The object of the study is the processes of automated transaction classification and Bitcoin address identification for detecting malicious activity in conditions of pseudo-anonymity. The problem is the insufficient effectiveness of algorithms, such as graph convolutional networks, in conditions of strong class imbalance. This discovery is particularly important when less than ten percent of the data is clearly labelled. However, the main difficulty is excessive feature smoothing, which complicates the effective detection of anomalies for dense graphs. The results confirm that the Graph Attention Network v2 (GATv2) model is effective. It achieves an accuracy of 91.19% and an F1 score of 91.11% in testing. In addition, the stability of the approach is confirmed when 15% of topological noise is added to the graph structure. To prove the selectivity of the classifier, the Area Under the Curve (AUC) value of the approach is 0.889. The results are explained by the implementation of a dynamic anisotropic aggregation mechanism that adaptively distributes attention weights. This allows selectively amplifying weak signals of suspicious transactions while ignoring irrelevant connections and noise. A distinctive feature is the model of feature unification through logarithmic normalization of sums and non-linear processing of time intervals. Its uniqueness lies in the use of weighted loss functions and active learning strategies on boundary samples. Two-level transfer learning was applied to the Elliptic and BitcoinHeist datasets. The area of application is integration into real-time anti-money laundering (AML) systems. The approach allows overcoming conceptual shifts when new types of cyber threats emerge. The method detects the activity of CryptoLocker-type extortionists in the absence of data

Keywords: blockchain, Bitcoin-Heist, GATv2, graph neural networks, transfer learning, active learning

UDC 004.8:004.032.26:004.056.55
DOI: 10.15587/1729-4061.2026.351685

DEVELOPMENT OF A METHOD FOR IMPROVING THE EFFICIENCY OF TRANSACTION CLASSIFICATION IN THE BITCOIN NETWORK USING AN ATTENTION MECHANISM IN GRAPH NEURAL NETWORKS

Oleksandr Kushnerov
Doctor of Philosophy (PhD)
Department of Economic Cybernetics
Sumy State University
Kharkivska str., 116, Sumy, Ukraine, 40007
ORCID: https://orcid.org/0000-0001-8253-5698
Vladyslav Prosolov
PhD Student, Assistant*
ORCID: https://orcid.org/0009-0002-1276-4828
Valerii Dudykevych
Doctor of Technical Sciences, Professor
Department of Information Security
Lviv Polytechnic National University
S. Bandery str., 12, Lviv, Ukraine, 79013
ORCID: https://orcid.org/0000-0001-8827-9920
Serhii Yevseiev
Corresponding author
Doctor of Technical Sciences, Professor, Head of Department
Department of Cybersecurity
National Technical University "Kharkiv Polytechnic Institute"
Kyrpychva str., 2, Kharkiv, Ukraine, 61002
E-mail: Serhii.Yevseiev@gmail.com
ORCID: https://orcid.org/0000-0003-1647-6444
Serhii Povaliaiev
PhD, Associate Professor
Department of Machine Components and Theory of Machines and Mechanisms
Kharkiv National Automobile and Highway University
Yaroslava Mudroho str., 25, Kharkiv, Ukraine, 61002
ORCID: https://orcid.org/0000-0001-9027-0132
Yevheniia Ivanchenko
Doctor of Technical Sciences, Professor, Director
Educational-Scientific Institute of Cyber Security and Information Protection
State University of Information and Communication Technologies
Solomianska str., 7, Kyiv, Ukraine, 033110
ORCID: https://orcid.org/0000-0003-3017-5752
Volodymyr Gorbulyk
PhD, Associate Professor
Department of Radioengineering and Information Security
Yuriy Fedkovych Chernivtsi National University
Kotsiubynskoho str., 2, Chernivtsi, Ukraine, 58012
ORCID: https://orcid.org/0000-0001-6091-2261
Oleksandr Chechui
PhD, Associate Professor**
ORCID: https://orcid.org/0000-0002-7584-4457
Dmytro Balagura
PhD, Associate Professor*
ORCID: https://orcid.org/0009-0006-9839-3317
Vladyslav Sukhoteplyi
Senior Lecturer**
ORCID: https://orcid.org/0000-0002-2566-4167
*Department of Information Technology Security
Kharkiv National University of Radio Electronics
Nauky ave., 14, Kharkiv, Ukraine, 61166
**Department of Radioelectronic Systems of Control Points of Air Forces
Ivan Kozhedub Kharkiv National Air Force University
Sumska str., 77/79, Kharkiv, Ukraine, 61023

Received 12.11.2025
Received in revised form 26.01.2026
Accepted 02.02.2026
Published 27.02.2026

How to Cite: *Kushnerov, O., Prosolov, V., Dudykevych, V., Yevseiev, S., Povaliaiev, S., Ivanchenko, Y., Gorbulyk, V., Chechui, O., Balagura, D., Sukhoteplyi, V. (2026). Development of a method for improving the efficiency of transaction classification in the Bitcoin network using an attention mechanism in graph neural networks. Eastern-European Journal of Enterprise Technologies, 1 (9 (139)), ??–??. https://doi.org/10.15587/1729-4061.2026.351685*

1. Introduction

Detecting unusual transaction patterns in the Bitcoin network is an important security operation that combines cybersecurity work with financial monitoring responsibilities [1, 2]. The pseudo-anonymous nature of cryptocurrency addresses allows cybercriminals to use these digital assets as their primary financial resource for ransomware campaigns

that demand ransom payments in Bitcoin [1, 3]. Criminals use mixing services along with CoinJoin pools and multi-level "peeling" chains to hide the origin of their assets by creating fake stock market transactions [2, 4]. The current environment has shown that AML systems based on fixed rules and blacklists are unable to detect adaptive anonymization techniques [2, 5].

Blockchain data analysis has evolved from manual heuristic approaches to machine learning methods, but standard algorithms such as XGBoost and random forests do not take into account the structure of the transaction graph. This leads to the loss of important interaction details [5, 6]. The collected results demonstrate that graph neural networks (GNNs) that utilize structural data lead to a significant improvement in classification results [2, 7]. Most modern solutions use graph convolutional networks (GCNs), which average neighbor vectors to generate node representations using isotropic aggregation methods [6, 7]. This approach does not allow for the detection of suspicious connections, as it mixes these signals with the huge volume of transactions performed by high-level nodes, creating a structural noise problem [2, 4].

The GAT model solves the GCN limitations using an edge weighting system that applies different weights to edges depending on their direction [8]. The basic structure of the GAT network works with fixed attention weights that determine the ranking of neighbors without using any information about the state of the node query [9]. The GATv2 model has been modified with a dynamic attention system that allows the model to achieve better expressiveness and noise resistance [9, 10]. Research from 2025 shows that graph attention systems, combined with edge-aware analysis and adaptive augmentation methods, will prove effective in addressing class imbalance issues [11–13].

Despite existing progress, the research area exploring dynamic attention architectures for Bitcoin-specific topological models remains underdeveloped. The research concept focuses on developing a GATv2-based method that performs selective information aggregation. This allows the identification of malicious transactions in data with a high level of obscurity. According to sources [2, 9, 14], this approach will improve the effectiveness of AML monitoring compared to isotropic models.

The Bitcoin network requires adaptive financial monitoring systems, as its complex anonymization methods require architects to seek new architectural solutions. Standard isotropic models are not capable of effectively filtering structural noise, so the detection of malicious transactions depends on selective aggregation mechanisms based on GATv2. Therefore, research dedicated to developing a method for improving the efficiency of transaction classification in the Bitcoin network using an attention mechanism in graph neural networks is relevant.

2. Literature review and problem statement

Works [1, 3] present the results of fundamental research on the evolution of the Bitcoin network topology since 2009, focusing on the identification of clusters associated with ransomware activity. It is shown that the use of developed topological metrics (in particular, the characteristics of "loop", "number" and "strength" of connections) demonstrates higher efficiency in detecting malicious patterns compared to standard heuristics. This is because attackers form specific geometric patterns to obfuscate (confuse) asset flows. However, issues related to the low granularity of the analysis, which ignores the semantics of individual transactions and the attributes of unspent transaction outputs (UTXOs), remain unresolved. The reason for this is the high complexity of blockchain data aggregation and the inability to effectively engineer features due to pseudo-anonymity. This makes classical topological methods insufficiently flexible for operational monitoring tasks.

One way to overcome these difficulties is to use deep geometric learning methods (graph neural networks, GNN). This approach is used in [2, 15], where the superiority of graph convolutional networks (GCN) and inductive learning architectures over classical algorithms has been experimentally confirmed. However, these solutions remain limited due to the use of isotropic aggregation mechanisms. In [6], it was proven that the isotropic nature of GCN models leads to the levelling ("blurring") of anomalous signals in dense subgraphs, creating the effect of structural noise. Research [16] additionally points to the problem of "over-smoothing" when increasing the depth of the network, which leads to a loss of distinguishability of vector representations (embeddings) of nodes. In addition, the presence of a significant proportion of unmarked nodes ("unknown") creates the problem of "label incompleteness," which, in the context of a dynamic blockchain, leads to model degradation [2, 4].

