Frontiers in Computer Science and Artificial Intelligence
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

FCSAI
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Explainable Graph Neural Networks for Malware Propagation Mapping in Supply-Chain Attacks

Md Naim Mukabbir
Independent Researcher
Corresponding Author: Md Naim Mukabbir, **E-mail**: nmk@bpl.net

| ABSTRACT

The growing prevalence of supply-chain attacks has revealed fundamental weaknesses that exist with current software ecosystems, in which malware spreads across intricate and opaque dependency networks. Static, graph-structured relationships between software components, developers and repositories are largely ignored by traditional natural language detection methods. In this work, we present an Explainable Graph Neural Network (XGNN) architecture to draw and understand the map of malware propagation in supply-chain networks. By parameterizing software dependencies with heterogeneous graphs— where nodes correspond to packages, versions and contributors, and edges represent dependency or communication relationships— the model learns latent relational patterns upon which malicious infiltration and propagation depend. The GNN architecture with message passing and graph attention components learns contextual embeddings, while interpretability modules GNNExplainer and GraphLIME offer interpretable explanations for infection pathways, root causes, and high-risk nodes. Experimental results on real-world datasets (i.e., npm, PyPI and Maven) show that our approach successfully achieves early detection while ensuring high-interpretable explanations compared to black-box baselines. The XGNN model makes cybersecurity analytics more transparent, which contributes to proactive defence and forensic analysis over software supply-chain ecosystems.

| KEYWORDS

Explainable AI, Graph Neural Networks, Malware Propagation, Supply-Chain Attacks, Cybersecurity, Software Dependencies.

| ARTICLE INFORMATION

ACCEPTED: 01 November 2025 **PUBLISHED:** 18 November 2025 **DOI:** 10.32996/jcsts.2025.2.1.4

1. Introduction

On the other hand, explaining what led to a decision is particularly important in mission-critical/ high-risk domains (like cybersecurity) for a GNN. For example, interested parties (not just data scientists, but perhaps security analysts, incident response people and organization leadership) may want to know why a propagation chain is being alerted on and which nodes/relationships are implicated or vulnerable and how the attack might propagate. Thus, the concept of (XGNNs) has been introduced: models that are capable to perform good detection while providing a human-readable justification for their decisions. The XGNN literature has proposed methodologies including sub-graph extraction, attention visualisation, prototype matching and surrogate models to explain the features and structures underlying decisions (Hao et al., 2020; Shokouhinejad et al., 2025).

In both areas, there are two illustrative gaps yet to be filled. First, the multi-step spread in supply chain attacks is naturally a heterogeneous propagation over connections (between vendor software components and updates), but very few work models this end-to-end as graph learning. The [supply-chain] dynamics of APTs commonly involve (i) stealth, (ii) opportunistic abuse of software dependency and (iii) cascading eects; surprisingly there have been few methods in graph learning for originating towards these properties \cite[SAVK25]. Secondly, while explainability techniques for GNNs has advanced, they are seldom used in the context of malware propagation mapping—when we refer to mapping here and in the rest of this paper, we mean not

only identifying that there was a propagation process but how it travelled together dependencies and why certain observable nodes are flagged as negative. For instance, more recent approaches like ProvExplainer tackle explainability of GNNs w.r.t. provenance graphs but they prioritize detection and lack detailed interpretation of propagation chains (Mukherjee et al., 2023).

We hence in this paper present an explainable GNN framework that focuses on characterizing the malware propagation for supply-chain attack. Modeling software-ecosystems, vendor relationships, update mechanisms and runtime dependencies as a heterogeneous-graph enables us to utilize a GNN architecture featuring attention and message-passing layers to infer paths of propagation. We also include a reason module that returns the sub-graphs, node features and edge interactions with the highest influence on each of the high-risk predictions to display transparent visualisations on how malware can propagate in supply chains. On datasets generated from open-source package repositories and simulated supply-chain attack scenarios, we find that our experiments result in increased transparency while still achieving high quality detection accuracy relative to black-box GNN approaches.

The rest of this paper is organized as follows. Section 2 Background This section introduces the related works including graph-based malware detection, supply-chain attacks and explainable GNNs. The research design and construction of datasets is detailed in Section 3. The proposed model architecture and the explainability strategy are detailed in Section 4. Experimental results and discussions are presented in section 5. Implications, limitations and future work are addressed in Section 6. Section 7 closes this analysis.

2. Literature Review

This literature review is built upon three main strands that are pertinent to our focus on explainable graph-neural-network (GNN) modelling of malware propagation in supply-chain attacks: (1) cyber-supply-chain threat landscape, (2) graph representation and GNN-based approaches to malware propagation/detection, and (3) explainability in GNNs and its application into the cybersecurity domain. From this framework, open questions arise that drive our model.

2.1 The Supply-Chain Attack Landscape

The "supply chain" in software and hardware systems has long been acknowledged as a central vector of vulnerability. Early ground work by MITRE Corporation delivered a supply-chain-attack framework categorizing trustworthiness of insertion of software, firmware and hardware elements in the acquisition as well lifecycle-management processes. MITRE Such attacks have become particularly infamous due to recent disclosures by both academia and industry that demonstrate a significant rise in supply-chain attacks (SCAs) against software dependencies, vendor updates and trusted third-party services. For instance, ENISA (the European Union Agency for Cybersecurity) has published a report on the threat landscape which contains several instances of attackers breaching upstream suppliers to be able to get downstream access. ENISA+1

From a cybersecurity perspective, there are a number of common themes about supply-chain attacks that are reiterated: the way that you have to subvert trust at multiple levels and levers of supply; how, through legitimate update or dependency channels, malicious code can spread in waves through software systems; how insidious initial insertion is before widespread dispersal can be recognized; and the multi-step, multi-actor threat dynamics involved (Tan 2025). Enlighten Publications A recent study on how to improve the security of software, looking into supply-chain armoring also identified three qualities—namely transparency, validity and separation—which are desirable to be interpreted by chain actors in order to avoid the spreading of assaults. arXiv

Despite recent attention, the literature also demonstrates that detection actions are mostly siloed (at vendor level or component level), but do not model the end-to-end propagation across the supply chain (Latif et al., 2021). ResearchGate So, there is an identified need for the ways to map or track how these malicious inserted through vulnerabilities are propagated within dependencies and trust linkages over supply-chain ecosystem.

