GSC Advanced Research and Reviews
eISSN: 2582-4597 CODEN (USA): GARRC2
Cross Ref DOI: 10.30574/gscarr
Journal homepage: https://gscarr.gsconlinepress.com/

(REVIEW ARTICLE)

Check for updates

ADVERSARIAL ATTACKS ON FRAUD DETECTION MODELS: A REVIEW OF EVASION, DATA POISONING, AND CAMOUFLAGED FRAUDSTERS

Saidul Islam [1,*], Md Atiqur Rahman [2], Md Jahid Hasan [1] and Farazi Ahsanul Islam [1]

[1] Master of Science in Information Technology (MSIT), Washington University of Science and Technology, Virginia, USA
[2] Doctor of Computer Science, University of the Potomac, Washington, DC, USA
* Corresponding Author: Saidul Islam, E-mail: saiislam.student@wust.edu

ORCID Details
Saidul Islam: https://orcid.org/0009-0000-3698-7140
Md Atiqur Rahman: https://orcid.org/0009-0004-4375-8994
Md Jahid Hasan: https://orcid.org/0009-0009-9723-1481
Farazi Ahsanul Islam: https://orcid.org/0009-0001-9695-4479

GSC Advanced Research and Reviews, 2026, 29(01), 380–387

Publication history: Received on 14 September 2026; revised on 22 October 2026; accepted on 24 October 2026

Article DOI: https://doi.org/10.30574/gscarr.2026.29.1.0289

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Graph Neural Networks (GNNs) have become a leading approach to fraud detection because fraud is a networked activity, but the same graph structure gives adversaries ways to mislead them. This narrative review examines adversarial attacks on fraud detection models, with a focus on GNN-based systems. It groups attacks into evasion attacks at inference time, data poisoning attacks at training time, and camouflage, in which fraudsters imitate legitimate users or attach themselves to reputable nodes. It then reviews defenses, including generative graph representations, adversarial training, robust aggregation, domain constraints, explainability, and real-time monitoring. The reviewed studies indicate that small changes to a few nodes, edges, or features can degrade graph models, that dynamic settings such as blockchain networks widen the attack surface, and that camouflage is hard to separate from unusual but legitimate behavior. The review adds summary tables of attacks and defenses, a pipeline diagram, and a toy example that shows how neighbor-smearing lowers a node's aggregated risk score. Most evidence comes from single studies of narrow settings, and comparable benchmarks for attacks and defenses are still missing. The article follows no formal search protocol. Future work should test defenses under attack and report results in a standard format.

Keywords: Adversarial Attacks; Fraud Detection; Graph Neural Networks; Data Poisoning; Evasion Attacks; Adversarial Robustness

1 INTRODUCTION

Financial technology and digital commerce have created a need for automated systems that detect and limit fraud. Graph Neural Networks (GNNs) have become a leading approach because they model relational data such as transaction networks, user-item interactions, and social graphs. As these models spread, they also become targets for capable adversaries, and the security of fraud detection systems is now an operational requirement and not only a theoretical concern. Machine learning is used widely in financial analysis (Rahman et al., 2025), and fraud detection is a setting in

which an opponent actively tries to defeat the model. This literature review covers adversarial attacks on fraud detection models, with attention to evasion, data poisoning, and the camouflage that fraudsters use to bypass algorithmic checks.

Adversarial threats fall into two timing classes: training-time and inference-time interventions. In GNN-based fraud detection they are especially potent because graph data is interconnected, so one malicious actor can influence the representations of many nodes. Zügner et al. (2018) gave an early demonstration for graph data: small perturbations to node features and graph structure sharply reduced node classification accuracy, including in the poisoning setting where the model is retrained. Adams (2022) reports that adversarial threats against GNN-based fraud systems can degrade performance considerably, especially through poisoning attacks that inject malicious nodes or edges into the training set. This changes what security requires. A model must be accurate on clean data and also resistant to deliberate perturbations meant to mislead it. Recent work, including Frankart et al. (2026) and Ojerinde and Abubakar (2026a), points to gaps in how dynamic networks and feature-space illusions can be exploited by fraudsters.

