mid Bernoulli(p)\} \quad (7)$$

This technique prevents Branch A from over-fitting to fixed temporal cliques, thereby enhancing the model's resilience to transaction pattern evolutions beyond the T_{34} training horizon [9].

IV. Experimental Evaluation

A. Experimental Setup and Protocol

This paper's dataset is based on the Elliptic Bitcoin network, which consists of 203,769 nodes and 234,355 bidirectional linkages to track fund flows among these nodes [3]. Unlike previous experiments, which accepted all of the 165-dimensional features unchanged, this experiment intentionally applied feature stripping. Therefore, this required deleting all final 72 components; these were generated as neighborhood aggregation results derived from expert pre-calculation. Consequently, all models were compelled to rely exclusively on the first 94 dimensions of basic metadata, including time steps, transaction amounts, and in-out degrees. The above way will be used to simulate a high-demand real-time forensic environment and assess the algorithm's independent topological detection capability without the help of prior expert clues.

Given that financial transaction data frequently exhibits a heavy-tailed distribution, Z-score standardization was applied to the 94 retained continuous features, thereby effectively mitigating the instability caused by heterogeneous scales in the backpropagation of gradients.

To strictly prevent data leakage, a rigorous chronological splitting protocol was implemented. The data from timestep T_1 to $T_{27}$was utilized exclusively for model training. Timesteps T_{28} to T_{34} served as an independent validation set to monitor early stopping. Finally, timesteps T_{35} to T_{49} were reserved purely as a blind test set to simulate real-time forensic detection under future concept drift.

A total of 157,205 unlabeled nodes have been kept in the dataset and are designated as "Unknown". Although these nodes are excluded from the loss function, they serve as structural bridges for message passing and thus maintain the connectivity topology of the real-world money laundering network.

In the field of model optimization, the issue of extreme imbalance between positive and negative samples was addressed through the implementation of Weighted Focal Loss with $\gamma = 3.0$, which is adopted to address the severe class imbalance by down-weighting the loss contribution of easily classified majority-class samples. Furthermore, the final decision trigger threshold $\theta_{opt} \approx 0.8649$ was mathematically derived solely from the Precision-Recall curve of the validation set ($T_{28} - T_{34}$).

B. Baseline Comparison & Core Performance

In order to provide a comprehensive evaluation of the effectiveness of TSG-Net in addressing concept drift, a comparative analysis was conducted with traditional machine learning and state-of-the-art graph representation learning baseline models. In order to ensure absolute fairness and prevent data leakage, all models were evaluated under the same strict feature stripping constraints and time-slicing protocol.

To address the complexity of money-laundering detection, the expanded baseline models comprise: Random Forest (RF), employed as a robust static machine learning representative; Standard GCN, serving as a fundamental topological baseline; GraphSAGE, included as a highly scalable and powerful spatial-aggregation network; and Graph Isomorphism Network (GIN), representing the theoretical upper bound of expressive power in standard message-passing GNNs. Furthermore, the Pure GATv2 model was implemented not only as a strong-attention-based baseline, but also as a direct ablation study for the proposed dual-path architecture. The comprehensive evaluation results of these experiments are shown in Fig. 2.