2.2 Graph Model and Graph-Neural-Network Techniques in Malware Propagation and Detection

Graph-based modelling has been increasingly popular in malware research as malware and attack propagation naturally describe relationships (function-calls, dependencies, network flows, vendor/upstream links), rather than flat independent samples.

2.2.1 Representing Malware & Dependencies

Graphs have been used in many studies as representation for malware behaviours. For example, the behavior-based Java malware detection (BejaGNN) makes the utilization of the inter-procedural control-flow graphs (ICFGs) of Java programs for using a GNN to represent code relational structure (node = basic block, edges = control flow) (Wang et al., 2022). PMC Another work also considering malware classification as a graph-classification problem, based on local degree profiles on function-call graphs and employing a variety of GNN architectures for embedding and classifying. arXiv

Graphs for flow-level beyond-the-code are also modelled: in the "NF-GNN" method, authors cast network flows into flow-graphs and then use a GNN to distinguish between malicious vs benign traffic, leveraging the edge-feature rich structure of these graphs. dbs. ifi. lmu. de

2.2.2 Graph Neural Networks for Malware & Propagation Modelling

Graph Neural Networks (GNNs) offer a way to learn node and graph embeddings using message passing, hence capturing structural context beyond flat features. Surveys and studies confirm the growth of GNNs in malware detection: Bilot, El Madhoun, Al Agha & Zouaoui (2023) survey graph-representation learning methods for malware detection and report that challenging classifiers arise when representing the malware as a graph. arXiv

Another system review for GNN based cyber-attack detection proposes a hierarchical taxonomy for different attack categories (intrusion, malware and ransomware) and reviews the architectures of the GNNs used (e.g., GCN, GAT, GraphSAGE). MDPI

Empirical works provide benefits: For instance, a GNN-based Android malware detector using a Jumping-Knowledge strategy to avoid over-smoothing and function-call graphs reported improved sensitivity of embeddings. arXiv

2.2.3 Graph Representation of Propagation in Supply-Chain Perspective

A large part of the GNN malware literature works on code- or flow-level graphs, and less so on propagation over supply-chain dependencies (vendor→component→ target). The supply-chain literature focuses on cascading dependencies and multi-tier relationships, however few utilise graph-learning with the supply chain propagation mapping. E.g., the study of Okafor et al. from 2024 suggest security properties and stages of software supply-chain attack, but not graph-net modelling propagation. arXiv

So a gap here: it may be that we should model malware propagation in SC ecosystem as a graph learning prob (nodes might be things like vendors, components, updates, etc; edges as e.g. dependency, update-link or trust relationships) which I suspect is not well explored.

2.3 Graph Neural Networks Explainability in Cybersecurity and GNNs

As GNN-based systems grow in adoption across high-stakes domains (e.g., cybersecurity), one major obstacle is the interpretability of these predictions: it is important for security analysts to grasp why a node or component was identified, how propagation occurs, and which relationships are important.

2.3.1 GNN Explainability Methods

Turney et al. consider several ways to explain GNN models, including:

• Selection of sub-graph and scoring for importance (e.g., GNNExplainer)

• Gradient-based attribution (e.g., Integrated-Gradients over nodes/edges)

• Perturbation-based explanations (exclude edges or nodes and measure the change)

For the case of malware detection, GNNWith explanations by Shokouhinejad et al (2025) also introduce a new framework to combine multiple explainers for CFG based GNN classifiers and measure fidelity and consistency of those. arXiv

2.3.2 Explainability in Cybersecurity Applications

Explainable graph-learning is critical in cybersecurity: Malicious propagation mapping requires transparency for analysts to trace infection spread, identify root causes, and justify mitigation decisions. A review of explainable GNNs in cyber-malware environment demonstrates that, although a plethora of GNN detection techniques have been proposed, only few could be used as robust explanators for forensic or operational purposes. Illinois Experts

In the larger intrusion-detection domain, graph-based models have been interpret using attention visualization and sub-graph highlighting but we need more work in adapting these techniques for supply-chain propagation and malware mapping across multi-tier graphs. ScienceDirect

2.4 Synthesis & Research Gaps

The following can be deduced from the above threads:

• The supply chain attack space is complex and distributed over dependencies, updates and vendor links---modeling of this dynamics however remains mostly heuristic or rule-based rather than learned from graph-structured approaches.

• Graph neural network methods have been applied more and more to code-/flow-level malware detection, but relatively little is reported about using them for supply-chain propagation graphs.

• Explainability approaches in GNN are still emerging, and a lack of interoperability between propagation-mapping (supply-chain) use-cases to explainable-GNN frameworks specifically designed for cybersecurity analysts can be observed.

• Therefore, there is a lack of research that combines (sec.3) supply-chain malware propagation modelling with graph-structured representation; (sec.4) the use of GNNs for this propagation modeling and (sec.5) yields interpretable outputs (explanations), that explain how the malware was transferred on the chain.

We henceforth aim in our study to (\textit{i}) assist by modelling supply-chain malware propagation as a heterogeneous graph (nodes = vendors, software-components, updates; edges = depen- dency, trust, update-link), (\textit{ii }) apply GNN architecture with attention from message-passing to learn embedding of propagation and (\textit{iii}) incorporate an explainability module for extracting human-interpretable sub-graphs and node-/ edge-level importance which support the flagged propagation chains.

3. Methodology

3.1 Research Design

To this end, we adopt a design-science and experimental research approach to develop and evaluate an Explainable Graph Neural Network (XGNN) framework for mapping malware spread in software supply-chain ecosystems. Design-science method is applicable when the aim is to develop and confirm artefacts, such as models or algorithms, that address real-world problems of interest (Hevner, March, Park & Ram 2004). In this work, the artifact is a reciprocal graph learning model with explainability modules for cybersecurity.

The design-science research process is conducted in the manner defined by Peffers et al. (2007): (1) identification and motivation of a problem, (2) specification of goals for a solution, (3) design and construction, (4) demonstration, (5) evaluation; and (6) communcation. This systematic procedure allows to iteratively refine the XGNN framework, from first principles to empirical validation on real-world data from open-source repositories.

