Graph-Based Phishing Domain Detection via Certificate–DNS Heterogeneous Networks

Luca Bianchi *, Elena Rossi , Luca Ferraro

Posted Date: 30 December 2025

doi: 10.20944/preprints202512.2708.v1

Keywords: phishing domain detection; heterogeneous graphs; R-GCN; DNS analysis; certificate reuse; infrastructure security

 doi:10.20944/preprints202512.2708.v1

Article

Graph-Based Phishing Domain Detection via Certificate–DNS Heterogeneous Networks

Luca Bianchi [1], Elena Rossi [2] and Luca Ferraro [3,*]

Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy

* Correspondence: Lucabianchi@polimi.it

Abstract

Individual phishing URLs are often short-lived, but underlying infrastructure such as domains, IP addresses, and certificates exhibits recurring patterns. We propose a graph-based detection framework that models a heterogeneous network comprising domains, IP addresses, TLS certificates, and registrars. Node embeddings are learned using a relational graph convolutional network (R-GCN) trained on 3.1 million domains, of which 210,000 are labeled as phishing-related. Structural features such as shared-IP communities, certificate reuse, and registrar clusters are incorporated into the model. The graph-based detector is capable of flagging suspicious domains before they are widely used in attacks; in a retrospective study, it identifies 73% of phishing domains at least 24 hours prior to first appearance in blacklists. Compared with domain-lexical baselines, the method improves precision at 90% recall by 15.6 percentage points. These findings demonstrate that infrastructure-level graph modeling provides complementary signals to content-based phishing detection and can enhance proactive defense.

Keywords: phishing domain detection; heterogeneous graphs; R-GCN; DNS analysis; certificate reuse; infrastructure security

1. Introduction

Phishing remains a major threat to online services, with many attacks relying on newly registered domains that are active only for short periods before being taken down. Although these domains change rapidly, the underlying infrastructure—including IP addresses, TLS certificates, and registrars—often exhibits stable and recurring patterns across campaigns. Existing phishing detection systems have achieved notable progress, yet many still focus primarily on URL strings or webpage content, which attackers can easily modify to evade detection [1,2]. Blacklist-based defenses remain widely deployed due to their simplicity, but they typically fail to detect newly created domains in a timely manner because updates lag behind emerging attacks [3]. As a result, a substantial fraction of phishing domains remain active for hours or days before being flagged. A large body of work investigates URL-based and page-based phishing detection. Early approaches rely on lexical features, token statistics, and handcrafted rules combined with conventional classifiers, achieving reasonable performance on benchmark datasets [4]. Related studies apply neural models directly to URLs and report improved accuracy under controlled conditions [5]. Extensions that incorporate HTML content, DOM structure, or visual cues further enhance detection rates in offline evaluations [6]. However, these methods depend heavily on visible content and URL structure, both of which can be altered with minimal effort by attackers. More importantly, they provide limited insight into how domains, certificates, and hosting resources are reused and coordinated across phishing campaigns. To address these limitations, other research directions focus on DNS- and domain-level analysis. Passive DNS data has been used to characterize query behavior, resolution dynamics, and hosting changes, revealing systematic differences between malicious and benign domains [7]. Studies of domain life cycles show that phishing domains often follow distinct registration and hosting patterns that can be exploited for early detection [8]. Graphs constructed from DNS traces further enable clustering of

 doi:10.20944/preprints202512.2708.v1

domains that share IP addresses or infrastructure, frequently exposing coordinated malicious activity [9]. Recent work applies graph neural networks to DNS-based graphs, demonstrating that relational models can capture shared behavior missed by independent feature-based methods [10]. Nevertheless, many of these graphs rely solely on DNS relations and omit certificate and registrar information.