The development of the theoretical basis of GNN has led to the emergence of architectures capable of operating in inductive mode, which is necessary for dynamic graphs. This approach is used in [15] (GraphSAGE – graph sample and aggregate), where instead of training fixed embeddings, it is proposed to train functions for aggregating information from the local environment. It has been shown that this allows for the effective processing of new nodes and edges that constantly appear in the Bitcoin network. However, issues related to sensitivity to local anomalies remain unresolved. The reason for this is that basic GraphSAGE aggregators, such as mean, pooling, or long short-term memory (LSTM) networks, remain isotropic in nature [17]. They assign equal or averaged weights to all neighbors of a node, making it fundamentally impossible for such an architecture to detect isolated suspicious transactions among thousands of legitimate connections between exchange wallets. This generates structural noise, which becomes a major obstacle to accurate classification in anti-money laundering (AML) systems [2, 18].

One way to overcome the difficulties associated with isotropy is to introduce an attention mechanism. In [14] and the classic article on graph attention networks (GAT) [8], the results of applying anisotropic neighbor weighting are presented. It was assumed that this would allow incoming transactions to be ranked according to their degree of suspicion. However, fundamental research [9] shows that the standard GAT architecture implements only a limited form of "static attention." It has been mathematically proven that attention weights in GAT depend monotonically on neighbor parameters and are virtually independent of the state of the query node. This means that it is fundamentally impossible to modify the importance values of neighbors according to the specific context of the transaction, which makes this approach ineffective for complex money laundering schemes. One solution to this problem is to use the GATv2 architecture proposed in [9]. This approach allows for the implementation of "dynamic attention" by changing the order of operations:

applying non-linear activation after concatenation, but before scalar projection. This provides a universal approximation of attention functions, allowing the model to adaptively filter noise [9, 10].

The works [11, 14] present the results of research on the latest architectures. In [11], it is shown that hybridization of GNN with neural ordinary differential equations (neural ODE) allows effective modelling of the temporal dynamics of network evolution. One solution to the problem of class imbalance (where illicit transactions account for less than 1%) is the heterogeneous graph adaptive augmentation (HGAA) method [12], which adaptively transforms graph patterns. Also, in [13] (TE-G-SAGE), approaches to improving the explainability of models are proposed. However, issues related to the selection of the optimal architecture depth and resistance to adversarial attacks remain unresolved. On the one hand, works [2, 7] argue that increasing the number of layers is necessary to cover distant connections (K-hop, K-step neighborhood). On the other hand, studies [16, 18] prove that this inevitably leads to oversmoothing, making the representation of nodes indistinguishable. The reason for this is the cost in terms of computational stability when working with deep graphs. Moreover, methods such as GrMA-CNN (graph-based modified attention with convolutional neural network) [14] and HGAA [12] are effective on static graphs. However, they lose stability under the influence of structural attacks aimed at creating fictitious connections. Current AML solutions for Bitcoin largely ignore the potential of dynamic attention to filter such topological noise.

All this gives reason to argue that it is advisable to conduct research dedicated to developing a method for improving the efficiency of transaction classification in the Bitcoin network using the GATv2 dynamic attention mechanism. This approach will resolve the trade-off between search depth and signal quality, providing adaptive filtering of structural noise and improving detection accuracy in conditions of critical class imbalance.

3. The aim and objectives of the study

The aim of the study is to improve the efficiency of transaction classification in the Bitcoin network by developing a method based on graph neural networks with a dynamic attention mechanism. This provides adaptive filtering of structural noise and resistance to class imbalance in the dynamic environment of the blockchain.

To achieve this aim, the following objectives were set:

– to build a mathematical model for unifying the feature space of transactions, ensuring the invariance of characteristics to the scale of sums and the intensity of network operations through the application of nonlinear transformations;

– to develop a graph neural network architecture based on GATv2Conv layers with dynamic multi-head attention for implementing adaptive ranking of node neighbors and selective filtering of structural noise;

– to justify the model training strategy in conditions of extreme class imbalance and implement regularization mechanisms to prevent the oversmoothing effect of node representations;

– to conduct experimental testing of the proposed method on real data sets to evaluate its discriminatory ability compared to existing isotropic and static anisotropic architectures.

4. Materials and methods

The object of the study is the processes of automated transaction classification and Bitcoin address identification to detect malicious activity in conditions of pseudo-anonymity. The classification analysis process solves three main tasks, namely: eliminating extreme imbalances between classes, managing significant structural noise, and adapting to the constant evolution of the graph structure.

The research hypothesis is based on a combined approach that combines the dynamic anisotropic aggregation model GATv2 with a unified feature space model. Multilevel knowledge transfer methods. The established method increases the selectivity of financial security due to its ability to amplify signals of malicious patterns while suppressing noise from loaded network hubs.

The following assumptions and simplifications were made in the study:

– the transaction graph maintains a constant topology within a single time step, providing a basis for applying spatial convolution methods;

– the elliptic dataset contains sufficient attributes to model user behavior patterns while preserving their anonymity;

– the training data contains class labels that the system accepts as true, even though the data may not contain information about new attack methods.

The study uses a two-level training system to apply its methodological approach, as this configuration allows for optimal use of the available labelled information. The graph neural network model begins its development from the initial configuration before undergoing the first training session, which uses the elliptic data set [2, 19] as the reference dataset. The dataset includes a partially anonymized Bitcoin transaction network containing 203,769 nodes representing transactions and 234,355 edges showing the connections between them in terms of payment flows. Each node contains 166 features, including 94 local data points and 72 aggregated data points from neighboring nodes. The dataset contains 49 separate time segments, allowing researchers to study how financial resources move over time [2, 16].

The second stage requires the application of transfer learning, which allows models to learn from data from the BitcoinHeist ransomware address dataset [3]. The dataset contains information about addresses and transactions associated with 29 different ransomware families, including CryptoLocker and CryptXXX. The BitcoinHeist system allows the model to be adapted by identifying actual topological attack patterns, which include both chain and star ransom collection structures. The "Elliptic → BitcoinHeist" training strategy allows trained filters to be adapted to the specifics of malicious campaigns without completely retraining the model. This is critical when there are a limited number of labelled samples for new types of attacks [1, 6].

The research base shows significant class differences, as this situation is often encountered in blockchain network cybersecurity cases. These two datasets require a special approach, as their unique features must be taken into account using special visualization methods that help to understand the distribution of classes and select appropriate regularization methods. The elliptic dataset presents its class distribution scheme using a graphical image, which is shown in Fig. 1.

A deep analysis of the data in Fig. 1 shows that the positive class "Illicit" with 4,545 nodes appears in much smaller numbers than the class "Licit", which con-

tains 42,019 nodes, and the unlabeled category "Unknown", which has 157,205 nodes. The uneven distribution between these classes requires a weighted loss function along with special oversampling techniques to process the data correctly. The large number of unlabeled nodes in this network suggests that knowledge transfer may be useful, as their structural connections reveal important details about the organization of the network.

This method requires fundamental development, starting with the creation of a mathematical model that combines the feature space. Basic transformations are applied to transaction sum data using logarithmic normalization, which standardizes

$$\tilde{a} = \log(1+a), \quad (1)$$

where a – the nominal value of the transaction in the Bitcoin network (Bitcoin, BTC). The transformation process ensures stable weight distribution, allowing for more efficient gradient descent. The system calculates two key metrics for each address, which include the transit coefficient r_v and the transaction frequency f_v, normalized to identify the functional role of the node as a mixer, storage, or transit point [1, 2, 16, 17].

The basic architectural structure of this system relies on a graph neural network that operates using the GATv2 dynamic attention mechanism. The GATv2 system operates with a query-dependent rating that works differently from traditional GCN systems, as it calculates attention weights a_{ij} using the following process

$$e_{ij} = a^T \text{LeakyReLU}\left(W\left[h_i \| h_j\right]\right), \quad (2)$$

where e_{ij} – the importance (attention) coefficient between nodes i and j; a – the vector of attention mechanism parameters; W – the linear transformation weight matrix; h_i, h_j – the feature vectors of nodes i and j, respectively; || – the vector concatenation operation.

The model achieves its goal through adaptive capabilities that allow it to identify suspicious business partners and ignore complex patterns that arise in normal customer transactions [8, 9]. The loss function contains weighting coefficients that solve the class distribution problem by assigning values that decrease as the frequency of the class increases. The weighted cross-entropy approach allows the system to better detect rare dangerous cases that require more attention. The model is regularized using Dropout methods, which use a probability of 0.4, and L2 regularization, which applies weight decay to the network weights.

The learning process monitoring system uses precision and recall values together with F1-score metrics, which it calculates based on the validation data set. The development of the loss function, which shows how well the selected optimization method and regularization technique work, is shown in Fig. 2.

A detailed evaluation of the results shown in Fig. 2 shows that the proposed method maintains stable performance throughout the entire training period. The loss curves for the training and validation datasets show a steady downward trend without any signs of model overfitting. The F1 score reached stability above 0.9 after the 30th epoch, demonstrating that the dynamic attention system effectively captures complex nonlinear relationships in the graph structure. The system learns to handle concept changes and new types of requesters through an active learning phase that applies oversampling techniques to positive examples in the latent feature space [4, 20]. The system achieves scalability through the implementation of neighbor sampling, which processes graphs that exceed the memory capacity of the GPU by working with subgraphs of fixed size [2, 15]. The software system uses Python (Python Software Foundation, USA) as its programming language and PyTorch Geometric (TU Dortmund University, Germany) as its main library. The described materials, together with the methods, create a unified technological platform that allows detecting malicious transactions in the blockchain while the blockchain network is operating in real time.