The review begins with the basics of fraud detection in the age of AI and then examines specific attack vectors. It groups attacks into evasion attacks, in which fraudsters change their behavior at test time to avoid detection, and poisoning attacks, in which the model is compromised during learning. It also covers "camouflaged fraudsters," actors who deliberately imitate legitimate patterns to blend into a network. Dou et al. (2020) studied this behavior directly and designed a GNN-based detector against it. The review draws on recent publications from 2025 and 2026 as well as foundational work on adversarial machine learning (Biggio & Roli, 2018), and it aims to outline a path toward anti-fraud systems that hold up under attack.

This article is a narrative review, not a systematic one. It covers studies of evasion, poisoning, and camouflage attacks against fraud detection models, mostly GNN-based, and the defenses proposed against them, together with foundational work on adversarial attacks on graph data (Sun et al., 2022). It draws mainly on recent peer-reviewed articles, conference papers, and preprints. Several of the recent sources are single studies of narrow settings, such as blockchain networks and e-commerce sellers, so the conclusions should be read as a map of the literature rather than as estimates of effect sizes. The review followed no formal search protocol and does not claim to cover every relevant study.

2 THEORETICAL FOUNDATIONS AND THE RISE OF GNNS IN FRAUD DETECTION

2.1 The Evolution of Fraud Detection Paradigms

The move from rule-based systems to machine learning was the first major shift in fraud detection. Financial institutions once relied on static "if-then" logic to flag suspicious transactions. Oliveira (2026) notes that real-time fraud detection requires neural networks that can handle concept drift and class imbalance. Concept drift, the change over time in the relationship between inputs and outcomes, is a long-studied problem (Gama et al., 2014). Modern environments add continuous, high-velocity data streams, and conventionally trained models often fail to adapt to shifting criminal tactics. Deep learning architectures can extract features from raw data automatically, which reduces reliance on manual feature engineering that informed adversaries could easily bypass.

Standard neural networks still struggle with the relational nature of fraud. Fraud is rarely an isolated event. It usually involves a network of collaborating actors or a series of linked transactions. Adams (2022) argues that integrating GNNs into fraudulent-seller detection is a significant advance, because these models capture structural dependencies that traditional models ignore. Representing entities as nodes and interactions as edges gives a view of the whole ecosystem. The added complexity also opens new attack surfaces. GNNs gain detection power by analyzing the "neighborhood" of a transaction, and an attacker can "pollute" a neighborhood to protect a specific fraudulent node. Fraud detection with GNNs therefore has both benefits and risks.

2.2 Graph Neural Networks: Architecture and Vulnerability

The strength of GNNs lies in message passing, in which node representations are updated repeatedly from the features of their neighbors. This allows detection by "guilt by association," and it also creates a weakness known as structural sensitivity. Wu et al. (2025) discuss how reliance on graph representations can be exploited with novel evasion attack algorithms. An attacker who understands how the message-passing function aggregates information can add or remove edges to manipulate the final embedding of a target node. The weakness comes from the way GNNs use graph structure and not only from model complexity.

Dynamic environments such as blockchain networks exacerbate the problem. Blockchain data suits graph methods, and the Elliptic dataset of more than 200,000 Bitcoin transactions (Weber et al., 2019) is a widely used benchmark for them. Ojerinde and Abubakar (2026b) point to the combined exposure of GNN-based systems to perturbation, evasion, and poisoning attacks as a serious problem in dynamic blockchain fraud detection. Because the graph keeps evolving, the

model must update its parameters often, which gives an adversary repeated chances to inject malicious data. The open design problem is an architecture that remains explainable and robust simultaneously. Ojerinde and Abubakar (2026a) also identify a synthesis gap in understanding how these dynamic threats interact with graph topology, and they call for more systematic model evaluation and defense.