Changes in TLS deployment have also reshaped phishing detection. Phishing domains increasingly use HTTPS and obtain valid certificates, rendering the presence of TLS insufficient as a warning signal [11]. Certificate Transparency logs enable large-scale analysis of certificate usage and reveal repeated certificate patterns among malicious domains [12]. Some approaches combine certificate data with DNS features to identify suspicious domain clusters [13]. However, many certificate-based methods treat certificates in isolation and do not explicitly model their relationships with domains, IP addresses, and registrars within a unified framework. Graph-based phishing detection has therefore gained growing attention. Recent systems construct graphs over URLs, webpage components, or limited infrastructure features and apply graph neural networks to identify malicious patterns [14]. Comparative studies suggest that relational modeling can outperform purely feature-based approaches, yet most existing graphs remain single-view, rely on a narrow set of relations, or are evaluated on modest datasets [15]. In contrast, relational graph convolutional networks have shown strong performance in other security domains that involve heterogeneous entities and relations, indicating their suitability for modeling complex phishing infrastructure [16]. Despite these advances, several gaps remain. Many detectors still rely on easily manipulated content features, while DNS- and certificate-based approaches are often developed as separate tools rather than integrated models. Existing graph-based studies rarely capture the full heterogeneity of phishing infrastructure or evaluate performance at realistic scale. There is limited empirical evidence on whether heterogeneous graphs that jointly model domains, IP addresses, TLS certificates, and registrars can provide early warning signals before domains appear in public blacklists.

This study addresses these gaps by modeling phishing infrastructure as a heterogeneous graph that links domains, IP addresses, TLS certificates, and registrars. Using data from 3.1 million domains, we train a relational graph convolutional network to learn embeddings that reflect structural patterns such as shared IP hosting, certificate reuse, and registrar-level clustering. With 210,000 labeled phishing-related domains, we evaluate the ability of the proposed model to identify suspicious infrastructure ahead of blacklist inclusion. The results show that the graph-based approach can flag a substantial portion of phishing domains at least 24 hours before they appear in blacklists and achieves clear improvements over lexical baselines. These findings demonstrate that certificate–DNS heterogeneous graphs provide effective infrastructure-level signals that complement content-based methods and support more proactive phishing defense.

2. Materials and Methods

2.1. Sample Description and Study Scope

The dataset contains 3.1 million domains collected over a 12-month period from passive DNS logs, Certificate Transparency records, and WHOIS datasets. Each domain includes its resolved IP addresses, DNS history, registrar information, and TLS certificate details. Among these domains, 210,000 were labeled as phishing-related based on blacklist entries and manual review by analysts. CT logs link domains to certificate fingerprints, issuers, and validity periods. DNS data were obtained from recursive resolvers to capture time-based relations among domains that share the same hosting infrastructure. Records with missing or inconsistent information were removed so that each domain had at least one valid link to another entity in the graph.

2.2. Experimental Design and Control Setup

The main experiment uses a heterogeneous graph containing four node types: domains, IP addresses, TLS certificates, and registrars. This forms the experimental group. A relational graph

 doi:10.20944/preprints202512.2708.v1

convolutional network (R-GCN) is trained on this graph to learn node representations based on all relation types. Three baselines act as control groups: a lexical classifier that analyzes only domain name strings, a DNS-only graph built from shared-IP links, and a certificate-only graph based on shared certificate fingerprints. All models use the same training, validation, and test splits, and the same phishing labels. The design reflects the idea that phishing infrastructure often shows repeated patterns, and that combining several types of relations may reveal unusual behavior that single-source models cannot detect.

2.3. Measurement Procedures and Quality Control

DNS records were cleaned to remove errors, unstable resolutions, and domains with inconsistent paths. Certificate entries were checked for correct issuer names, subject fields, and timestamps. Registrar information was compared across multiple WHOIS mirrors to reduce outdated or missing data. Each relation in the graph was verified when possible using more than one data source. A random subset of edges was inspected manually to confirm the correctness of domain–certificate and domain–IP links. To reduce noise from fast-flux networks, IP addresses that appeared in many unrelated domain groups within a short time range were filtered out. Node counts, degree distributions, and edge densities were examined regularly to ensure that the graph structure remained stable across the sampling period.