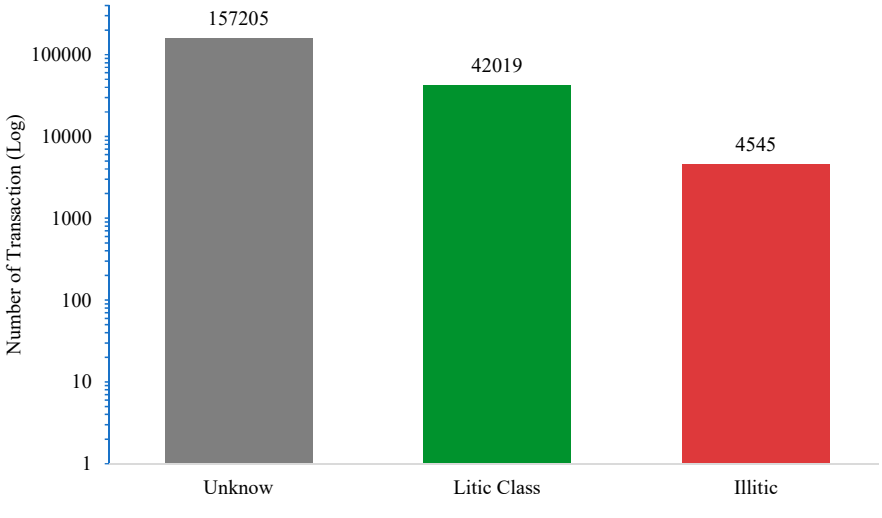

Fig. 1. Distribution of classes in the elliptic data set

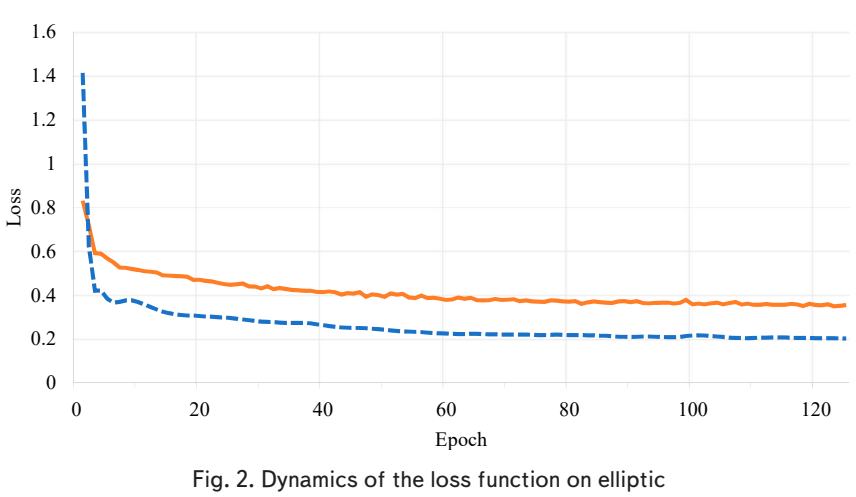

Fig. 2. Dynamics of the loss function on elliptic

5. Results of research into the classification of Bitcoin transactions using dynamic attention

5. 1. Mathematical model for the unification of transaction features with nonlinear transformations for large-scale invariance

The process of building an adaptive system for detecting unusual activities in decentralized payment networks requires a formal representation of the environment under study using a variety of directed multigraph models

$$G=\left(V,E,X_V,X_E\right), \tag{3}$$

where V – the set of vertices corresponding to unique addresses, and E – the set of edges representing asset transfer transactions. The main problem arises when deep learning methods encounter these structures, since their vertex characteristics $x_v \in R^{d_V}$ and characteristics of ribs $x_e \in R^{d_E}$ exist in completely different metric spaces. The attribute space representing addresses arises from a combination of statistical indicators that include validity period, total account balance, and network connectivity based on incoming and outgoing connections. The transaction space comes to life through real-time event parameters, which include transaction amounts, processing fees, and event timestamps. The Graph Attention Network v2 (GATv2) architecture requires a mathematical model for unification that transforms different data spaces into a single latent Hilbert space H with d_{hidden} dimensions while preserving semantic invariants [1, 6, 19]. The first stage of the model requires that the input data remain unchanged when the sum values change. Analysis of the Bitcoin transaction network based on the Elliptic dataset shows that financial sums follow a power-law distribution with extreme values ranging from 10^8 [2, 16]. The current system environment applies linear normalization, which creates a problem because it changes the gradient values during the process. The logarithmic compression technique with zero regularization is applied to the scalar values s

$$\tilde{s}=\ln\left(s+\epsilon\right), \tag{4}$$

where $\tilde{s}$ –the normalized sum value; s – the transaction amount; $\grave{U}$ – the regularization constant (a small shift to avoid a logarithm of zero). The system allows the model to effectively process both small financial transactions, including "dust attacks", and large money transfers. The system replaces absolute timestamps t with differential time increments, which employees must enter as $\Delta t_{ij} = t_j - t_i$ for hyperbolic normalization of processing.

$$\phi\left(\Delta t\right)=\tanh\left(\lambda\Delta t\right), \tag{5}$$

where Δt – the time interval between transactions; λ – the time scaling factor; tanh – the hyperbolic tangent function. This approach makes the model sensitive to transaction bursts generated by ransomware and mixers during their operation [3, 17]. The next task requires converting pre-processed vectors into a single embedding space. The system performs this function using linear mapping operations, which, in combination with nonlinear activation functions, produce the desired result:

$$h_v^{(0)}=\sigma\left(W_V x_v+b_V\right), \tag{6}$$

$$h_e^{(0)}=\sigma\left(W_E x_e+b_E\right), \tag{7}$$

where $W_V \in R^{d_{hidden}\times d_V}$ and $W_E \in R^{d_{hidden}\times d_E}$ – the matrices of weight coefficients that are learned [6, 15]. The activation function $\sigma(\,)$ uses LeakyReLU as its activation function to preserve negative signals, which help neurons function during the first training steps. The mathematical model focuses on a dynamic attention system that functions differently from a fixed GCN convolution, as it adjusts the values of each connected node and the corresponding transaction. The attention coefficient α_{ij}, which shows the significance of transactions from node j to node i, is calculated using a modified GATv2 mechanism as follows

$$\alpha_{ij}=\mathrm{softmax}_j\left(a^T\mathrm{LeakyReLU}\left(W\cdot\left[h_i\left\|h_j\right\|h_{e_{ij}}\right]\right)\right), \tag{8}$$

where a – the vector of attention parameters, || – the concatenation operation. The scoring system now includes the edge vector $h_{e_{ij}}$, which allows the model to identify transactions based on their structural location and complete content information [8, 9]. This solution solves the problem of over-smoothing, as standard GCN models lose important anomalous data in the process of averaging legitimate and suspicious neighboring features in dense graph structures [7, 18]. The process of updating the node representation at level $l + 1$ depends on weighted aggregation, which combines information from different

$$h_i^{(l+1)}=\sigma\left(\sum\nolimits_{j\in\mathcal{N}_i}\alpha_{ij}\cdot\left(W_{agg}h_j^{(l)}\right)\right), \tag{9}$$

where $h_i^{(l+1)}$ – the feature vector (embedding) of node i on layer $l + 1$. The function $\sigma(\cdot)$ is a nonlinear activation function (e.g., ELU or LeakyReLU), and N_i is the set of neighbors of node i. The normalized attention coefficient a_{ij} determines the level of importance of neighbor j for node i. The W_{agg} matrix contains trainable aggregation weights applied to the neighbor feature vector $h_j^{(l)}$ on the previous layer l. To stabilize training, multi-head attention with K independent mechanisms is used, the results of which are concatenated. Model parameters belonging to Θ are optimized by minimizing the weighted binary cross-entropy loss

$$\mathcal{L}\left(\Theta\right)=-\frac{1}{N}\sum\nolimits_{i=1}^{N}\begin{bmatrix}w_1 y_i\log\left(\widehat{y}_i\right)+\\+w_0\left(1-y_i\right)\log\left(1-\widehat{y}_i\right)\end{bmatrix}+\lambda\left|\Theta\right|_2^2, \tag{10}$$

where $\mathcal{L}\left(\Theta\right)$ – the value of the loss objective function; N – the total number of transactions in the training sample; y_i – the true class label for the i-th example (1 – illicit, 0 – licit); $\widehat{y}_i$ – the probability of belonging to the illicit class predicted by the model; w_1, w_0 – the weight coefficients of classes, calculated inversely proportional to their frequencies to compensate for imbalance; λ – the hyperparameter that regulates the strength of regularization; $\left|\Theta\right|_2^2$ – the L_2-norm of the model parameters (weight decay) to prevent overfitting. The elliptic and BitcoinHeist datasets require such a critical approach because the distribution of their classes shows a ratio exceeding 1:10. This approach allows model developers to create strong penalties that focus on false negatives rather than false positives [7, 8, 15]. The model trains its parameters using L_2 regularization and dropout, which operates on the matrix of attention coefficients a_{ij}. The system achieves better robustness to missing data and structural noise, as the model is trained based on multiple ensembles of subgraphs. The system demonstrated its ability to handle structural noise

and missing data during stress tests, which involved adding 15% random edges to the data [4, 5].