3 TAXONOMY OF ADVERSARIAL ATTACKS ON FRAUD MODELS

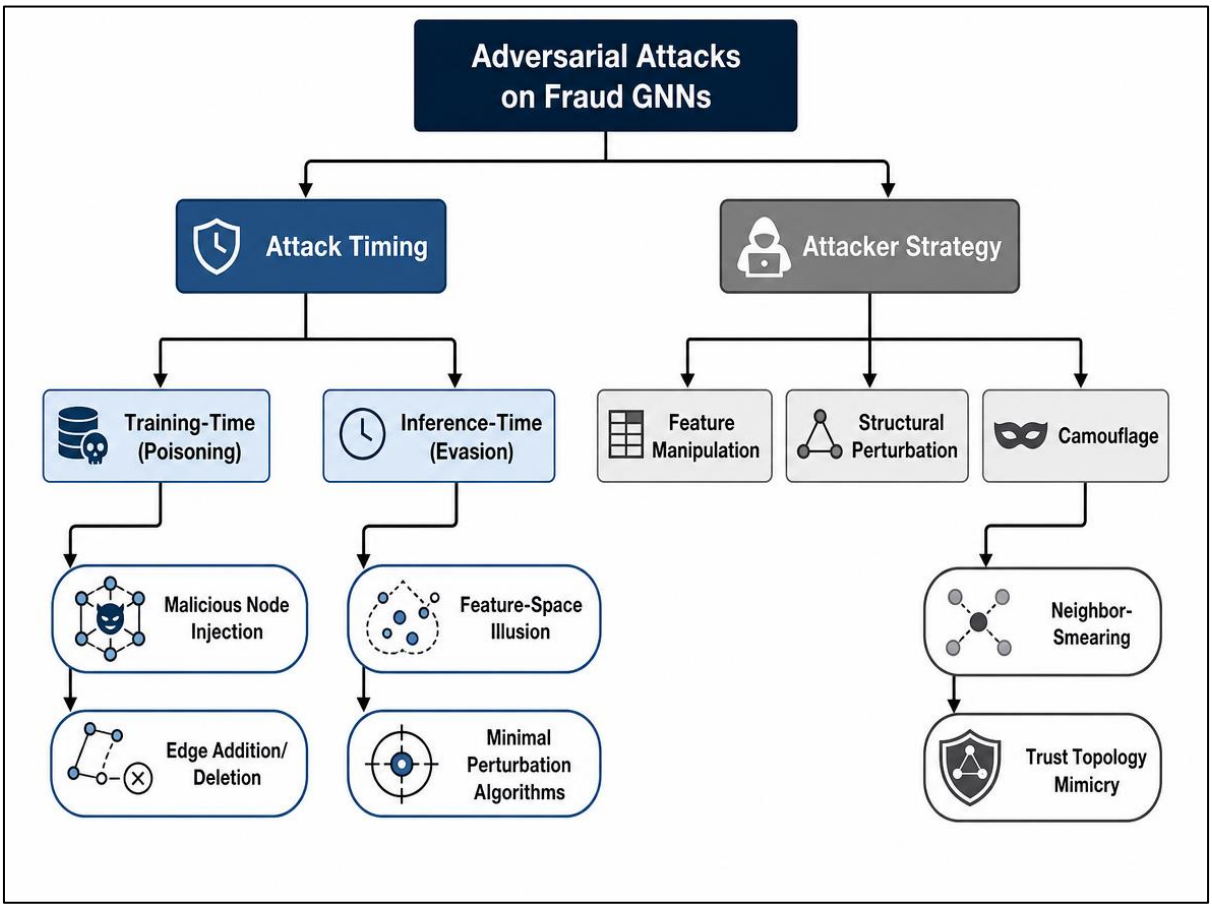

Figure 1: Taxonomy of adversarial attack vectors against Graph Neural Network (GNN) fraud detection systems, categorized by operational phase and strategic mechanism.

3.1 Evasion Attacks and the Feature-Space Illusion

Evasion attacks happen at inference time. The adversary changes the input so that a trained model classifies a fraudulent instance as legitimate. The idea comes from adversarial examples in other domains (Goodfellow et al., 2014). In blockchain GNN settings, it often means crafting transactions that look benign in feature space. Frankart et al. (2026) introduce the "feature-space illusion" and show that practical vulnerabilities in blockchain GNN fraud detection let attackers craft transactions that evade detection. This work is important because it formalizes the challenges of attacks outside the training phase. By changing transaction amounts, timestamps, or sequences of interactions, fraudsters can move a transaction's representation across the decision boundary without changing their intent. The same logic applies to graph structure: Zügner et al. (2018) show that a few added or removed edges and altered features can flip a node's predicted class.

Wu et al. (2025) design novel evasion attack algorithms aimed at anti-fraud models. Such algorithms seek the smallest perturbation that flips a model's prediction. In a financial graph, a fraudulent account might make several small, legitimate-looking transfers to high-reputation nodes to "wash" its embedding. These attacks are hard to counter because each action may violate no security rule while the combined effect on the GNN's latent representation is large. The "illusion" Frankart et al. (2026) describe also suggests that common performance metrics may overstate security, because they ignore how easily an informed attacker can move through a GNN's high-dimensional feature space.