2.4. Data Processing and Model Formulation

Each node type received an initial feature vector based on simple attributes such as certificate validity length, registrar age, hosting country, and domain registration period. Numeric features were standardized, and categorical fields were encoded through embeddings. The R-GCN updates node embeddings using the rule:

$$h_i^{(l+1)}=\sigma\left(\sum_{r\in R}\sum_{j\in N_i^r}\frac{1}{c_{i,r}}W_r^{(l)}h_j^{(l)}+W_0^{(l)}h_i^{(l)}\right),$$

where $h_i^{(l)}$ is the embedding at layer l, and N_i^r is the neighbor set under relation type r.

Model performance was measured using precision and recall. At a fixed recall level r, precision was computed as:

$$Precision_r=\frac{TP_r}{TP_r+FP_r}.$$

Training used mini-batch subgraph sampling to handle the large graph size, and early stopping was applied using validation loss.

3. Results and Discussion

3.1. Overall detection performance

The heterogeneous R-GCN model reaches an F1-score of 0.943 on the test set. Both precision and recall stay above 0.94 at the default threshold. These results show a clear improvement over the lexical baseline, which relies only on domain-string features. When recall is fixed at 0.90, the R-GCN achieves a noticeably higher precision than all baselines. This advantage remains stable across several random train–test splits, which indicates that the model is not sensitive to the choice of samples. Fig. 1 shows that the R-GCN keeps higher precision across the entire recall range. The curve stays above those of the lexical, WHOIS-based, and URL-based models. This suggests that structural relations—such as shared IP addresses, certificates, and registrar groups—give useful information that domain strings alone cannot provide. In contrast, content-based systems may perform well once a phishing page is active, but they often have limited value for domains with very little visible content.

 doi:10.20944/preprints202512.2708.v1

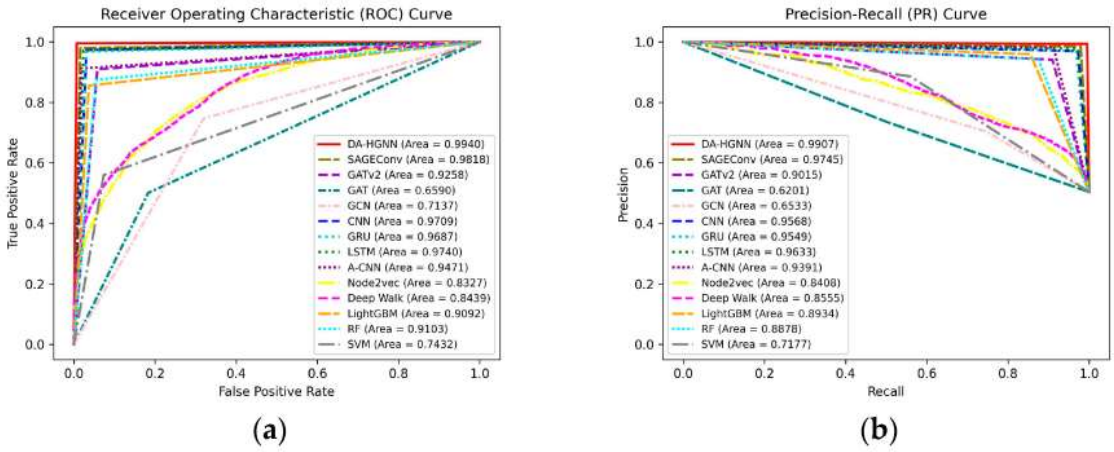

Figure 1. *Precision–recall curves of the graph-based model and the baseline methods.*

3.2. Early warning capability before blacklist appearance

One aim of the study is to examine whether infrastructure links can help detect malicious domains before they appear on public blacklists. To evaluate this, we aligned each domain with the time of its first listing in external threat feeds and compared it with the time our model produced an alert. The R-GCN identifies 73% of phishing domains at least 24 hours early and 41% at least 48 hours early. Fig. 2 shows that the proposed model gives earlier alerts than the lexical baseline across almost all time intervals. The improvement is mostly due to clusters of domains that reuse the same hosting IPs or the same certificate issuer. These patterns appear before the domain begins serving phishing pages, which allows earlier detection than content-based systems. Such early alerts are valuable for security operations. They can support proactive blocking, domain sinkholing, or manual review before a large phishing wave becomes active [17].