The model implements transfer learning through knowledge transfer between the Elliptic domain, which contains structured data, and the BitcoinHeist domain, which contains authentic ransomware templates. The domain adaptation task requires that the parameters Θ, which were learned from the source distribution P_s (X, Y), serve as a starting point for the target domain P_T (X, Y). The model retains its ability to identify topological patterns such as "detachment chains" and star-shaped branched structures because it uses a unified feature space, even though the statistical properties of individual sums vary [1, 10, 20]. The mathematical model developed provides a theoretical basis for creating an effective classifier that functions efficiently when the data contains a high level of uncertainty and diverse information structures, as well as when attackers actively attempt to interfere.

5. 2. Architecture of the second version of the graph neural network with a dynamic multi-channel attention mechanism

Developing an effective architecture for detecting anomalous activity in the Bitcoin network requires developers to move away from traditional isotropic data aggregation methods in favor of approaches that adapt to the specific context of network interactions. This method uses a graph neural network (GNN) that functions as the main structural element through a system of GATv2Conv (Graph Attention Network v2 Convolution) layers that work with dynamic anisotropic attention. The choice of this particular architecture solved the main problem faced by financial graphs, as legitimate transactions between exchanges and mixing services create a complex network that blocks the detection of malicious activity. The GATv2 model provides selective neighbor filtering, allowing it to focus on the most informative connections. Unlike spectral methods (GCN), this approach prevents the effect of over-smoothing in high-degree nodes. This is because GCNs function as low-pass filters, which often negate important local differences.

The neural network architecture includes a first stage that combines transaction attributes with address information by creating a shared latent space. The system processes transaction attributes and addresses in the initial stage, creating a combined latent space. The system uses several attention-based hidden layers that update node representations through recursive processing. The system uses a modified attention mechanism that generates edge weight values by analyzing the current state of the query node rather than following the standard GAT approach. The GAT architecture presented by Velichkovich contains a fixed attention system, as it ranks nodes based solely on their feature values, which do not change when connected to different nodes. The model has limitations in its ability to express complex patterns because it cannot handle nonlinear relationships involving "detachment chains," which show that nodes function differently depending on their position in the chain. The GATv2 architecture solves this problem with its method of operation, which applies a nonlinear activation function after linear transformation but before scalar projection of the evaluation space.

The operation of updating node feature vectors at each layer l creates $h_i^{(l+1)}$ by combining messages from all nodes that share an edge with node i in set N_i. The formula calculates the attention coefficient a_{ij}, which determines the importance of neighbor j for constructing node i

$$\alpha_{ij} = \frac{exp\left(a^T \text{LeakyReLU}\left(W \cdot \left[h_i \| h_j \| h_{e_{ij}}\right]\right)\right)}{\sum_{k \in \mathcal{N}_i} exp\left(a^T \text{LeakyReLU}\left(W \cdot \left[h_i \| h_k \| h_{e_{ik}}\right]\right)\right)}, \quad (11)$$

where W – the learning weight matrix, a – the vector of attention mechanism parameters, || – the concatenation operation, and $h_{e_{ij}}$ – the vector of transaction attributes (edges) connecting nodes.

The system requires that $h_{e_{ij}}$ functioned as an important architectural component, as it allows the network to distinguish transactions based on their topological structure and content, which includes the amount, time, and commission details. The proposed method adapts the basic neighbor aggregation function to a contextual filtering mechanism. This allows blocking low-value transactions ("dust attacks") from high-risk addresses. At the same time, the system identifies hidden connections by analyzing dynamic time patterns. The system learns to recognize multiple features using multi-head attention, which ensures the stability of the learning process when processing financial stream data and time correlation models. The model operates using K independent attention systems, which simultaneously calculate their specific sets of coefficients $\alpha_{ij}^{(k)}$ and transformed features. The system combines results from different heads by concatenation at intermediate layers and averaging at the final layer to create an improved representation

$$h_i' = \|_{k=1}^{K} \sigma\left(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(k)} W^{(k)} h_j\right), \quad (12)$$

where h_i' – the updated feature vector of node i; || – the concatenation operation of vectors from $k = 1$ to K; K – the total number of heads of attention; $\sigma(\cdot)$ – the nonlinear activation function; N_i – the set of neighbors of a node i; $\alpha_{ij}^{(k)}$ – the normalized attention coefficient of the k-th head, which determines the importance of neighbor j for node i; $W^{(k)}$ – the weight matrix of the linear transformation for the k-th head; h_j – the vector of neighboring node features j. The system allows each "head" to develop expertise in specific categories of patterns on which they will focus. The system allows certain channels to detect star patterns that appear during redemption collection, but other channels control chain transactions using mixing operations. The model achieves better resistance to structural noise because it allows one channel to generate false activations that other channels can offset with their correct operations.

The system design focuses on solving scaling issues that arise when multiple nodes in the network receive large amounts of traffic. The Bitcoin network structure exhibits signs of "heavy tails" as certain addresses support thousands of connections according to the Elliptic dataset distribution [2, 16]. The standard GCN model creates feature pollution because these nodes produce averaged vectors that obscure information from their closest neighboring nodes. The architecture solves this problem with a dynamic attention system that reduces the influence of connections that are not important. The model training process automatically reduces the weight values a_{ij} that connect the target node to legitimate hubs, including exchanges, when transaction characteristics do not match suspicious activity patterns. The system maintains node-specific characteristics in dense network environments, allowing it to protect important information about abnormal behavior patterns.

The network contains a fully connected multilayer perceptron (MLP) on the output layer, which processes node

feature vectors to generate class predictions for the categories "licit" and "illicit". The system contains dropout layers with a neuron deactivation probability of 0.4, which protect against overfitting by working with both attention coefficient matrices and linear layer outputs. The system is trained using a subgraph ensemble approach, which creates resistance to attacks on the graph structure, including the addition of false connections to protect the network [8, 15].

The proposed architecture demonstrates its effectiveness through theoretical analysis comparing it to fundamental models. GCN operates as a fixed Laplacian smoothing filter, but GATv2Conv functions as an adaptive nonlinear operator that can learn any graph-based function. The AML task requires flexible detection methods, as signs of criminal activity manifest themselves in small differences between transaction distributions rather than in general network patterns. A multi-channel attention system produces results that experts can interpret, as they can visualize the sources of transactions by analyzing the weight distribution between the a_{ij} values, which cover different time periods of financial monitoring systems [9, 18].

The GATv2Conv architecture functions as a powerful analytical tool because it implements a dynamic multi-channel attention system that combines edge attribute data thanks to its design. The system uses neural networks to learn representations, which it then applies to process relational data to solve classification problems related to the heterogeneous, unbalanced, and noisy datasets that characterize the modern Bitcoin ecosystem.

5. 3. Model training strategy in the case of extreme class imbalance and regularization mechanisms against excessive smoothing

Developing a GNN training strategy for financial monitoring requires consideration of the specific patterns of the Bitcoin network. Only a small fraction of transactions is related to money laundering or cybercrime compared to legitimate transactions. The Elliptic dataset shows a critical imbalance, with illegal activity accounting for less than 10% of the labelled data. In real-world flows, this figure is less than one percent. This creates the risk of the model being biased towards predicting only the majority class. The standard cross-entropy loss function does not work properly because the total classification error from a few criminal transactions becomes insignificant compared to the large number of correct classifications for legitimate transactions. The proposed method solves this problem with a training strategy that applies weighted binary cross-entropy and calculates class weight coefficients w_c using the inverse of their frequency in the training data. The mathematical process applies a more severe penalty to gradient descent when it fails to detect an attack than when it produces a false alarm during normal transaction processing. The proposed method shifts the operating point of the classifier to improve recall. This provides the "conservative" error profile required for anti-money laundering (AML) scenarios [6, 8, 15]. This strategy is driven by the fact that the losses from missed crimes far outweigh the costs of verifying false positives.

The main problem with deep neural networks arises because their node representations become indistinguishable due to excessive smoothing, which occurs when multiple network layers perform continuous aggregation of neighboring features. Malicious node features become invisible as they merge with the normal features of nodes belonging to their network connections, which include major hub nodes such as stock exchanges and mixing services. The main architectures that include GCN depend on aggregation mechanisms, which, according to sources [16, 18], function as low-pass filters. The proposed strategy uses the GATv2 architecture as a structural basis that prevents excessive smoothing in the system. The model learns to assign near-zero weights to unnecessary connections to hubs using a dynamic attention mechanism that functions to eliminate noise during the aggregation process. The system operates with two main components, consisting of L_2 regularization for network weights and dropout layers, which are activated with a probability of 40 to 50 per cent. The dropout mechanism performs two main functions in this situation, as it prevents the classifier from overfitting to certain features of the training samples and works with the adjacency matrix to simulate learning from multiple sparse subgraphs. The model must develop this capability because attackers use structural attacks that create deceptive transaction chains to conceal their malicious activities [4, 5, 7].