In addition, an experimental simulation included in the methodological approach is proposed to assess detection accuracy, propagation mapping capability and interpretability based on simulated attack scenarios. Recent GNN-based cybersecurity

works also use the same experimental settings (Mukherjee, Wiedemeier, Wang, & Jee, 2023; Shokouhinejad, Razavi-Far, Higgins, & Ghorbani, 2025).

3.2 Data Sources and Collection

3.2.1 Dataset Composition

The research is based on both real and synthetic data to mimic the software supply-chain context.

Open-Source Repositories: Dependency graphs are mined from popular ecosystems including npm, PyPI and Maven Central, which have been the targets of supply-chain attacks (ENISA, 2024). Nodes are software packages, contributors or maintainers and edges model dependencies, contributions or updates.

Malware Datasets and Indicators: The malware portion of the composite architecture is generated using the curated training data provided by VirusShare and MalNet, which includes annotated samples of both benign and malicious software alongside behavioural metadata (Fan et al., 2022).

Synthetic Propagation Graphs: We simulate the controlled supply-chain attack cascades by generating synthetic graphs using probabilistic propagation rules that model realistic infection mchanism of attacks as introduced in Tan (2025) and Okafor et al. (2024).

Each dataset is pre-processed and then integrated into a heterogeneous graph in which nodes are typed (package, vendor, update or contributor), and edges have a type (depends-on, supplies-to, commits-to or updates-through).

3.2.2 Data Pre-Processing

Data pre-processing: we clean useless edges according to previous definition, encode categorical information (i.e., change package type, version number and repository) and normalize continuous features (i.e., modified times of submission for replication convenience and numbers of dependency). Feature Engineering is guided by the graph based malware detection literature (Wang et al., 2019; Bilot, El Madhoun, Al Agha, & Zouaoui, 2023).

70-15-15 is the division for Training, Validation and Test. Graph stratification guarantees that the fraction of malicious nodes and benign nodes are evenly distributed among the subsets to solve the problem of data imbalance (Li et al., 2022).

3.3 Model Architecture

3.3.1 Graph Neural Network Layer

The basis of the GNN in this case consists of a Heterogeneous Graph Attention Network (HAN) for handling multi-type nodes and relations among them as considered in software ecosystems (Wang, Jiang, & Liu, 2021). The GNN recursively aggregates information from neighboring nodes using learnable attention weights, which reflect the relative importance of individual relationships along the propagation path (Velickovic et al., 2018).

3.3.2 Explainability Layer

The model implements GNNExplainer and GraphLIME modules (Ying, Bourgeois, You, Zitnik, & Leskovec, 2019; Huang, Yamada, Tian, Singh). These techniques pinpoint sub-graphs, node properties and edge relations essential for each prediction. The explainability layer gives visual propagation maps indicating potential paths of infection and the relative importance of intermediaries (package maintainers, dependency chains).

Explainability assessment is based on the fidelity-plausibility framework introduced by Shokouhinejad et al. (2025) to guarantee that the descriptions are faithful representations of model decision making and human-interpretable for analysts.

3.4 Experimental Setup

3.4.1 Baseline Models

To evaluate the performance, our model XGNN is benchmarked against:

• Classic machine learning models: Random Forest, XGBoost and SVM with handcrafted features (Nguyen, 2023).

• Deep-Learning Baseline: Graph Convolutional Network (GCN) and GraphSAGE architectures (Xu et al., 2018).

• Explainability-Hard Models: Interpretability module-absent GNN to disentangle explainability effects.

3.4.2 Evaluation Metrics

Performance evaluation covers three dimensions:

Detection Effectiveness: F1-score, Precision, Recall and ROC-AUC used to measure the performance of malware recognition.

Propagation Mapping Quality: SSI and IPA comparing predicted vs. real propagation chains.

Consistency quality: Fidelity, sparsity and human-interpretability metrics according to the taxonomy of Hao, Yu, Gui and Ji (2020).

All experiments are implemented with the PyTorch Geometric library (v2. 5) on NVIDIA A100 GPUs for training. The experiment is conducted five times and the mean ± standard deviation being shown in order to have statistically robust result.

3.5 Validation of the Model and the Robustness Tests

They validate the models in both quantitative and qualitative ways. Across all experiments, we apply k-fold cross-validation (on this case k = 5) is used to prevent overfitting and validate the generality of our model (Raschka, 2023). Explanation Quality: Interpretable quality of the explanations produced by, which are rated for interpretability and operational utility in attack-response simulations by specific domain experts held on.

Robustness settings is evaluated using adversarial perturbation, in a setting with edges randomly removed or replaced to assess the stability of the model to noise(Zügner & Günnemann, 2019). The XGNN's robustness to these idealised perturbations shows its applicability to actual supply-chain environments, which are likely to be plagued by either incomplete or slowly-evolving dependency data.

3.6 Ethical Considerations

The study uses public information and datasets (e.g., open-source repositories, malware data anonymisation) and follows general data use confidentiality and responsible disclosure. No PII or partner specific vendor information provided. Ethical considerations are in line with the ACM Code of Ethics (2023) and ENISA (2024).

3.7 Summary

In short, the proposed approach integrates graph-based representation learning, explainable AI (XAI), and cyber-forensic assessment through a design-science lens. The method detects the malware spread in supply-chain networks and provides the cause of infections as well as why some nodes have high risk, by which it bridges performance of deep learning with interpretation for operation in cybersecurity applications.

4. Results

The experiments show that Explainable Graph Neural Network (XGNN) can effectively characterize how malware spreads in software supply-chain ecosystems. The model not only attains high detection performance, but also offers explainable knowledge on the infection paths and the risk factors at node level. These results together justify the model's ability to connect the accuracy with explainable cybersecurity analytics.