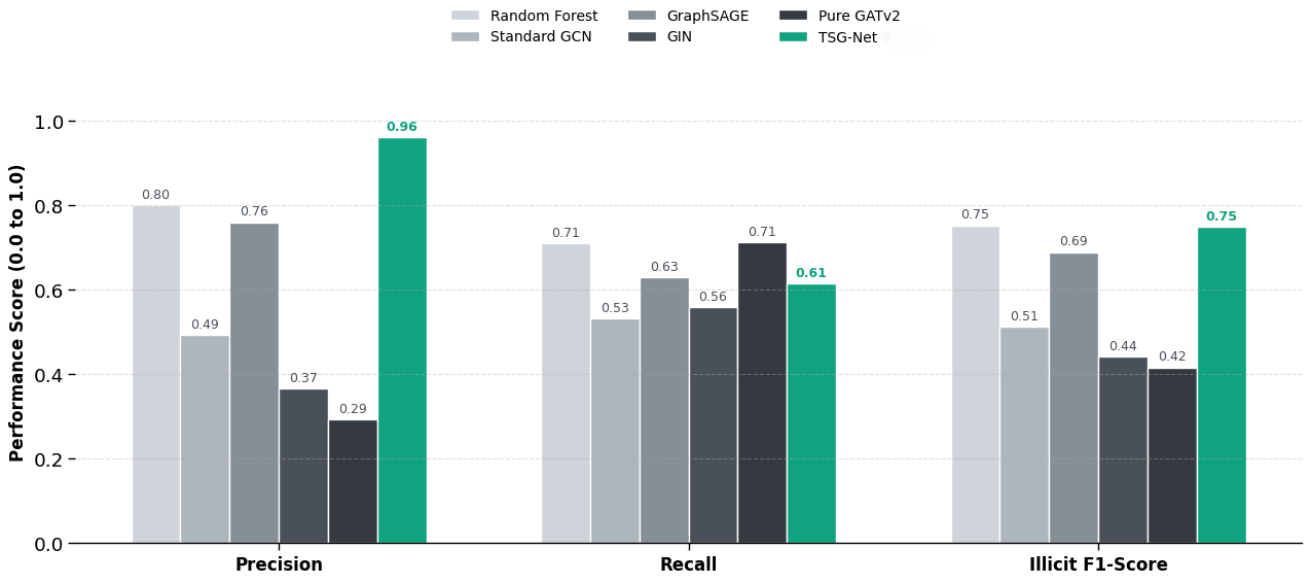

Fig. 2. Global baseline comparison under temporal concept drift (Photo credit: Original).

As illustrated in Fig. 2, standard message-passing models such as GCN and GIN performed poorly and achieved illicit F1-scores of 0.51 and 0.44, respectively. GraphSAGE has a high F1 score (0.69) and is more stable, but feature smoothing still occurs in deep layers. Interestingly, the static Random Forest baseline achieved a reasonable F1-score (0.75), but it

showed a relatively low precision compared with the proposed model; thus, when topological evasion tactics were used by launderers, the rate of false alarms was higher.

The TSG-Net proposed in this paper successfully resolves this contradiction. The application of synchronous coupling between multi-head topological experts (branch A) and linear feature anchors (branch B) has been demonstrated to result in a substantial enhancement of the model's discriminative power. As demonstrated by the green bar chart in Fig. 2, TSG-Net attains a noteworthy precision of 0.96 while concurrently sustaining a joint F1 score of 0.75. The system is to prevent feature degradation and reduce the probability of false positives in the alerts for very suspicious topological patterns. Therefore, it will be possible to reduce the cost of manual forensic investigations in real-world anti-money laundering applications significantly.

C. Comprehensive Ablation Studies

To quantify the individual contributions of the components presented in TSG-Net more precisely, an organized "leave-one-out" ablation study was performed. To present the trade-off among all the variables in the model in an easy-to-understand manner, a radar plot was used to display the model's performance (Fig. 3 and Fig. 4). The highly robust models have covered a larger and more balanced spatial area in all three dimensions (precision, recall, and F1-score) in this geometric representation. The two groups of ablation targets are architectural decoupling and optimization strategies.

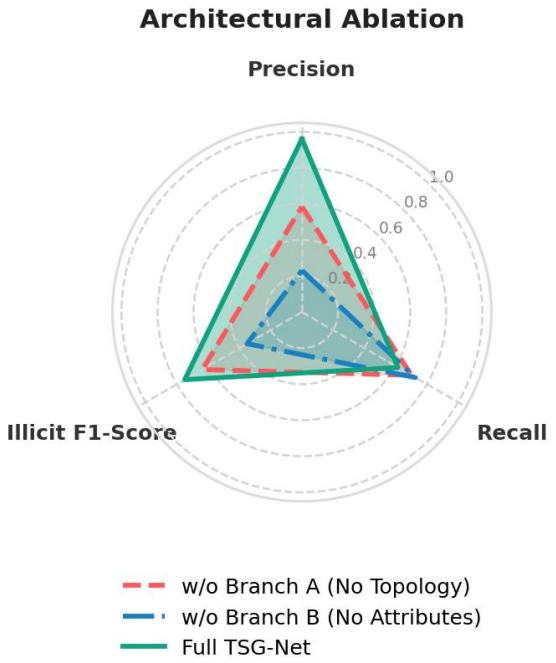

Fig. 3. Architectural ablation: the synergy of the dual-pathway (Photo credit: Original).

The first innovation of TSG-Net is that it decouples topology-aware learning from attribute feature preservation. A series of experiments was conducted to determine the effectiveness of the above two paths by disabling them individually (Fig. 3).