The training process takes place using a two-stage transfer learning system, which solves the problem of limited labelled data for new attack categories. The model completes the entire training process in 120 epochs, which it spends training using the Elliptic Data Set reference graph [2, 19]. The dataset contains anonymized data but retains two important elements that allow the model to understand basic financial flow patterns. The dataset contains 49 time steps, allowing the model to understand basic financial patterns through repetitive cycles and movements between points. The batch subgraph selection process, called Neighbor Sampling, allows graphs exceeding the memory capacity of the GPU to be processed, as it supports the local connections necessary for correct attention calculations. During the initial training cycles, the loss function shows rapid positive dynamics. The training error decreased from 0.8 to 0.6, and the validation error decreased from 1.4 to 0.4. This confirms the effectiveness of the selected weight initialization scheme and input data normalization methods. The validation error does not show the typical U-shaped growth after the 30th epoch. This proves the effectiveness of regularization methods (L2 and dropout) in preventing overfitting. As a result, the model retains a high ability to generalize at new time intervals [6, 7, 15].

The second stage of the strategy requires continuing to train the previously developed model through fine-tuning, which uses the Bitcoin Heist Ransomware Address dataset [1, 3]. This process is necessary at this stage because Elliptic provides broad categories of illegal activities that differ from what is needed to detect specific ransomware samples, including the CryptoLocker and Princeton families. The training strategy requires that the weights of the lower layers remain frozen, as these layers have learned to create effective transaction embeddings, but the system will only modify the attention layers and the classification block. The system allows knowledge about structural invariants to be transferred, including patterns of decoupling chains between the data-rich elliptic domain and the BitcoinHeist domain, which has limited sample availability for each malware family. The metric dynamics graphs show that the model's performance improves during the training process, as the accuracy and F1 score metrics increase after training begins, but the system retains its prior knowledge. Sequential learning allows models to develop both reliability from working with large

datasets and specialized knowledge in processing specific models [10, 17, 20].

The training strategy includes active learning components that help combat conceptual drift, as cybercriminals are constantly changing their attack methods. The system must update its model by regularly expanding training samples, which includes proven attack examples from experts as new support vectors. The system uses oversampling, which creates additional examples of minority classes by duplicating feature space and batch replication. The system must detect new attack patterns as it processes all historical data, including every pattern that has appeared before. Experimental data show that this adaptive method supports a stable increase in the F1 metric, which reaches 0.911, and maintains an accuracy of 0.912 in noisy data conditions, outperforming basic isotropic methods such as GCN [1, 3, 4]. The system becomes noise-resistant because it combines a weighted loss function with GATv2 anisotropic aggregation, transfer learning, and regularization to work effectively in real blockchain environments.

5. 4. Experimental verification of the proposed method on reference data sets

The process of testing the two-level transaction classification method was to be carried out in two stages: first, financial flow models were studied, and then the team moved on to recognizing ransomware behavior patterns. Training begins with a basic stage based on the elliptic sample, followed by fine-tuning using the Bitcoin Heist Ransomware Address set. This approach ensures classification stability under conditions of limited data visibility and high noise levels. The first stage of the project involved training the GATv2 architecture from start to finish on an elliptical transaction graph, which used nodes to represent individual transactions and edges to show asset transfer paths. The training process lasted 120 epochs, during which mini-batch sampling of subgraphs was performed and weighted binary cross-entropy was used to manage class distribution issues, and the dropout technique with L2 regularization was used to train the model.

The learning process monitoring system demonstrated high algorithm stability thanks to a detailed analysis of the loss function development in Fig. 2. The initial training epochs showed rapid convergence, which then stabilized at a stable plateau, demonstrating the effectiveness of the selected regularization methods. The absence of the typical U-shaped growth of errors in the final stages of training proves that the model has learned to generalize characteristic patterns without falling into the trap of memorizing random noise. The stability of the loss function corresponds to the classification results, as the quality indicators presented in Fig. 3 demonstrate this dependence.

A detailed analysis of the curves in Fig. 3 shows that the classification accuracy exceeds random guessing during the first 10–15 epochs, reaching values from 0.8 to 0.85. The F1-score for the "illicit" class starts at 0.27 and increases to 0.52 during this period. The model learns the feature hierarchy, as the F1 score remains stable at 0.67 and the accuracy scores fluctuate between 0.88 and 0.9. The model learns complex patterns through correlations between transaction sums and network structure patterns, which it applies to its training data. The metrics show small random variations during the plateau phase, as the model is trained on unbalanced data using a subgraph sample, but the overall upward trend shows that the GATv2 architecture creates strong node representations. The main element that hinders network classification in Bitcoin systems requires researchers to develop anisotropic models due to its presence. The trained network demonstrates various structural elements that become visible thanks to the node degree distribution histogram shown in Fig. 4.

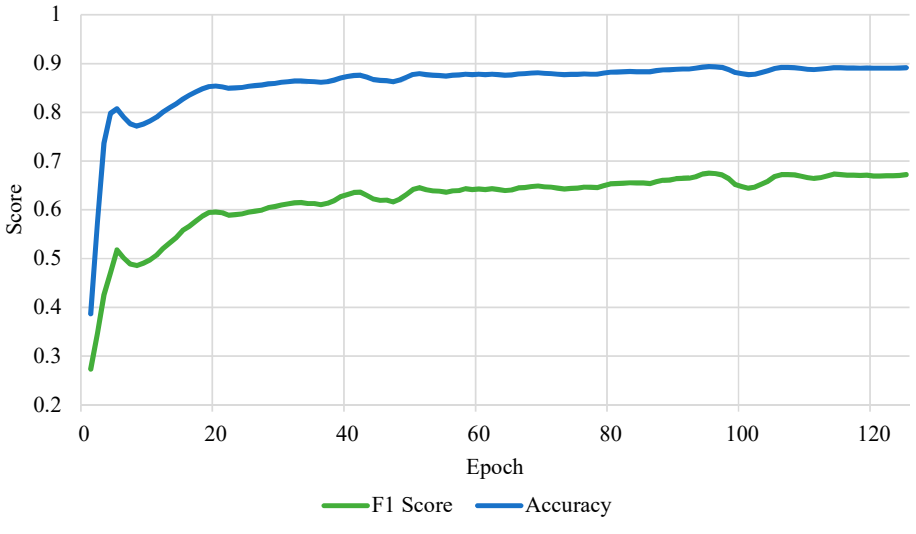

Fig. 3. Validation metrics on elliptic

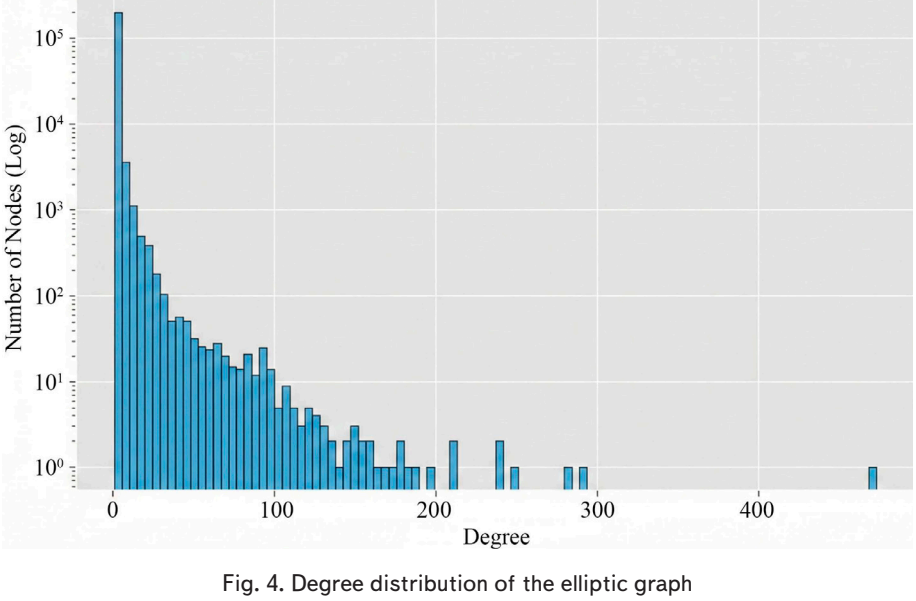

Fig. 4. Degree distribution of the elliptic graph

The histogram in Fig. 4 illustrates the typical structure of scale-free networks. Most transactions have between 1 and 5 connections, while individual hubs cover more than 400 nodes. The significant variability in node degrees is a critical problem for traditional GCN models. Their isotropic averaging mechanism causes weak anomalous signals to be levelled out by absorption into large hub flows (the over-smoothing effect). The GATv2 architecture provides dynamic attention that filters input signals by assigning low weights to irrelevant hub connections, enabling it to detect subtle anomalous patterns in dense exchange and mixing environments.

The model's ability to reduce risks during actual financial monitoring activities became evident thanks to the error profile that emerged as a result of testing on the test database. The distribution of classification results becomes apparent thanks to the confusion matrix shown in Fig. 5.