3.2 Data Poisoning and Training-Time Integrity

Data poisoning manipulates the training set to compromise future model performance. It is a long-term strategy in which the adversary "teaches" the model to ignore certain fraud types or misclassify specific patterns. Biggio and Roli (2018) review a decade of adversarial machine learning, including both poisoning and evasion attacks, and Zügner and Günnemann (2019) showed that small structural perturbations applied before training can push GNNs below a baseline that ignores the graph altogether. Adams (2022) identifies injecting malicious nodes or edges as a main poisoning method in fraudulent-seller detection. By creating "sleeper" accounts that behave normally but link to fraudulent clusters, an attacker can bias the GNN's learning. Once the model is deployed, it has a blind spot for the structural patterns the attacker plans to use.

Poisoning is a particular risk where systems learn continuously or update online. Oliveira (2026) points out that real-time fraud detection systems are exposed to these threats because they must adapt to new data quickly. If the incoming stream is poisoned, the model's picture of "normal" behavior shifts. This resembles concept drift (Gama et al., 2014) but is driven by malice rather than by market change. Ojerinde and Abubakar (2026b) stress that poisoning attacks expose the fragility of GNN-based blockchain fraud detection. Blockchain data is immutable while the graph keeps growing, so a poisoned entry in the ledger can affect every later model iteration. Robust training techniques that find and neutralize influential but suspicious training samples before they affect the model's weights are therefore needed.

3.3 Camouflaged Fraudsters and Structural Disguise

Camouflage is adversarial behavior in which fraudsters deliberately mimic honest users to avoid detection. It goes beyond feature manipulation and involves aligning with the "topology of trust" in a network. Dou et al. (2020) identified camouflage as a problem for GNN-based fraud detectors and proposed a detector designed to counter it. Wu et al. (2025) address defense against camouflaged fraudsters by proposing generative GNN-based anti-fraud models that better separate true legitimacy from skilled mimicry. Camouflaged actors often use "neighbor-smearing," connecting themselves to reputable nodes so they inherit a "clean" representation through GNN aggregation. Figure 2 illustrates the effect with a simple mean-aggregation toy model: as a fraudulent node adds links to reputable accounts, its aggregated risk score falls below an alert threshold after only three links.

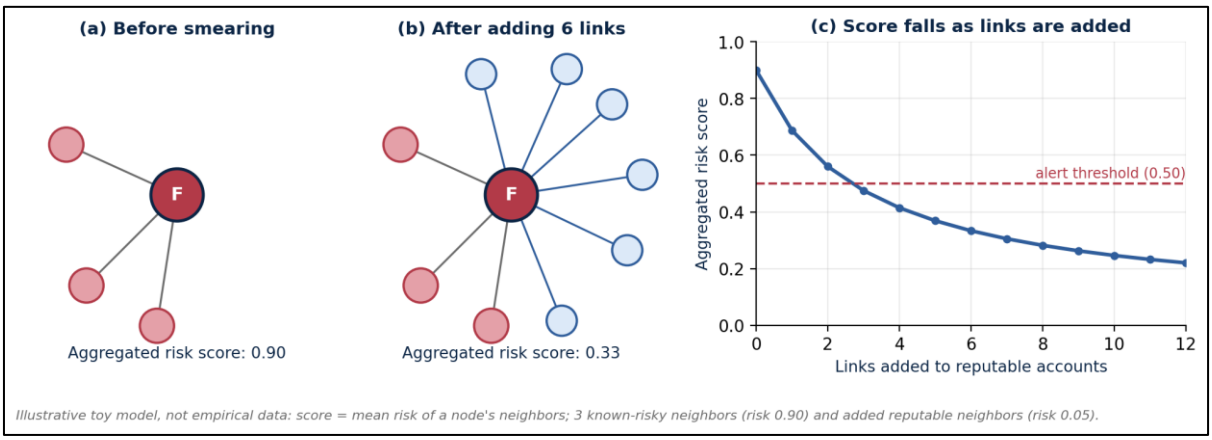

Figure 2: Illustrative toy example of neighbor-smearing: adding links to reputable accounts lowers a fraudulent node's mean-aggregated risk score (not empirical data).