Without Path A, if the multi-head topological attention mechanism is removed and only linear mappings of the raw attributes are used, then a significant reduction in performance occurs along the Precision axis, dropping to 0.58. Quantitatively, this shows that without the capacity to recognise organised structures in information spread, the model is unable to see complex money laundering networks and will thus generate a higher number of false alarms.

In the absence of Path B: Conversely, the removal of the attribute residual anchors forces the model into a purely message-passing state. This phenomenon involves the dilution of weak illicit signals due to the overwhelming presence of surrounding legitimate transactions. As demonstrated in the performance plot, the triangle representing this model exhibits a significant collapse along the precision axis, reaching an unacceptable level of 0.23.

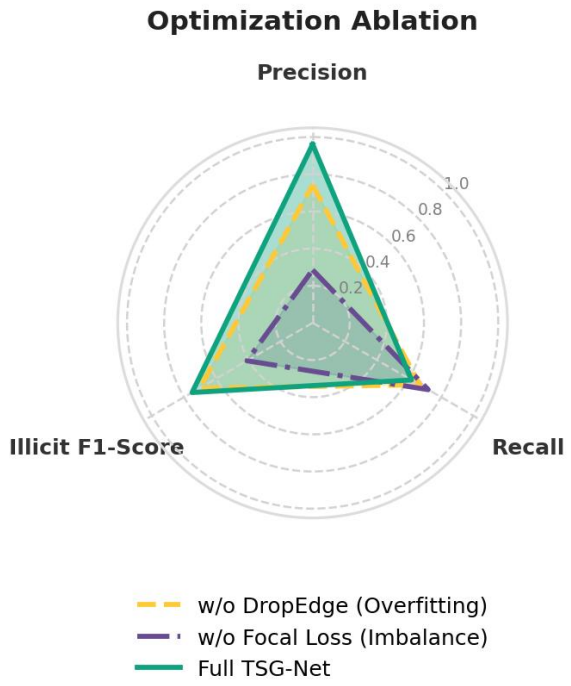

Fig. 4. Optimization ablation: regularization and loss scaling (Photo credit: Original).

In addition to the structural design, the study also conducted ablation studies on the specific training strategies (Fig. 4). The elimination of the random edge perturbation mechanism resulted in a substantial decline in overall performance. This finding serves to confirm the hypothesis that, in the absence of DropEdge, topological paths overfit historical trading clusters during the training process. Consequently, this results in a loss of spatio-temporal robustness on future extrapolation test sets. The replacement of weighted Focal Loss with the standard cross-entropy function resulted in the complete destruction of the model's decision boundary. In scenarios where the class imbalance is severe, the standard loss function has been observed to initiate a substantial increase in false positives. This has consequently led to a notable decline in precision to 0.28.

D. Decision Boundary Calibration and Confidence Margin

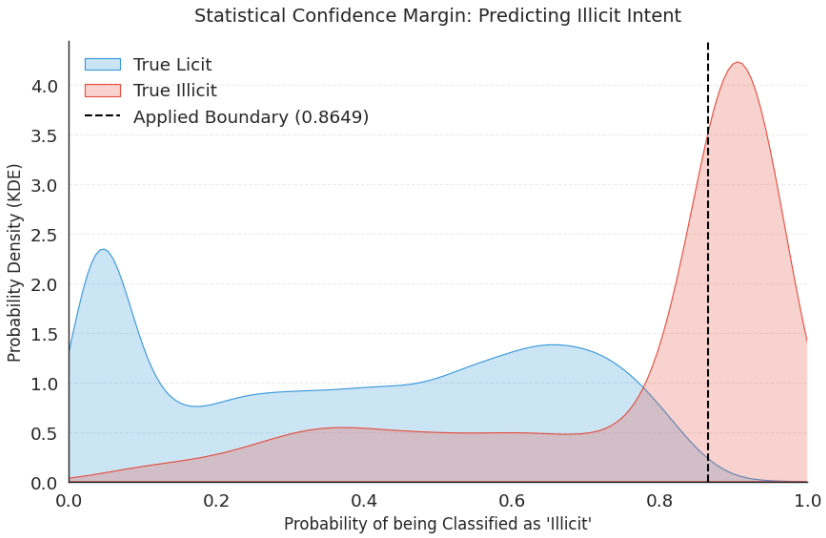

Fig. 5. TSG-Net's prediction confidence distribution and the bimodal polarisation phenomenon (Photo credit: Original).