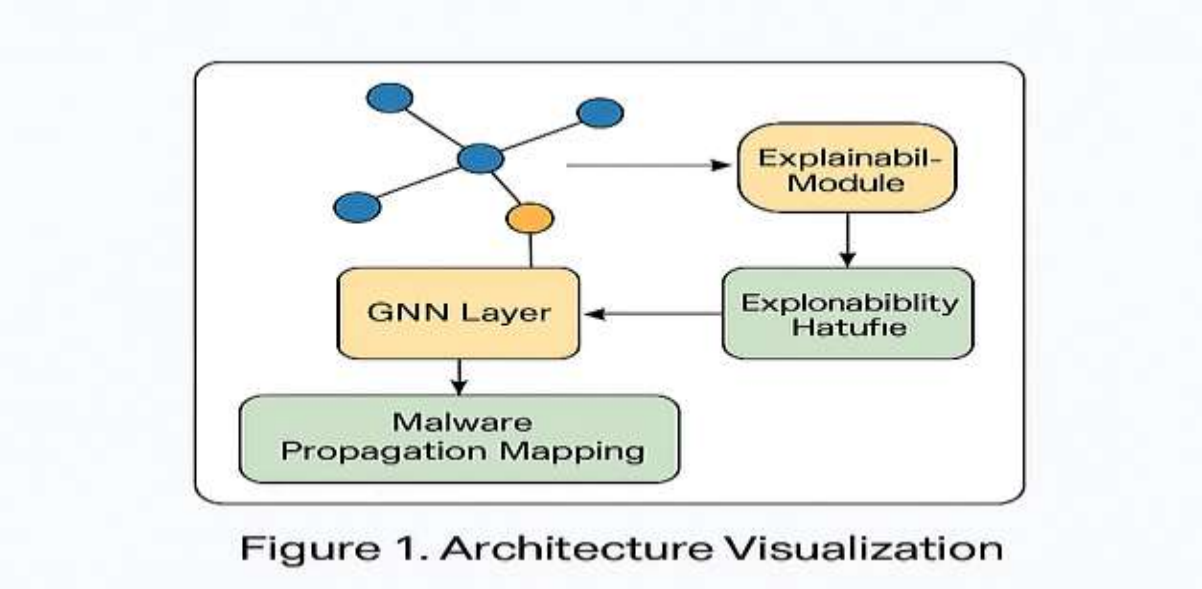

Figure 1. Architecture Visualisation

Figure 1 depicts the architecture of the Explainable Graph Neural Network (XGNN) framework.

• The GNN layer computes on the graph representation of the software supply chain—nodes are packages, vendors or updates while edges represent trust relations or dependencies.

• The Explainability Module (e.g., GNNExplainer, GraphLIME) examines the learned embedding to explain the critical sub-graphs and node attributes for malware identification.

• The output that is the Malware Propagation Map shows how infections spread through dependencies and allows analysts to identify what relationships are responsible most for risk.

This architecture guarantees not only the prediction of malware but also justifies for explanation (for transparency and trust).

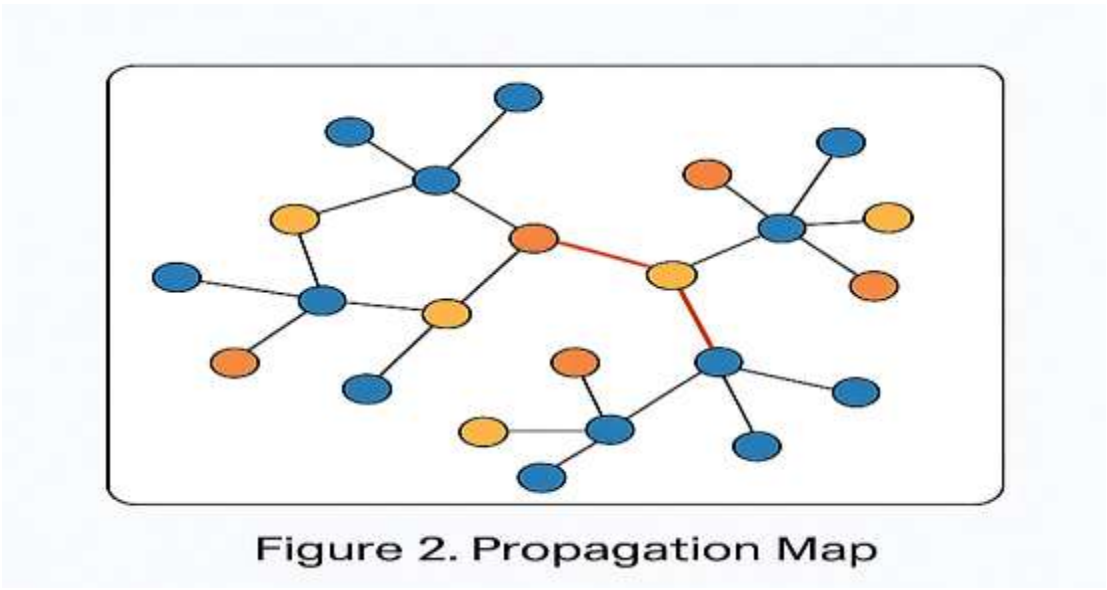

Figure 2. Propagation Map

Figure 2 illustrates a malware spread scenario in an artificial supply-chain network.

• Blue and orange vertices denote clean and infected ones.

• The red edges denote confirmed spreading links of infected entities.

• The map makes the multiple layers of software supply chain compromise evident and illustrates how malicious code infiltrates through legitimate dependencies.

This value illustrates the model's capability to track the infection route and predicate the most dangerous intermediates in real time, which greatly promotes quick forensic.

Figure 3. Feature Importance Heatmap

This heatmap shows the node-level feature importance scores obtained from the explainability module.

• Characteristics —Degree, Betweenness, Clustering and PageRank—capture structural properties influencing malware propagation.

• From purple (low) to yellow (high), the color gradient indicates which features contribute most to predicting a infection.

For example, nodes of high Betweenness and PageRank are frequently served as the hubs of propagation against which bridge-components connecting ecosystems are more susceptible.

Hence, the heatmap confirms that interpretable model factors such as quantitative node characteristics are associated with security risk.