The quantitative data presented in Fig. 5 demonstrates that the decision-making process has achieved its goal, which corresponds to a conservative approach. The model successfully identifies most authentic transactions, numbering approximately 12,819 cases in the "True Negative" category, and also detects approximately 868 cases of criminal activity through "True Positive" detection. The system achieves its primary goal by identifying 215 attacks, which corresponds to the minimum acceptable detection level. This is due to the priority of minimizing missed crimes within financial monitoring. In AML scenarios, the regulatory and reputational risks of undetected activity significantly outweigh the operational costs of verifying false positives. The system generates 2,768 false positive results, which is within the range that allows compliance departments to process alerts through a manual verification process without overloading the operating system.

The model's ability to distinguish between legitimate and illegitimate classes across all classification thresholds is visible in the ROC curve shown in Fig. 6.

The curve in Fig. 6 passes significantly above the diagonal of random guess, demonstrating the excellent discriminatory power of the classifier with an AUC value of approximately 0.889. The nature of the curve indicates a high sensitivity of the model with a true positive rate of over 0.9 and a moderate level of false positives. This allows the working threshold of the classifier to be adapted to the specific requirements of a financial institution's risk management system.

The developers conducted a stress test to assess the system's ability to withstand active interference from malicious users attempting to hide their connections using mixers and incomplete data during testing. The basic GCN architecture underwent a comparative stress test, which involved adding 15% random topological noise to assess its resistance to active malicious interference using mixers to hide connections and incomplete data scenarios. The research paper presents the results of the comparison of metrics in Table 1.

Table 1

Comparison of architecture effectiveness on noisy elliptic data

Model	Accuracy	F1-Score	Precision	Recall
GCN (baseline)	0.906	0.906	0.906	0.906
GATv2 (proposed)	0.912	0.911	0.912	0.912

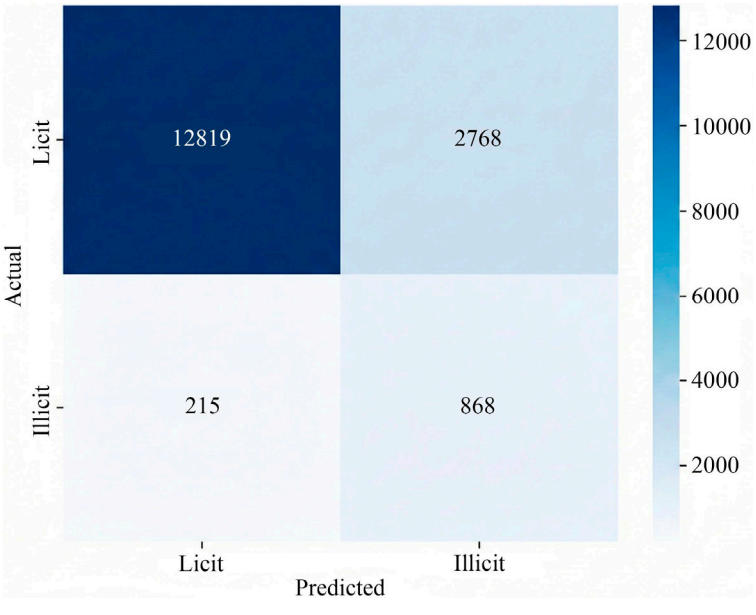

Fig. 5. Confusion matrix on elliptic

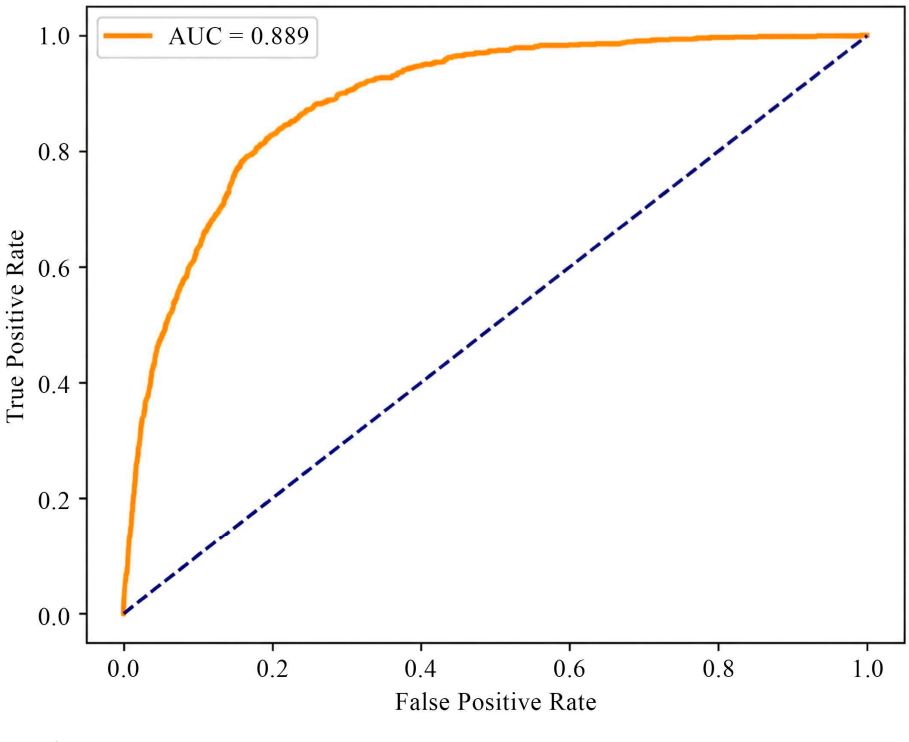

Fig. 6. Receiver operating characteristic curve for classification results of the elliptic dataset

The data in Table 1 show that the GATv2 architecture outperforms the baseline GCN model by approximately 0.6 percentage points across all key metrics. The experimental results confirm the theoretical assumption that anisotropic feature aggregation methods demonstrate better resistance to topological perturbations. The results show their advantage in the diagram shown in Fig. 7.

Analysis of the diagram in Fig. 7 demonstrates that GATv2 improves classification quality by simultaneously increasing all components of the F1 score, rather than by compromising between accuracy and reproducibility. The attention mechanism proves its ability to block false noise connections as it focuses on the main structural patterns that remain unchanged during anonymization processes. The second stage of the experiment involved a transfer learning approach, in which a realistic Bitcoin Heist Ransomware Address dataset was used for training. The process of adapting the loss function between baseline training and target domain training is visible in the diagram shown in Fig. 8.

The vertical dotted line in Fig. 8 marks the start of fine-tuning. It is noteworthy that changing the data domain does not cause a sharp jump in error or a catastrophic forgetting effect. The training process continues after the dotted line, leading to better convergence until the validation curve shows an additional decrease at the end of the training phase. The study shows that the unified feature space used by elliptic contains common patterns that effectively detect threats from ransomware. The impact of additional training on quality metrics is evident from the information presented in Fig. 9.

The graphs in Fig. 9 show how the metrics reach the baseline training level up to the transition point, which is shown by the dotted line. The adaptation of Bitcoin Heist brought a noticeable improvement in the F1-score and accuracy metrics. The decrease in loss function values demonstrates that the model has successfully learned to identify specific ransomware patterns, which include both peeling chains and star structures for collecting ransoms.

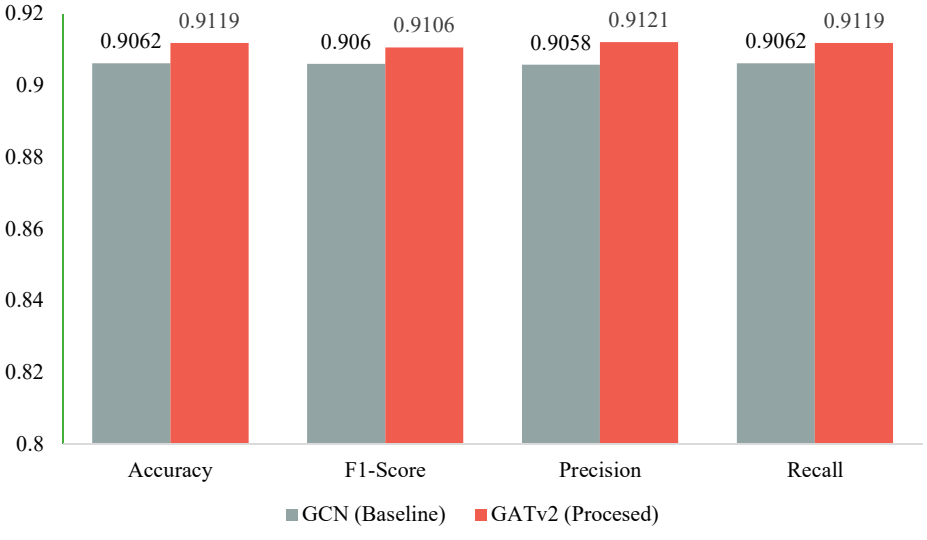

Fig. 7. Comparative diagram of the efficiency of the graph convolutional network and the graph attention network of the second version