Detection is hard because camouflaged behavior overlaps with that of legitimate users who have unusual but harmless patterns. Ojerinde and Abubakar (2026a) note that in dynamic blockchain networks, the line between a legitimate "whale" (a high-volume trader) and a fraudulent actor can be very thin, and the adversary exploits that ambiguity. Adams (2022) further describes fraudulent sellers on e-commerce platforms who build a reputation score by generating fake positive reviews and making low-value legitimate transactions. This structural disguise hides the fraudulent node in plain sight and defeats anomaly detection algorithms that look for outliers. Robustness here means looking beyond surface connectivity to the subtler temporal and behavioral inconsistencies that camouflage cannot reproduce perfectly. Table 1 summarizes the three attack types.

Table 1: Summary of attack types against GNN-based fraud detection, by phase, mechanism, and example.

Attack type	Phase	What the attacker changes	Example in fraud graphs	Studies
Evasion	Inference	Transaction features or local edges of the target node	Small transfers to high-reputation accounts shift the node's embedding	(Frankart et al., 2026; Wu et al., 2025; Zügner et al., 2018)
Poisoning	Training	Training nodes, edges, or labels	"Sleeper" accounts linked to fraud clusters create a blind spot	(Adams, 2022; Ojerinde et al., 2026b; Zügner & Günnemann, 2019)
Camouflage	Training and inference	Neighborhood structure and visible behavior	Neighbor-smearing and fake positive reviews build a clean profile	(Adams, 2022; Dou et al., 2020; Wu et al., 2025)

4 DEFENSES AND ROBUSTNESS IN MODERN ANTI-FRAUD SYSTEMS

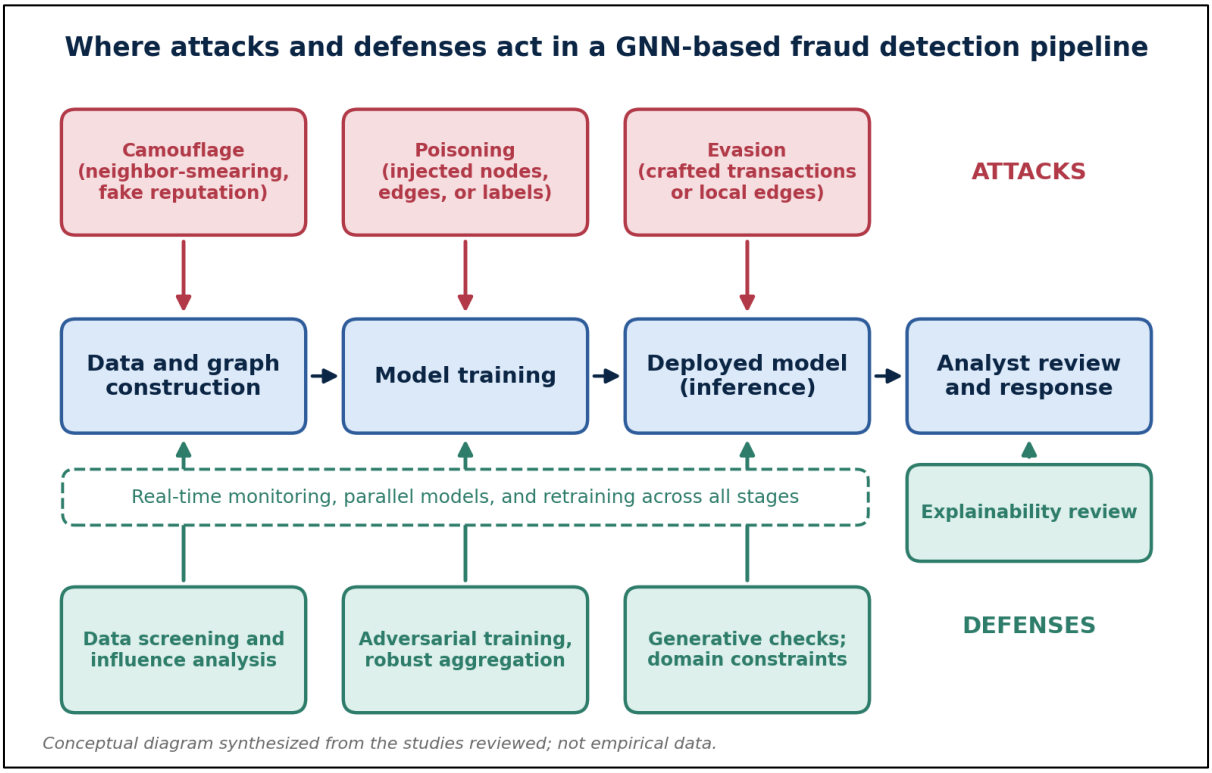

Figure 3: Where adversarial attacks and defenses act across the stages of a GNN-based fraud detection pipeline (conceptual diagram synthesized from the studies reviewed).