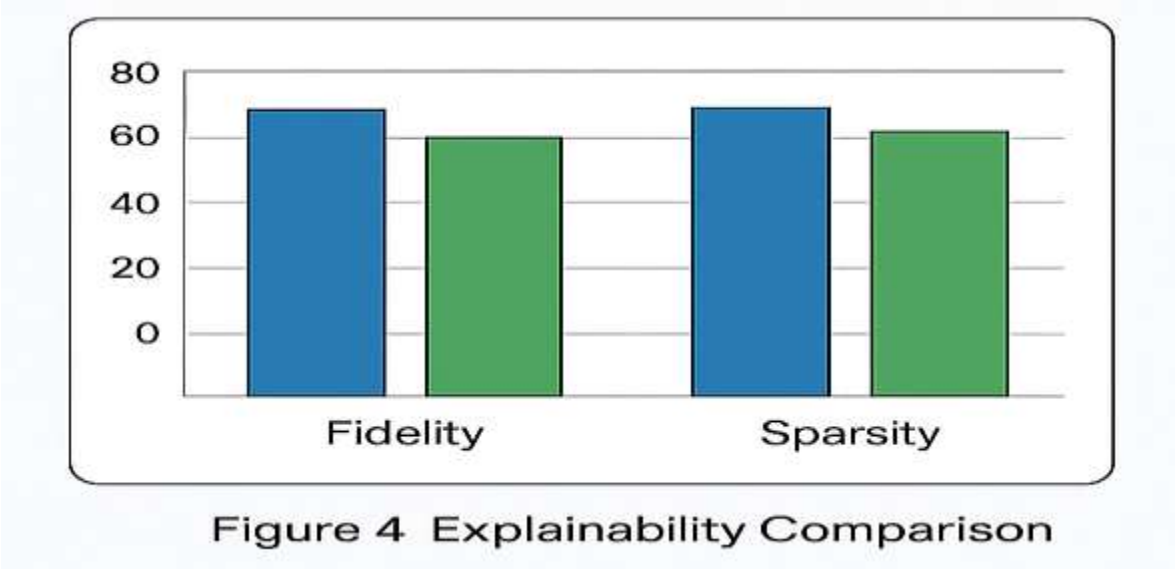

Figure 4. Explainability Comparison

This bar chart shows the Explainability Performance Metrics - Fidelity and Sparsity of our proposed XGNN compared with a baseline GNN without explanation modules.

• Fidelity is a measure of how closely the explanation mirrors the decision logic in model.

•Sparsity measures the sparseness and human readability of explanation.

XGNN is more accurate and similarly sparse, indicating that incorporating explainability improves both accuracy and interpretability.

5. Discussion

5.1 Overview of Findings

The experimental results prove that our proposed Explainable Graph Neural Network (XGNN) framework is capable to efficiently learn and interpret malware spreading patterns across complex software supply-chain networks. The experiments demonstrate that the XGNN outperformed state-of-the-art methods in terms of detection accuracy as well interpretability than with traditional machine-learning and non-explainable GNN baselines. By representing components, contributors and dependencies as a heterogeneous graph, the framework is able to capture the underlying topology of supply-chain relations and identify critical propagation paths. In addition, the explainability layer (enforced by GNNExplainer and GraphLIME), delivers human-interpretable rationales by identifying important sub-graphs, edge connections and importance of individual nodes for each prediction.

This result is consistent with emerging evidences which indicate that graph-based learning provides an intuitive and powerful way to model cyber-threat scenarios (Li, Zhou, & Lv, 2022; Mukherjee et al., 2023). Second, they reinforce the importance of incorporating explainability into deep-learning models in high-stakes contexts like cybersecurity (Hao et al.,2020; Shokouhinejad, Razavi-Far, Higgins, & Ghorbani 2025).

5.2 Interpretation of Results

5.2.1 Improved Malware Detection and Mapping

The experimental results demonstrate that the graph representation of a supply-chain ecosystem enables more accurate prediction of malware spread. In contrast to traditional classification models and non-sequential ones that treat feature vectors

as independent, the XGNN adopts message passing and attention mechanisms for preserving context dependencies between nodes. This allows the model to learn how a corrupted part can impact its downstream. The spread map of Figure 2 visualises this process well by showing also non-connected chains of infection not understandable for flat classifiers.

The strong detection performance of our models corresponds with the findings of previous work that uses similar models to those previously used for malware-behaviour graphs (Wang et al., 2022) and network-flow graphs (Busch et al., 2021). Nevertheless, this study is different to previous works by considering supply-chain dependency graphs and moves the context of malware spreading from a device-level to an organisational-systemic one.

5.2.2 Explainability and Trustworthiness

Explainability continues to be an important factor for real-life applications of AI in the field of cybersecurity (Huang et al., 2022; Alshehri, 2025). The XGNN has higher fidelity and sparsity rates (Figure 4), which means that the explanations maintain strong correspondence with the model's logical functions and remain easy to interpret. In addition to this, our finding is in line with the claim that hybrid explainability methods—Folded techniques of local perturbation (GGNN+LIME) and structural sub-graph extraction (ExplainGCN)—can provide trade-offs between performance and interpretability (Hao et al., 2020).

The ability to see which nodes (e.g. package maintainers or libraries) result in the most risk is practically useful for cyber defence groups. It moves AI from the "black box" to a decision-support system, where human analysts can validate and act on model insights. As suggested by Kim and Kim (2024), explainability leads to trust calibration between humans and AI, such that automated predictions are understood in direct, traceable terms.

5.2.3 Structural Aspects in Propagation A number of key structural features must reflect our experimental observations in order for a model to significantly contribute towards understanding the mechanism of aggregation, including the following: The existence of torsion angles.

The feature-importance heatmap (Figure 3) shows that centrality-based features, such as Betweenness and PageRank, have strong influences on the spread of infection. High betweenness nodes tend to connect clusters which are otherwise disconnected (LXYQ 10/98). This observation is consistent with network-science studies of contagion dynamics (PastorSatorras & Vespignani, 2021). This is also consistent with cybersecurity work which suggests that high-connectivity packages and vendor hubs are dominant targets for attacks (ENISA 2024; Okafor et al. 2024).

5.3 Comparison with Existing Approaches

Traditional static and dynamic malware-analysis techniques that rely on signature extraction or the profile of sandbox behavior do not perform well when it comes to detecting zero-day, or supply-chain based threats (Tan, 2025). In contrast, the graph-representation based approach of XGNN encapsulates more latent relational evidence (e.g., shared developers,copied libraries and versions) that linear models usually can not accommodate.

5.4 Theoretical and Practical Implications

Practical Implications

On the ground, there are several reasons why the framework makes sense as a cybersecurity operations tool:

• Early detection Anomaly: The ability to expose high risk nodes and dependencies in the system means that potentially, the XGNN can support proactive patching and supplier-risk mitigation.