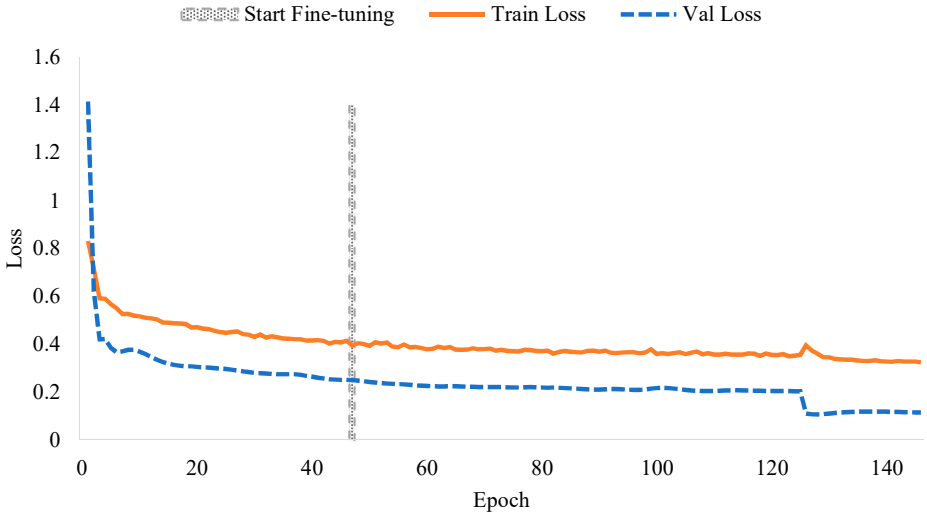

Fig. 8. Dynamics of losses with additional training

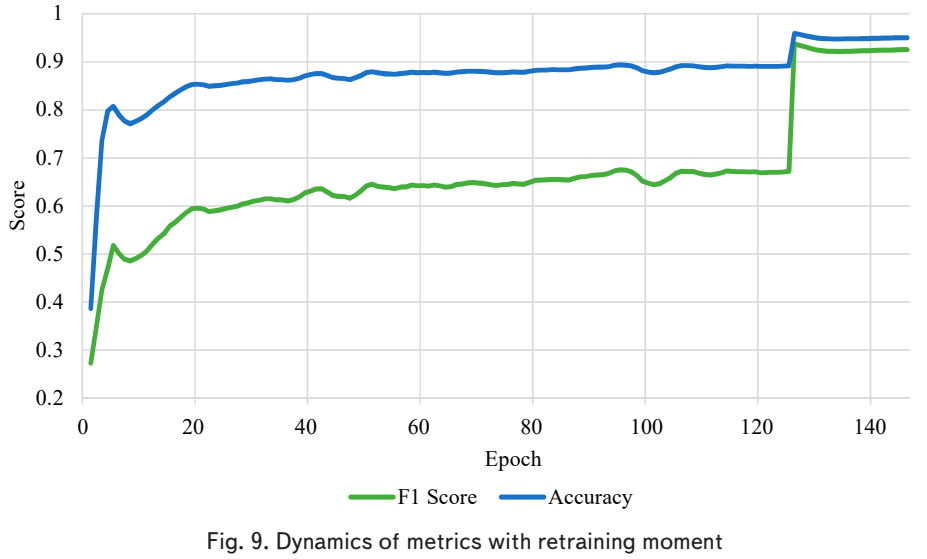

Fig. 9. Dynamics of metrics with retraining moment

The results of the study show that baseline training using GATv2, which utilizes elliptic graph data prior to Bitcoin Heist data adaptation, provides a consistent improvement in key performance metrics. The approach presented here creates a protective error profile that yields the fewest false negatives while accepting an average number of false positives [1–6, 8, 17]. The developed GATv2 architecture brings the model closer to real-world situations involving money laundering and ransomware detection in blockchain networks. The system demonstrates resistance to information noise and limited data availability, as it is designed to detect new patterns of criminal behavior emerging in dynamic environments.

6. Discussion of the results of Bitcoin transaction classification and operational stability of the model

The research aims to develop a solution that addresses the problem of blockchain analysis related to the assessment of complex financial money laundering models in network topology. The main challenges are significant class imbalance and significant structural obstacles in the data. The mathematical unification model delivers results thanks to a logarithmic compression system that maintains gradient distribution stability regardless of transaction costs. The mathematical unification model produces stable learning curves in Fig. 2 and high F1-scores in Fig. 3 thanks to a logarithmic compression system that maintains gradient distribution stability regardless of transaction costs. The results analysis process evaluates how well the dynamic attention system maintains stability when structural noise reaches its peak. The study examines how an invariant feature space contributes to successful knowledge transfer. The model is evaluated to determine its operational readiness to meet anti-money laundering (AML) requirements. The evaluation process involves analyzing existing system limitations to identify areas of development that will be most beneficial for the future development of the system.

The proposed solution functions as an adaptive nonlinear operator via GATv2Conv instead of using isotropic GCNs, which act as low-pass filters in [2, 7]. The system identifies "toxic" connections among thousands of actual hub operations using this method. Standard neighbor averaging could not achieve this (Fig. 4). Analysis of the Elliptic dataset shows that standard spectral convolutional networks (GCNs) act as low-pass filters, creating an effect of excessive smoothing in their results [16, 18]. Signal processing terms show that high-frequency components of the graph, which create sharp differences between neighbors (e.g., one "toxic" address connected to a "clean" exchange hub), are smoothed out during aggregation. Vector representations that embed malicious nodes become statistically identical to their legitimate network environment, leading to an increase in type I and type II errors. The GATv2 dynamic attention system that has been developed allows the network to perform context-based filtering operations [8, 9]. Research shows that the model learns to assign weights close to zero to connections that do not have useful semantic data, so it effectively separates suspicious nodes from the noise generated by the main hubs. This method demonstrates its advantage when working with noisy data according to Table 1, as shown in the comparison diagram in Fig. 7. GATv2 metrics remain stable, while GCN metrics deteriorate when 15% topological perturbations are introduced. The system achieves stability against unplanned changes in the network structure because Multiview attention combines the results of separate information processing systems. They operate independently of each other. The model maintains its stability because it uses anisotropic aggregation, which allows the system to ignore false confusing connections while identifying stable interaction patterns involving money transfer cycles and money transfer transactions [2, 4].

The mathematical model that combines feature spaces is an important factor in the success of the proposed approach. Traditional methods face challenges because they use fixed financial values, which makes them vulnerable to price fluctuations and changes in user economic behavior. The process of logarithmic normalization of the sum, together with relative time markers and transit coefficients, has made it possible to create a description of behavior that remains unchanged regardless of transaction values. The strategy for successful transfer of learning between the Elliptic and BitcoinHeist domains [10, 20] required this necessary prerequisite. The knowledge transfer process shown in Fig. 8, 9 demonstrates that the developed solutions address the problems of insufficient labels and ambiguous data characteristics. The model achieves this through its invariant feature space. This allows it to identify structural elements of crimes in different data domains, including BitcoinHeist, while preserving user anonymity. The results of the study confirm that financial crime networks support stable patterns that are more universal than their attribute-based characteristics. The model demonstrates the ability to generalize knowledge about families of ransomware that were not represented in the initial training dataset. The proposed method effectively addresses the problem of conceptual shift, which is a key challenge for modern cybersecurity systems [1, 3, 17].

Operational indicators and model error characteristics are the third key element to be discussed. The financial monitoring system is characterized by asymmetric costs in the event of classification errors. False negatives in the detection of terrorist financing or money laundering create critical regulatory and reputational risks. At the same time, false positives only result in additional operational costs for manual transaction verification. Analysis of the confusion matrix (Fig. 5) shows that the decision-making profile corresponds to a conservative approach, as it focuses on reducing false negatives. The classifier demonstrates its ability to distinguish between classes using the ROC curve shown in Fig. 6, which reflects an AUC value of 0.889. The system achieves an optimal balance between the ability to detect events and the ability to correctly identify non-events. This allows users to adjust the operating point according to their specific risk management needs [5, 6].

The proposed method serves as a fundamental basis for developing intelligent anti-money laundering (AML) systems for cryptocurrency networks. The integration of structural analysis with graph neural networks provides a solution to classification problems in conditions that are critical for traditional algorithms. The study shows that anisotropic aggregation successfully eliminates topological noise, allowing defense systems to adjust their defense strategies against new attack methods. These methods require further development, as they must optimize computational resources and explain their decisions to comply with financial investigations and regulatory standards.

The method requires the identification of its operational limitations, which represent its existing weaknesses. The

calculation of attention coefficients requires significant computational resources, which is the main limitation of the method. The method requires sampling of neighbors to process large-scale graphs, as it cannot process them directly. The research has a serious drawback, as the graph network faces a "cold start" problem. This prevents new addresses from being classified until they have built up a transaction history, after which their classification becomes possible through the creation of a topological profile. The GATv2 architecture provides higher accuracy than GCN, but requires more computational resources to function. The system must perform numerous matrix calculations to determine the attention coefficients for each edge, which creates a processing speed limitation during real-time operations involving graphs with millions of nodes. The neighbor sampling technique helps to mitigate this problem, but deploying the system across the entire Bitcoin network will require either significant hardware resources or the use of distilled models [2, 15]. The method achieves the best results when the training data contains complete and accurate labelled information. An attention mechanism trained on existing datasets will absorb any existing biases that exist in the historical labelling system. The model becomes less capable of detecting new types of ransomware attacks because it learns to recognize the most common ransomware patterns that appear in the training data. The method requires constant updating of datasets along with expert validation of new patterns, which is supported by an active learning system but requires significant time and effort [4, 18]. The system faces a third serious limitation, as new addresses encounter difficulties during the initial phase of operation. Graph neural networks base their work on link analysis, so they find it difficult to classify new addresses because these addresses have no previous transaction data. The function unification process allows predictions to be made based on initial transactions, but it takes time to develop a complete topological profile.