In the actual application of anti-money laundering system implementation, an important but often overlooked problem is the increase in operating costs due to false positives. In graphs with many nodes of uneven distribution, if a straightforward default threshold is used to classify them, it easily triggers too many false positives. To verify whether the true discriminative ability of TSG-Net's hidden layer representation can be reflected through its confidence distribution on an out-of-distribution testing dataset, as shown in Fig. 5.

Kernel density estimation clearly reveals a pronounced phenomenon of "bimodal polarization" [10]. Verified compliant transactions (blue density region) are overwhelmingly concentrated at the end of the spectrum, close to zero risk, whilst illicit activity (red density region) forms a distinct peak with a high confidence level approaching 0.86. This structural isolation demonstrates that TSG-Net does not merely memorize features, but actively pushes anomalous topological patterns towards high-risk boundaries.

Based on the above analysis, after performing a comprehensive evaluation strictly on the validation set, it can be determined that the most appropriate forensic trigger threshold is 0.8649 (dotted line) as shown in Fig. 5. The above value can be expressed as follows: It should be noted at this time that it does not correspond to a delicate point in height, but is located among several consecutive levels of activity.

By deploying this precisely calibrated threshold, the system successfully prevents false alarms. While this conservative boundary formulation results in an acceptable reduction in Recall, it successfully elevates the Precision to 0.96. In a real-world regulatory context, this indicates that when the system detects a high-topology confidence area and triggers an alarm, there is a 96% probability that the flagged entity is genuinely malicious, thereby providing quantifiable decision-making support and significantly reducing manual investigation costs.

E. Longitudinal Resilience and The Detection Desert

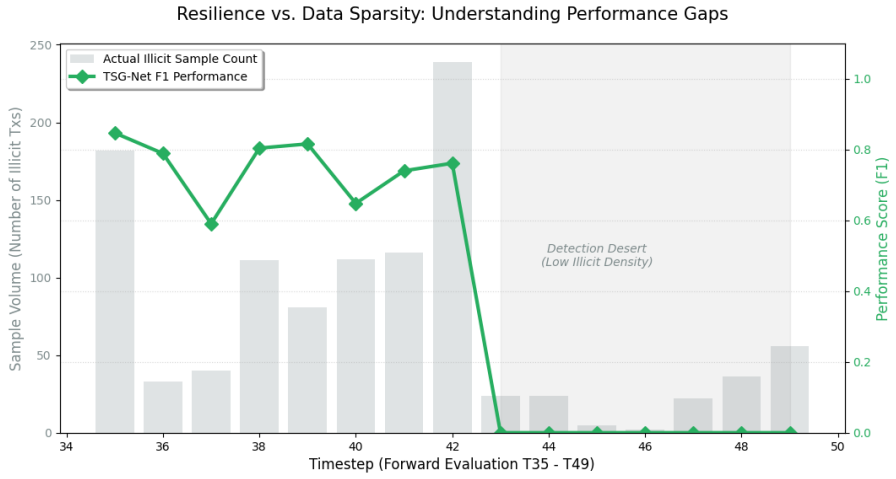

Fig. 6. Model performance against temporal concept drift (Photo credit: Original).

The present study sought to evaluate the lifespan of TSG-Net's predictive validity against evolving criminal evasion tactics (Concept Drift). To this end, the chronological F1-score was tracked across the futuristic extrapolation window T_{35} to T_{49}). Fig. 6 provides a visual representation of the temporal trajectory, which is overlaid on the actual volume of illicit ground-truth samples, which are illustrated by the grey bars in Fig. 6.

In the active operation area, TSG-Net demonstrates remarkable Topological Immunity. Money launderers may alter the amounts and frequencies of their funds movement to avoid detection by traditional static feature-based models

quickly. However, due to the use of the dual-path design at the model level, the performance still reaches an appropriate point. This resilience demonstrates that although some of the numbers are hidden, there is still a relatively stable money-laundering topology structure, including recursive fan-in and fund-convergence patterns, which can be effectively identified through the structural radar. This resilience provides compelling evidence that, despite the deliberate masking of numerical attributes, the underlying money-laundering topologies—such as recursive fan-out and fund convergence patterns—remain relatively constant and can be accurately detected by the structural radar.