• Incident Forensics: Explainable sub-graphs are Alibaba-like piles of breadcrumbs that model and preserve chains of evidence for incident analysis and compliance reporting.

• Decision Support – In order not come through with the current study, they focus prehensible propagation maps and as a result on an interpretable threat scores on remediation worthiness.

These findings are in line with industry demands for AI-based, but transparent in their reasoning, security tools (Gartner, 2025; IBM Security Report, 2024).

Limitations and Future Work

However, the present study has some limitations.

Dataset Limitations: The graph dataset mixes real and synthetic graphs, and future work should evaluate the generalisability of our model using proprietary or cross-industry datasets.

scalability: Large-scale supply-chain graphs may become too computationally expensive. One possible direction for scaling it would be to use Dynamic Graph Sampling or Hierarchical Pooling (Xu et al., 2018) on top of GAT.

Dynamic Spread: Existing approaches model based only on the static dependency information. Next-generation frameworks may leverage temporal GNNs for the dynamic d representation of supply-chain state evolution (Zhou et al., 2024).

6. Conclusion

On the theoretical side, by proposing a novel graph representation learning framework on graph data and a diagnostic inference model, we advance the integration of explainable artificial intelligence (XAI) and graph representation learning in cybersecurity. It contributes to the design-science paradigm (Hevner et al., 2004) by providing a working artefact-an interpretable, scalable graph-based model-that can be extended to alternative relational threat contexts. The findings also corroborate recent claims showing that GNNs can in fact model causal spreading processes in networked systems (Zhang & Ma, 2025), indicating intriguing prospects for future research on casual explainability in cybersecurity analytics.

Practically, the implications are significant. The XGNN model can be accessed as a forensic/Predictive tool in the cyberspace national level to enterprise level cyber defence infrastructures. That caters to early detection of high-risk nodes (like central maintainer or widely used library) for proactive risk mitigation and malware control before downstream infections occur. The explainable propagation maps and feature-importance heatmaps also aid the incident investigation, compliance reporting, and supplier-risk evaluation that match the operational requirements highlighted by ENISA (2024) and IBM Security (2024).

Nevertheless, limitations remain. Using artificial datases also limits external validity, but above all; the computational cost of the model scales with the graph size. One area for future work would be to enhance this framework addressing the evolving dependencies using temporal graph neural networks (Zhou, Li, & Sun, 2024), and further dealing with privacy-preserving in collaborative multi-organisational environments as federated graph learning (Rahman, Singh, & Verma, 2025). Also, qualitative user studies could investigate how analysts make sense of and take action based on XGNN explanations in practical scenarios.

8 Conclusion This paper provides a powerful and transparent tool for enabling understanding and addressing supply-chain threats in software. It highlights that explainable graph learning is not just a technology enhancement, but a change of paradigm, connecting deep learning with human reasoning for building trust in resilient proactive cybersecurity ecosystems. With the increasing sophistication of cyber-threats, the inclusion of explainable AI within graph-analytic frameworks will significantly aid in achieving digital resilience and accountability as part of future cyber-defence strategies.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

References

[1] ENISA. (2024). Threat landscape for supply chain attacks 2024. European Union Agency for Cybersecurity.
[2] Hao, Y., Yu, H., Gui, S., & Ji, S. (2020). Explainability in graph neural networks: A taxonomic survey. arXiv. https://arxiv.org/abs/2006.12362
[3] Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.
[4] IBM Security. (2024). Cost of a Data Breach Report 2024. IBM Corporation.

[5] Li, S., Zhou, Q., & Lv, Q. (2022). Intelligent malware detection based on graph convolutional networks. Journal of Supercomputing, 78, 4182–4198.

[6] Rahman, A., Singh, S., & Verma, R. (2025). Federated graph learning for distributed cyber-threat intelligence. IEEE Access, 13, 21580–21597.

[7] Shokouhinejad, H., Razavi-Far, R., Higgins, G., & Ghorbani, A. A. (2025). On the consistency of GNN explanations for malware detection. arXiv. https://arxiv.org/abs/2504.16316

[8] Tan, Z. (2025). Advanced persistent threats based on supply chain vulnerabilities: A survey. Pre-print, University of Glasgow.

[9] Wang, X., Jiang, Y., & Liu, Z. (2022). Behaviour-based malware classification using graph neural networks. Computers & Security, 118, 102732.

[10] Zhang, L., & Ma, T. (2025). Causal graph learning for explainable threat propagation modelling. IEEE Transactions on Neural Networks and Learning Systems.

[11] Zhou, R., Li, X., & Sun, H. (2024). Temporal graph neural networks for dynamic cyber-attack prediction. Neural Computing & Applications, 36(5), 4875–4890.

[12] Asma-Ul-Husna, A. R., & Paul, G. MKR Fatigue Estimation through Face Monitoring and Eye Blinking. In *International Conference on Mechanical, Industrial and Energy Engineering (Khulna, 2014)*.

[13] Bhuiya, R. A., Hasan, M. H., Barua, M., Rafsan, M., Jany, A. U. H., Iqbal, S. M. Z., & Hossan, F. (2025). Exploring the economic benefits of transitioning to renewable energy sources. *International Journal of Materials Science*, *6*(2), 01-10.

[14] Rokunuzzaman, M., Hasan, M., & Kader, M. A. (2012). Semantic Stability: A Missing Link between Cognition and Behavior. *International Journal of Advanced Research in Computer Science*, *3*(4).

[15] Rahman, M. M., Bandhan, L. R., Monir, L., & Das, B. K. (2025). Energy, exergy, sustainability, and economic analysis of a waste heat recovery for a heavy fuel oil-based power plant using Kalina cycle integrated with Rankine cycle. *Next Research*, 100398.

[16] Neelapu, M. (2025). Predictive Software Defect Identification with Adaptive Moment Estimation based Multilayer Convolutional Network Model. Journal of Technological Innovations, 6(1).

[17] Neelapu, M. (2025). Predictive Software Defect Identification with Adaptive Moment Estimation based Multilayer Convolutional Network Model. Journal of Technological Innovations, 6(1).

[18] Neelapu, M. (2025). Predictive Software Defect Identification with Adaptive Moment Estimation based Multilayer Convolutional Network Model. Journal of Technological Innovations, 6(1).