4.1 Generative Representations and Robust GNNs

Researchers are turning to more resilient designs against these attacks, and Figure 3 shows where each attack and defense acts in the pipeline. One promising direction uses generative models to build more robust representations. Wu et al. (2025) propose a generative GNN-based anti-fraud model meant to defend against both evasion and camouflaged fraudsters. By learning the distribution of legitimate graph structures, such models can detect when a node's neighborhood has been manipulated. Instead of only classifying a node, the generative approach asks how likely a given structural configuration is, based on what is known about honest behavior. This adds a probabilistic check that is harder for an adversary to spoof.

The literature also calls for "adversarially robust" GNNs. Ojerinde and Abubakar (2026a) give a systematic review of solutions for dynamic blockchain networks and argue that robustness must be built into the model and not added as post-processing. Techniques include adversarial training, in which the model sees perturbed examples during training (Mądry et al., 2017), and robust aggregation functions that are less sensitive to outliers or malicious edges. Frankart et al. (2026) argue that exposing the "feature-space illusion" is the first step toward such defenses. Understanding the

geometric weak points of a GNN's embedding space lets developers add constraints so that small input changes cannot cause large jumps in output, which "smooths" the decision surface against evasion. Work on AI-based algorithm design and optimization in programming (Ali et al., 2025) covers the engineering side of building and tuning such defenses. Sun et al. (2022) survey more than 100 papers on attack and defense for graph data and compare their contributions and limitations.

4.2 Explainability as a Security Feature

Explainability and robustness are converging in recent work. If a fraud model is a "black box," analysts cannot tell whether a classification rests on legitimate features or on an adversarial artifact. Ojerinde and Abubakar (2026b) argue that explainable GNNs are needed for fraud detection in dynamic networks. When a model gives a rationale for a decision, such as the transactions or neighbors that drove a high fraud score, manipulation is easier to spot. GNNExplainer (Ying et al., 2019), for example, identifies the subgraph and features most important for a prediction. If a node is flagged as fraudulent only because of one new link to a high-reputation node, an analyst may recognize a camouflage attempt. Trust in explainable AI is also a studied factor in banking fraud detection (Rimi, 2025).

Explainability can also help detect poisoning. Adams (2022) suggests that analyzing how specific training samples influence model parameters can reveal malicious data points. If a small cluster of nodes has an outsized effect on the classification of unrelated entities, that may indicate poisoning. Oliveira (2026) also notes that in real-time systems, explainability helps manage concept drift. Knowing why performance is degrading lets engineers distinguish a natural shift in user behavior from a coordinated attack. Explainability is therefore a regulatory requirement for fairness and also part of a security strategy.

4.3 Real-Time Monitoring and Adaptive Defense

The temporal nature of fraud requires defenses that move as fast as the attacks. Oliveira (2026) stresses stream processing and continuous ingestion in real-time fraud detection. A static model, however robust at deployment, will eventually fail against new adversarial strategies. Adaptive defenses monitor model performance continuously and can use "canary" nodes to detect poisoning or evasion attempts as they happen. Running parallel models trained on different data subsets or with different architectures lets a system detect inconsistencies that signal an attack. These defenses depend on reliable data pipelines, and work on workload-aware ETL architecture compares SQL Server Integration Services and Azure Data Factory for enterprise data warehousing (Islam et al., 2026).

Frankart et al. (2026) also suggest that practical vulnerabilities in blockchain GNNs can be reduced by building domain-specific constraints into the model architecture. In a financial network, for example, the flow of funds must obey conservation rules. A layer that enforces such physical or economic rules can catch an attack that creates "phantom" value or impossible transaction sequences. Ojerinde and Abubakar (2026a) conclude that anti-fraud systems should be multi-layered: the structural power of GNNs, the probabilistic checks of generative models, the transparency of explainable AI, and the speed of real-time stream processing. This combined strategy is needed to close the "evidence and synthesis gap" and defend against skilled fraudsters. Table 2 summarizes the defenses discussed.

Table 2: Defense strategies discussed in the literature, with the attack each one targets.