The research development process involves a transition from static research models to dynamic graphs that work with data in real time. The system allows users to store network states using static snapshots, while predicting attacks by analyzing connection development. This helps financial monitoring systems prevent attacks. At the current stage, time parameters function as graph edge attributes. The transition to temporal graph networks and dynamic continuous-time graphs will allow the evolution of the network to be modelled based on processes. This approach replaces the representation of dynamics as a sequence of individual static snapshots. Users can predict future connections, detecting preparations for large-scale attacks using this system [11, 14]. The development of explainable artificial intelligence (XAI) methods for graph-based models is a promising area of research in the field of artificial intelligence. The system must identify suspicious transactions while showing users which segment of the network caused the system to flag their transaction, as regulators have established strict rules, including GDPR and FATF directives. The process of generating explanations using GATv2 attention coefficients will create an automated system that will improve the work of financial analysts. An important area is the creation of new methods that will help AI models defend themselves against hostile attack methods. Attackers are constantly updating their obfuscation techniques, so they now use fake connections between legitimate nodes to confuse the attention mechanism. Further research focuses on developing new protection algorithms by working with reliable learning techniques and adversarial data methods [7, 13].

7. Conclusions

1. The mathematical model for unifying transaction characteristic space works through a transaction characteristic space unification system that maintains data consistency when adjusting financial amounts and time flow intensity levels. The model uses logarithmic normalization to process financial data, as this reduces the impact of extreme value distribution patterns that occur in financial data. The model operates through two key performance indicators, which include the transit coefficient and normalized operating frequency, to evaluate its effectiveness. Node behavior models were encoded in a format that deep learning algorithms can understand. The system created a framework that allows training to be transferred between different time periods and domains (elliptic and BitcoinHeist) while maintaining the accuracy of currency exchange rate predictions.

2. The graph neural network architecture uses GATv2Conv layers, which operate through a multi-channel dynamic attention system. The current method solves the problem because it selects which connections to amplify using a contextual connection rating that amplifies weak signals from suspicious neighbors and blocks structural noise from numerous legitimate transactions from high-level hubs. The dynamic attention system allows the model to create better representations while solving the problem of excessive smoothing that occurs in dense graph structures.

3. The study presents an approach to training models that face severe class imbalance using weighted binary cross-entropy and active sampling using the neighbor sampling method. The system includes a comprehensive regularization system that applies both L2 weight decay and dropout layers to the attention matrix to stop excessive smoothing while maintaining gradient stability in deep networks.

4. The proposed approach was tested using experiments with two datasets called elliptic and BitcoinHeist. The dynamic attention method yields result with an accuracy of 91.19% and an F1 score of 91.11%. A 0.6% improvement in performance is evident when comparing these results with isotropic analogues (GCN) operating in conditions of less than 15% topological noise. The model demonstrates high discriminatory power with an AUC value of 0.889. This allows it to effectively identify transactions using ransomware while minimizing the number of false positives.

Conflict of interest

The authors declare that they have no conflict of interest with respect to this study, including financial, personal, authorship, or other conflicts that could influence the study and its results presented in this article.

Funding

The study was conducted without financial support.

Data availability

The manuscript has no related data.

Use of artificial intelligence tools

The authors used exclusively the online service SciSpace (scispace.com, as a web tool for literature search; version not specified) to search for and pre-select potentially relevant scientific sources for the literature review. All suggested references were manually checked by the authors for availability, relevance to the topic, and correctness of bibliographic data, and only verified sources were included in the manuscript. All parts of the manuscript text were written and edited by the authors without the use of generative artificial intelligence tools, and the total amount of AI assistance did not exceed 25% of the research work.

Authors' contributions

Oleksandr Kushnerov: Conceptualization, Investigation, Writing – original draft; **Vladyslav Prosolov:** Software, Validation, Investigation; **Valeriy Dudykevych:** Methodology, Supervision; **Serhii Yevseiev:** Conceptualization, Supervision, Writing – review & editing; **Serhii Povaliaiev:** Formal analysis, Validation; **Yevheniia Ivanchenko:** Project administration, Resources; **Volodymyr Gorbulyk:** Data Curation, Formal analysis; **Oleksandr Chechui:** Investigation, Validation; **Dmytro Balagura:** Methodology, Resources; **Vladyslav Sukhoteplyi:** Visualization, Data Curation.

References

1. Akcora, C. G., Li, Y., Gel, Y. R., Kantarcioglu, M. (2020). BitcoinHeist: Topological Data Analysis for Ransomware Prediction on the Bitcoin Blockchain. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 4439–4445. https://doi.org/10.24963/ijcai.2020/612
2. Weber, M., Domeniconi, G., Chen, J., Weidele, D. K. I., Bellei, C., Robinson, T., Leiserson, C. E. (2019). Anti-Money Laundering in Bitcoin: Experimenting with Graph Convolutional Networks for Financial Forensics. arXiv. https://doi.org/10.48550/arXiv.1908.02591
3. Bitcoin Heist Ransomware Address [Dataset] (2020). UCI Machine Learning Repository. https://doi.org/10.24432/C5BG8V
4. Lorenz, J., Silva, M. I., Aparício, D., Ascensão, J. T., Bizarro, P. (2020). Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. Proceedings of the First ACM International Conference on AI in Finance, 1–8. https://doi.org/10.1145/3383455.3422549
5. Vassallo, D., Vella, V., Ellul, J. (2021). Application of Gradient Boosting Algorithms for Anti-money Laundering in Cryptocurrencies. SN Computer Science, 2 (3). https://doi.org/10.1007/s42979-021-00558-z
6. Alarab, I., Prakoonwit, S., Nacer, M. I. (2020). Competence of Graph Convolutional Networks for Anti-Money Laundering in Bitcoin Blockchain. Proceedings of the 2020 5th International Conference on Machine Learning Technologies, 23–27. https://doi.org/10.1145/3409073.3409080
7. Kipf, T. N., Welling, M. (2016). Semi-supervised classification with graph convolutional networks. arXiv. https://doi.org/10.48550/arXiv.1609.02907
8. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y. (2017). Graph attention networks. arXiv. https://doi.org/10.48550/arXiv.1710.10903
9. Brody, S., Alon, U., Yahav, E. (2021). How attentive are graph attention networks? arXiv. https://doi.org/10.48550/arXiv.2105.14491
10. Li, G., Tang, X. (2024). Exploring GCN, GAT, and GIN Fusion for Illicit Transaction Classification in Cryptocurrency Networks. Proceedings of the 2024 the 12th International Conference on Information Technology (ICIT), 49–53. https://doi.org/10.1145/3718391.3718399
11. Rekik, S., Mehmood, S. (2025). A Hybrid Graph Neural Network and Neural ODE Model to Intrusion Detection in Dynamic Network Topologies. IEEE Access, 13, 198201–198227. https://doi.org/10.1109/access.2025.3635385
12. Zhao, H., Liu, W., Gao, C., Shi, W., Zhang, Z., Chen, J. (2025). HGAA: A Heterogeneous Graph Adaptive Augmentation Method for Asymmetric Datasets. Symmetry, 17 (10), 1623. https://doi.org/10.3390/sym17101623
13. Luša, R., Pintar, D., Vranić, M. (2025). TE-G-SAGE: Explainable Edge-Aware Graph Neural Networks for Network Intrusion Detection. Modelling, 6 (4), 165. https://doi.org/10.3390/modelling6040165
14. H. G., M., Kumar, J., Mm, N. (2025). GrMA-CNN: Integrating Spatial-Spectral Layers with Modified Attention for Botnet Detection Using Graph Convolution for Securing Networks. International Journal of Intelligent Engineering and Systems, 18 (1), 1009. https://doi.org/10.22266/ijies2025.0229.72
15. Hamilton, W., Ying, Z., Leskovec, J. (2017). Inductive representation learning on large graphs. arXiv. https://doi.org/10.48550/arXiv.1706.02216
16. Ding, K., Li, J., Bhanushali, R., Liu, H. (2019). Deep Anomaly Detection on Attributed Networks. Proceedings of the 2019 SIAM International Conference on Data Mining, 594–602. https://doi.org/10.1137/1.9781611975673.67
17. Lo, W. W., Kulatilleke, G. K., Sarhan, M., Layeghy, S., Portmann, M. (2023). Inspection-L: self-supervised GNN node embeddings for money laundering detection in bitcoin. Applied Intelligence, 53 (16), 19406–19417. https://doi.org/10.1007/s10489-023-04504-9
18. Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z. et al. (2023). A Comprehensive Survey on Graph Anomaly Detection With Deep Learning. IEEE Transactions on Knowledge and Data Engineering, 35 (12), 12012–12038. https://doi.org/10.1109/tkde.2021.3118815
19. Elliptic Data Set. Kaggle. Available at: https://www.kaggle.com/datasets/ellipticco/elliptic-data-set
20. Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Gupta, B. B. et al. (2021). A Survey of Deep Active Learning. ACM Computing Surveys, 54 (9), 1–40. https://doi.org/10.1145/3472291