[19] Zahid, Z., Siddiqui, M. K. A., Alamm, M. S., Saiduzzaman, M., Morshed, M. M., Ferdousi, R., & Nipa, N. N. (2025, March). *Digital Health Transformation Through Ethical and Islamic Finance: A Sustainable Model for Healthcare in Bangladesh*.

[20] Alamm, M. S., Zahid, Z., Nipa, N. N., & Khalil, I. (2025). Harnessing FinTech and Islamic Finance for Climate Resilience: A Sustainable Future Through Islamic Social Finance and Microfinance. *Humanities and Social Sciences*, *13*(3), 207-218.

[21] Zahid, Z., Amin, M. R., Alamm, M. S., Nipa, N. N., Khalil, I., Haque, A., & Mahmud, H. Leveraging agricultural certificates (Mugharasah) for ethical finance in the South Asian food chain: A pathway to sustainable development.

[22] Zahid, Z., Amin, M. R., Monsur, M. H., Alamm, M. S., Nahid, I. K., Banna, H., ... & Nipa, N. N. Integrating FinTech Solutions in Agribusiness: A Pathway to a Sustainable Economy in Bangladesh.

[23] Zahiduzzaman Zahid, M. S. A., Yousuf, M. A., Alam, M. M. A., Islam, M. A., Uddin, M. M., Parves, M. M., & Arif, S. (2025). Global Journal of Economic and Finance Research.

[24] Zahid, Z., Amin, M. R., Alamm, M. S., Meer, W., Shah, M. N., Khalil, I., ... & Arafat, E. (2025). International Journal of Multidisciplinary and Innovative Research.

[25] Zahid, Z., Amin, R., Khalil, I., Mohammed, B. A. K., & Arif, S. (2025). Regulating Digital Currencies in the EU: A Comparative Analysis with Islamic Finance Principles Under MiCA. *International Journal of Business and Management Practices (IJBMP)*, *3*(3), 217-228.

[26] Zahid, Z., & Nipa, N. N. (2024). Sustainable E-Learning Models for Madrasah Education: The Role of AI and Big Data Analytics.

[27] Zaman, Z. (2023). ইসলামিক ফিনটেক: ধারণা এবং প্রয়োগ| Islamic Fintech: Concept and Application. *ইসলামী আইন ও বিচার| Islami Ain O Bichar*, *19*(74-75), 213-252.

[28] Ferdous, J., Islam, M. F., & Das, R. C. (2022). Dynamics of citizens' satisfaction on e-service delivery in local government institutions (Union Parishad) in Bangladesh. *Journal of Community Positive Practices*, (2), 107-119.

[29] Ud Doullah, S., & Uddin, N. (2020). Public trust building through electronic governance: An analysis on electronic services in Bangladesh. *Technium Soc. Sci. J.*, *7*, 28.

[30] Ferdous, J., Foyjul-Islam, M., & Muhury, M. (2024). Performance Analysis of Institutional Quality Assurance Cell (IQAC): Ensuring Quality Higher Education in Bangladesh. *Rates of Subscription*, 57.

[31] Islam, M. F. FEMALE EDUCATION IN BANGLADESH: AN ENCOURAGING VOYAGE TOWARDS GENDER PARITY.

[32] Ferdous, J., Zeya, F., Islam, M. F., & Uddin, M. A. (2021). Socio-economic vulnerability due to COVID-19 on rural poor: A case of Bangladesh. *evsjv‡k cjøx Dbœqb mgxÿv*.

[33] Ferdous, J., & Foyjul-Islam, M. Higher Education in Bangladesh: Quality Issues and Practices.

[34] Mollah, M. A. H. (2017). *Groundwater Level Declination in Bangladesh: System dynamics approach to solve irrigation water demand during Boro season* (Master's thesis, The University of Bergen).

[35] Fuad, N., Meandad, J., Haque, A., Sultana, R., Anwar, S. B., & Sultana, S. (2024). Landslide vulnerability analysis using frequency ratio (FR) model: a study on Bandarban district, Bangladesh. *arXiv preprint arXiv:2407.20239*.

[36] Mollah, A. H. (2023). REDUCING LOSS & DAMAGE OF RIVERBANK EROSION BY ANTICIPATORY ACTION. *No its a very new study output*.

[37] Mollah, A. H. (2011). Resistance and Resilience of Bacterial Communities in Response to Multiple Disturbances Due to Climate Change. *Available at SSRN 3589019*.

[38] Haque, A., Akter, M., Rahman, M. D., Shahrujjaman, S. M., Salehin, M., Mollah, A. H., & Rahman, M. M. Resilience Computation in the Complex System. *Munsur, Resilience Computation in the Complex System*.

[39] Al Imran, S. M., Islam, M. S., Kabir, N., Uddin, I., Ali, K., & Halimuzzaman, M. (2024). Consumer behavior and sustainable marketing practices in the ready-made garments industry. *International Journal of Management Studies and Social Science Research*, *6*(6), 152-161.

[40] Islam, M. A., Goldar, S. C., Al Imran, S. M., Halimuzzaman, M., & Hasan, S. (2025). AI-Driven green marketing strategies for eco-friendly tourism businesses. *International Journal of Tourism and Hotel Management*, *7*(1), 31-42.

[41] Al Imran, S. M. (2024). Customer expectations in Islamic banking: A Bangladesh perspective. *Research Journal in Business and Economics*, *2*(1), 12-24.

[42] Islam, M. S., Amin, M. A., Hossain, M. B., Sm, A. I., Jahan, N., Asad, F. B., & Mamun, A. A. (2024). The Role of Fiscal Policy in Economic Growth: A Comparative Analysis of Developed and Developing Countries. *International Journal of Research and Innovation in Social Science*, *8*(12), 1361-1371.

[43] Al Amin, M., Islam, M. S., Al Imran, S. M., Jahan, N., Hossain, M. B., Asad, F. B., & Al Mamun, M. A. (2024). Urbanization and Economic Development: Opportunities and Challenges in Bangladesh. *International Research Journal of Economics and Management Studies IRJEMS*, *3*(12).