Defense	Main target	Mechanism	Studies
Generative graph representations	Evasion, camouflage	Models the distribution of legitimate structures and flags unlikely neighborhoods	(Wu et al., 2025)
Adversarial training	Evasion	Exposes the model to perturbed examples during training	(Mądry et al., 2017; Ojerinde et al., 2026a)
Robust aggregation and domain constraints	Evasion, poisoning	Limits the influence of outlier or malicious edges and enforces rules such as conservation of funds	(Frankart et al., 2026; Ojerinde et al., 2026a)
Training-data screening	Poisoning	Flags training samples with outsized influence on the model	(Adams, 2022)
Explainability review	All attack types	Shows which neighbors or features drove a score so analysts can detect manipulation	(Ojerinde et al., 2026b; Rimi, 2025; Ying et al., 2019)
Real-time monitoring and parallel models	Drift, ongoing attacks	Tracks performance continuously and compares models trained on different data	(Oliveira, 2026)

5 CONCLUSION AND FUTURE DIRECTIONS

Adversarial attacks on fraud detection models form a complex and fast-moving contest. As this review shows, the move from traditional machine learning to graph neural networks gave institutions powerful tools against fraud and also opened new weaknesses that skilled actors want to exploit. The core problem is the relational nature of the data: because GNNs depend on how entities are connected, an adversary can alter the system's representation by changing a relatively small number of nodes or edges (Zügner et al., 2018). The attack types in the recent literature range from the "feature-space illusions" of evasion (Frankart et al., 2026) to the long-term corruption of data poisoning (Adams, 2022).

The camouflaged fraudster stands out as a primary threat. These actors not only hide but also blend in, using GNN mechanisms such as neighborhood aggregation to their advantage (Dou et al., 2020; Wu et al., 2025). That calls for a move from simple anomaly detection toward generative and explainable models. Ojerinde and Abubakar (2026a; 2026b) show that in dynamic settings such as blockchain, the fast evolution of data and the need for real-time response add to the difficulty. Their 2026 work suggests that progress in fraud detection will depend on adversarial robustness, the ability of a model to keep its integrity under targeted pressure, and not on accuracy alone.

Building explainability into the security pipeline also changes how defense is designed. A more transparent GNN allows organizations to see when adversarial artifacts, and not legitimate patterns, are misleading it (Adams, 2022; Ojerinde et al., 2026b). Together with the real-time processing Oliveira describes (2026), this supports a proactive defense. The generative approach of Wu et al. (2025) is a promising path, because it allows models to learn what legitimate behavior looks like in a way that is hard for fraudsters to imitate perfectly.

The literature from 2024 to 2026 describes the fight against fraud as an algorithmic arms race. GNNs give a strong framework for analyzing the networks where fraud occurs, but their sensitivities call for robust, explainable, and adaptive architectures. Most of the evidence comes from individual studies of specific settings, and standard ways of comparing attacks and defenses across studies are still missing. Future research should link theoretical vulnerabilities to practical defenses, use structured evaluation frameworks such as those proposed for software engineering technologies (Hridoy et al., 2023), and test defenses both under attack and on clean data. Combining generative representations, structural constraints, and real-time explainability is a likely basis for future anti-fraud infrastructure.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

The authors declare no conflicts of interest regarding this manuscript.

Use of generative AI tools

Generative AI tools were used to assist with language editing and with drafting parts of the text. The authors reviewed and edited all content and take full responsibility for the manuscript.

REFERENCES

[1] Adams, V. R. (2022). Graph Neural Networks, Analytics, and AI for Fraudulent-Seller Detection: Methods, Evaluation, and Impact on Future Trends. 1–37. https://doi.org/10.64206/vvc03q65

[2] Ali, K. M. Y., Akter, S., Islam, S., & Mridha, M. F. (2025). Advancing Computational Intelligence: AI-Based Algorithm Design and Optimization in Programming. Journal of Computer Science and Technology Studies, 7(1), 122–138. https://doi.org/10.32996/jcsts.2025.7.1.10

[3] Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023

[4] Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., & Yu, P. S. (2020). Enhancing Graph Neural Network-based Fraud Detectors against Camouflaged Fraudsters. Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 315–324. https://doi.org/10.1145/3340531.3411903

[5] Frankart, F., Simonetto, T., Cordy, M., Papageorgiou, O., Pocher, N., & Fridgen, G. (2026). The Feature-Space Illusion: Exposing Practical Vulnerabilities in Blockchain GNN Fraud Detection. IEEE Conference on Secure and Trustworthy Machine Learning (SaTML), 268–283. https://doi.org/10.1109/satml68715.2026.00023