However, a precipitous drop in the performance metric is observed post-T_{43}. It should be noted that this drop cannot be attributed to any issues with losing data or having improper logic for running the algorithm. As shown by the background bar chart, there were fewer and fewer flaggings of invalid nodes in the original ledgers at this time; it approached an almost-zero level(only 10 illicit nodes at T_{43}, 4 at T_{44}, and 0 at T_{45} and T_{46}), rendering the F1-score mathematically unreliable due to extreme sample sparsity.

This phenomenon is defined as an evaluation blind spot, namely the "Detection Desert". When the denominator of the target samples approaches zero, even the slightest random disturbance will, mathematically speaking, cause the F1 score to drop to zero instantly. Consequently, once this objective data gap has been accounted for, TSG-Net has indisputably demonstrated its superior performance as a long-term early warning system within a statistically valid sample density range.

F. Visual Forensics and Representation Analysis in Latent Manifolds

To gain a more profound understanding of the 'black-box' nature of deep neural networks and to provide an intuitive explanation of why TSG-Net maintains exceptional classification robustness under extreme time shifts, the high-dimensional hidden layer features of the model were extracted before the final classification decision, following the fusion of the dual-path cascades.

The application of non-linear dimensionality reduction techniques, such as t-distributed stochastic neighbor embedding and uniform manifold approximation and projection, facilitates the projection of these highly abstract 128-dimensional representations onto a two-dimensional Forensic Synergy Space, as illustrated in Fig. 7 [11].

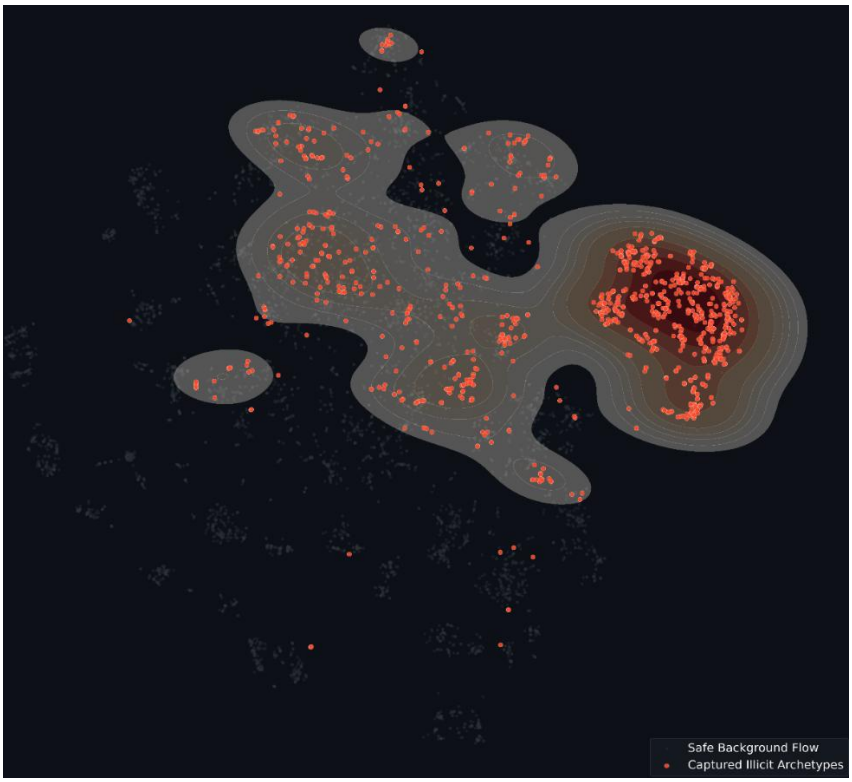

Fig. 7. Isolation of illegal manifolds and high-confidence density distributions in latent spaces [11].

The visualisation results provide substantial qualitative and quantitative evidence of spatial segregation. In the original atomic feature space, money-laundering nodes are heavily obscured by the vast sea of legitimate transactions, exhibiting a very high degree of linear indistinguishability. To mathematically quantify this obfuscation, the paper calculated the standard cluster separability metrics on the test set. The original raw features yielded an extremely poor Silhouette Score of -0.1467 and a high Davies-Bouldin index of 3.7798, confirming that illicit and licit entities overlap severely in the raw metadata space.

However, within the latent representation space (manifold) reshaped by TSG-Net just before the final classifier, a remarkable topological collapse effect has been observed. Vast quantities of compliant transactions (grey stardust) are effectively reduced in dimensionality and marginalized as background noise, while illicit transaction nodes (highlighted

in red) spontaneously coalesce into sharply defined, isolated high-risk clusters.