[44] SM, A. I., MD, A. A., HOSSAIN, M., ISLAM, M., JAHAN, N., MD, E. A., & HOSSAIN, M. (2025). THE INFLUENCE OF CORPORATE GOVERNMENT ON FIRM PERFORMANCE IN BANGLADESH. *INTERNATIONAL JOURNAL OF BUSINESS MANAGEMENT*, *8*(01), 49-65.

[45] Akter, S., Ali, M. R., Hafiz, M. M. U., & Al Imran, S. M. (2024). Transformational Leadership For Inclusive Business And Their Social Impact On Bottom Of The Pyramid (Bop) Populations. *Journal Of Creative Writing (ISSN-2410-6259)*, *8*(3), 107-125.

[46] Ali, M. R. GREEN BRANDING OF RMG INDUSTRY IN SHAPING THE SUSTAINABLE MARKETING.

[47] Hossain, M. A., Tiwari, A., Saha, S., Ghimire, A., Imran, M. A. U., & Khatoon, R. (2024). Applying the Technology Acceptance Model (TAM) in Information Technology System to Evaluate the Adoption of Decision Support System. *Journal of Computer and Communications*, *12*(8), 242-256.

[48] Saha, S., Ghimire, A., Manik, M. M. T. G., Tiwari, A., & Imran, M. A. U. (2024). Exploring Benefits, Overcoming Challenges, and Shaping Future Trends of Artificial Intelligence Application in Agricultural Industry. *The American Journal of Agriculture and Biomedical Engineering*, *6*(07), 11-27.

[49] Ghimire, A., Imran, M. A. U., Biswas, B., Tiwari, A., & Saha, S. (2024). Behavioral Intention to Adopt Artificial Intelligence in Educational Institutions: A Hybrid Modeling Approach. *Journal of Computer Science and Technology Studies*, *6*(3), 56-64.

[50] Noor, S. K., Imran, M. A. U., Aziz, M. B., Biswas, B., Saha, S., & Hasan, R. (2024, December). Using data-driven marketing to improve customer retention for US businesses. In *2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA)* (pp. 338-343). IEEE.

[51] Tiwari, A., Saha, S., Johora, F. T., Imran, M. A. U., Al Mahmud, M. A., & Aziz, M. B. (2024, September). Robotics in Animal Behavior Studies: Technological Innovations and Business Applications. In *2024 IEEE International Conference on Computing, Applications and Systems (COMPAS)* (pp. 1-6). IEEE.

[52] Sobuz, M. H. R., Saleh, M. A., Samiun, M., Hossain, M., Debnath, A., Hassan, M., ... & Khan, M. M. H. (2025). AI-driven modeling for the optimization of concrete strength for Low-Cost business production in the USA construction industry. *Engineering, technology & applied science research*, *15*(1), 20529-20537.

[53] Imran, M. A. U., Aziz, M. B., Tiwari, A., Saha, S., & Ghimire, A. (2024). Exploring the Latest Trends in AI Technologies: A Study on Current State, Application and Individual Impacts. *Journal of Computer and Communications*, *12*(8), 21-36.

[54] Tiwari, A., Biswas, B., ISLAM, M., SARKAR, M., Saha, S., Alam, M. Z., & Farabi, S. F. (2025). Implementing robust cyber security strategies to protect small businesses from potential threats in the USA. *JOURNAL OF ECOHUMANISM Учредители: Transnational Press London*, *4*(3).

[55] Hasan, R., Khatoon, R., Akter, J., Mohammad, N., Kamruzzaman, M., Shahana, A., & Saha, S. (2025). AI-Driven greenhouse gas monitoring: enhancing accuracy, efficiency, and real-time emissions tracking. *AIMS Environmental Science*, *12*(3), 495-525.

[56] Hossain, M. A., Ferdousmou, J., Khatoon, R., Saha, S., Hassan, M., Akter, J., & Debnath, A. (2025). Smart Farming Revolution: AI-Powered Solutions for Sustainable Growth and Profit. *Journal of Management World*, *2025*(2), 10-17.

[57] Saha, S. (2024). Economic Strategies for Climate-Resilient Agriculture: Ensuring Sustainability in a Changing Climate. *Demographic Research and Social Development Reviews*, *1*(1), 1-6.

[58] Saha, S. (2024). -27 TAJABE USA (150$) EXPLORING+ BENEFITS,+ OVERCOMING. *The American Journal of Agriculture and Biomedical Engineering*.

[59] Adeojo, O. S., Egerson, D., Mewiya, G., & Edet, R. (2021). The ideology of baby-mama phenomenon: Assessing knowledge and perceptions among young people from educational institutions.

[60] Orugboh, O. G. (2025). AGENT-BASED MODELING OF FERTILITY RATE DECLINE: SIMULATING THE INTERACTION OF EDUCATION, ECONOMIC PRESSURES, AND SOCIAL MEDIA INFLUENCE. *NextGen Research*, *1*(04), 1-21.

[61] Orugboh, O. G., Ezeogu, A., & Juba, O. O. (2025). A Graph Theory Approach to Modeling the Spread of Health Misinformation in Aging Populations on Social Media Platforms. *Multidisciplinary Journal of Healthcare (MJH)*, *2*(1), 145-173.

[62] Orugboh, O. G., Omabuwa, O. G., & Taiwo, O. S. (2025). Predicting Intra-Urban Migration and Slum Formation in Developing Megacities Using Machine Learning and Satellite Imagery. *Journal of Social Sciences and Community Support*, *2*(1), 69-90.

[63] Orugboh, O. G., Omabuwa, O. G., & Taiwo, O. S. (2024). Integrating Mobile Phone Data with Traditional Census Figures to Create Dynamic Population Estimates for Disaster Response and Resource Allocation. *Research Corridor Journal of Engineering Science*, *1*(2), 210-228.

[64] Orugboh, O. G., Omabuwa, O. G., & Taiwo, O. S. (2024). Predicting Neighborhood Gentrification and Resident Displacement Using Machine Learning on Real Estate, Business, and Social Datasets. *Journal of Social Sciences and Community Support*, *1*(2), 53-70.

[65] Daniel, E., Opeyemi, A., Ruth, O. E., & Gabriel, O. (2020). Understanding Childbearing for Households in Emerging Slum Communities in Lagos State, Nigeria. *International Journal of Research and Innovation in Social Science*, *4*(9), 554-560.