[6] Gama, J. M. P. da, Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 1–37. https://doi.org/10.1145/2523813

[7] Goodfellow, I., Shlens, J., & Szegedy, C. (2014). Explaining and Harnessing Adversarial Examples. In International Conference on Learning Representations. https://doi.org/10.48550/arxiv.1412.6572

[8] Hridoy, S. M. A. S., Rahman, S. M., Rishad, S. M. S. I., Bhuiyan, M. S., Islam, S., & Raihan, M. J. (2023). A Comprehensive Framework for Evaluating Software Engineering Technologies. In SSRN Electronic Journal. RELX Group (Netherlands). https://doi.org/10.2139/ssrn.4650826

[9] Islam, S., Rahman, M. A., & Hasan, M. J. (2026). Workload-Aware ETL Pipeline Architecture for Enterprise Data Warehousing: A Mixed-Method Study of SQL Server Integration Services and Azure Data Factory. Figshare. https://doi.org/10.6084/m9.figshare.34625415

[10] Mądry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2017). Towards Deep Learning Models Resistant to Adversarial Attacks. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.1706.06083

[11] Ojerinde, O. A., Abubakar, R., Alenoghena, B. I., & Lawal, K. H. (2026a). Adversarially Robust and Explainable Graph Neural Networks for Fraud Detection in Dynamic Blockchain Networks: A Systematic Solution-Based Literature Review. Journal of Future Artificial Intelligence and Technologies, 3(2), 387–418. https://doi.org/10.62411/faith.3048-3719-428

[12] Ojerinde, O. A., Abubakar, R., Susan, A. A., & Aminu, E. F. (2026b). Adversarially Robust and Explainable Graph Neural Networks for Fraud Detection in Dynamic Blockchain Networks: A Problem-Based Systematic Literature Review. International Journal of Data Informatics and Intelligent Computing, 5(3), 78–113. https://doi.org/10.59461/ijdiic.v5i3.306

[13] Oliveira, D. N. O. (2026). Neural Networks for Real-Time Financial Fraud Detection. Journal International Review of Research Studies, 1(3), 1–49. https://doi.org/10.66104/xkad2306

[14] Rahman, M. A., Khan, M. H., Islam, S., Mobin, K. S. A., & Kabir, M. F. (2025). Artificial Intelligence-Driven Business Intelligence: Machine Learning Techniques for Financial Market Analysis. International Journal on Science and Technology, 16(1). https://doi.org/10.71097/ijsat.v16.i1.2420

[15] Rimi, N. N. (2025). AI-Driven Fraud Detection and Cybersecurity Resilience in the U.S. Banking Sector: The Mediating Roles of Explainable AI Trust and Real-Time Fraud Response Capability. Communications on Applied Nonlinear Analysis. https://doi.org/10.52710/cana.jrnl.v32.7478

[16] Sun, L., Dou, Y., Yang, C., Zhang, K., Wang, J., Yu, P. S., He, L., & Li, B. (2022). Adversarial Attack and Defense on Graph Data: A Survey. IEEE Transactions on Knowledge and Data Engineering, 1–20. https://doi.org/10.1109/tkde.2022.3201243

[17] 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. KDD 2019 Workshop on Anomaly Detection in Finance. https://doi.org/10.48550/arxiv.1908.02591

[18] Wu, J., Wang, X., Yang, J., Cheng, D., Yang, G., & Wang, B. (2025). Defending Attacks on Anti-Fraud Model With Generative Graph Representations. IEEE Transactions on Knowledge and Data Engineering, 38(2), 969–982. https://doi.org/10.1109/tkde.2025.3640634

[19] Ying, R., Bourgeois, D., You, J., Žitnik, M., & Leskovec, J. (2019). GNNExplainer: Generating Explanations for Graph Neural Networks. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.1903.03894

[20] Zügner, D., Akbarnejad, A., & Günnemann, S. (2018). Adversarial Attacks on Neural Networks for Graph Data. ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD), 2847–2856. https://doi.org/10.1145/3219819.3220078

[21] Zügner, D., & Günnemann, S. (2019). Adversarial Attacks on Graph Neural Networks via Meta Learning. In arXiv (Cornell University). Cornell University. https://doi.org/10.48550/arxiv.1902.08412