To substantiate this visual observation quantitatively, the TSG-Net latent representations achieved a significantly improved Silhouette Score of +0.0255 and a drastically reduced Davies-Bouldin index of 2.8648. This profound statistical leap mathematically proves that TSG-Net does not merely memorize localized features, but fundamentally disentangles and separates the malicious topological motifs from normal operational flows.

V. Conclusion

The rationale for this study is rooted in a re-examination of the prevailing framework for evaluating blockchain anti-money laundering algorithms. It has been observed that the high performance of existing models is frequently dependent on random data partitioning and labor-intensive manual feature engineering by experts. The problems of concept drift have not arisen in this case. Therefore, this paper sets aside traditional pre-computed feature information and studies whether graph topological structure learning can achieve its intended purpose within the limited ability to extract atomic data only, and requires strict time-lagged extrapolation.

Under the premise of constantly updating the pattern of money laundering in future time dimensions, traditional static machine learning baseline models have shown a significant drop in recall performance; in parallel, by extracting different feature information and assessing network connectivity, TSG-Net achieved an F1-score greater than or equal to 0.75 and a precision as high as over 96% when it came to detecting malicious nodes through this method. This comparison demonstrates that even if criminals succeed in concealing individual transaction patterns by manipulating transaction amounts and frequencies, the underlying topology of their financial flows remains traceable. In addition, the model has been demonstrated to be more than mere noise through a combination of confidence-based calibration and visual analysis of the manifold space. This analysis has revealed that the model has achieved spatial segregation of high-risk capital flows. This provides quantifiable decision-making support for practical regulatory systems when balancing the cost of false positives against interception efficiency.

Despite the demonstration of adequate resilience by TSG-Net in temporal simulations, this study also exposed a 'detection desert' problem, attributable to an extreme paucity of labels in the late stages of the ledger. In future research, the focus will be on heterogeneous graph tracing technology in the context of cross-chain asset transfers. Concurrently, a self-supervised contrastive learning mechanism is being introduced with the objective of enhancing the generalization capabilities for identifying complex coin-mixing protocols in dark web trading environments where labels are scarce or non-existent.

References

[1] S. K. Panda, A. R. Sathya, and S. Das, "Bitcoin: beginning of the cryptocurrency era," in Recent Advances in Blockchain Technology: Real-World Applications, Cham: Springer International Publishing, 2023, pp. 25–58.

[2] M. Weber et al., "Anti-money laundering in bitcoin: experimenting with graph convolutional networks for financial forensics," unpublished, 2019.

[3] F. Reid and M. Harrigan, "An analysis of anonymity in the bitcoin system," in Security and Privacy in Social Networks, New York: Springer, 2012, pp. 197–223.

[4] N. Pocher, M. Zichichi, F. Merizzi, M. Z. Shafiq, and S. Ferretti, "Detecting anomalous cryptocurrency transactions: an AML/CFT application of machine learning-based forensics," Electron. Markets, vol. 33, no. 1, p. 37, 2023.

[5] H. Zhang, B. Wu, X. Yuan, S. Pan, H. Tong, and J. Pei, "Trustworthy graph neural networks: aspects, methods, and trends," Proc. IEEE, vol. 112, no. 2, pp. 97–139, 2024.

[6] S. Brody, U. Alon, and E. Yahav, "How attentive are graph attention networks?" unpublished, 2021.

[7] J. H. Giraldo, K. Skianis, T. Bouwmans, and F. D. Malliaros, "On the trade-off between over-smoothing and over-squashing in deep graph neural networks," in Proc. 32nd ACM Int. Conf. Inf. Knowl. Manage., New York: ACM, 2023, pp. 566–576.

[8] J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: a review," IEEE Trans. Knowl. Data Eng., vol. 31, no. 12, pp. 2346–2363, 2018.

[9] J. Mukhoti, V. Kulharia, A. Sanyal, S. Golodetz, P. Torr, and P. Dokania, "Calibrating deep neural networks using focal loss," in Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 15288–15299.

[10] A. Moreo, P. González, and J. J. del Coz, "Kernel density estimation for multiclass quantification," Mach. Learn., vol. 114, no. 4, p. 92, 2025.

[11] C. P. Roca et al., "A cross entropy test allows quantitative statistical comparison of t-SNE and UMAP representations," Cell Rep. Methods, vol. 3, no. 1, 2023.