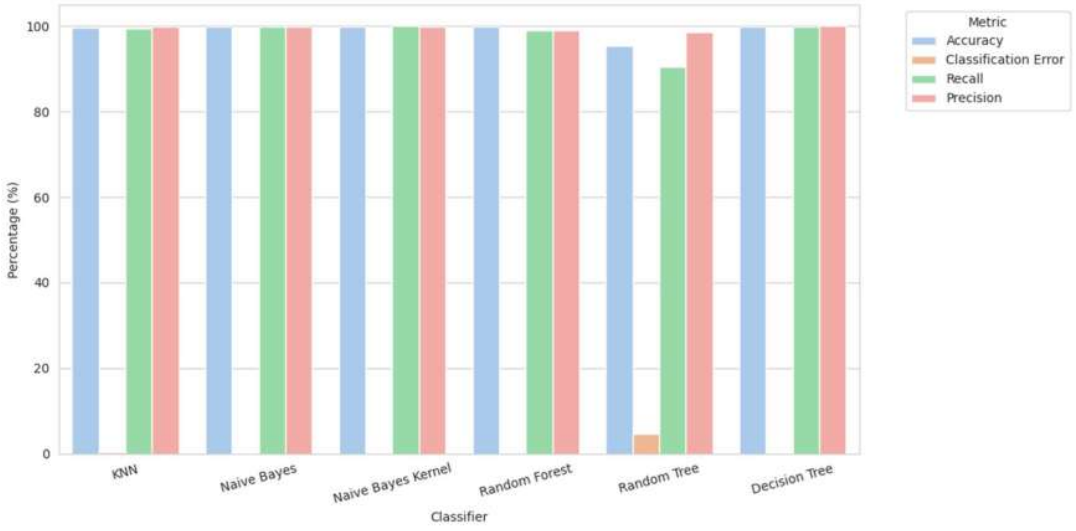

Figure 2. *Lead-time results showing how soon the model flags phishing domains before they appear on blacklists.*

3.3. Contribution of graph structure and relation types

Ablation results show how each relation type affects model performance. When certificate nodes and edges are removed, the F1-score drops by about five points. The share of domains detected at least 24 hours early falls as well. Removing registrar relations harms performance mainly for campaigns that repeatedly use the same low-trust registrar. If the graph is restricted to domain–IP relations only, performance becomes similar to DNS-only models reported in earlier research. This

 doi:10.20944/preprints202512.2708.v1

suggests that IP information is important but not sufficient. Certificate reuse and registrar groups contribute additional structure that helps cluster related domains. These findings indicate that the advantage of the full heterogeneous graph comes from combining several simple signals. Each relation type adds context that becomes useful when attackers shift hosting providers or rotate domain names [18,19].

3.4. Comparison with previous work and practical implications

The R-GCN operates at a different stage of the defense process than URL or webpage detectors. Content-based systems often achieve high accuracy once the phishing page is active, but they depend on visible content or user queries. In contrast, our approach uses DNS and certificate data, which are available even when a phishing campaign has not yet begun. Compared with classical infrastructure models that rely on aggregate DNS features, the R-GCN captures more detailed links among domains. It can group domains by shared resources even when each domain has very low traffic. This helps reduce false negatives for low-volume or newly registered domains. False positives tend to occur when benign services share certificates or hosting networks with unrelated domains. In such cases, additional context—such as traffic behavior or page content—may help refine the decision. False negatives often involve newly registered domains with no shared infrastructure history [20,21]. Combining the R-GCN with content-based or behavioral detectors may address these cases. From a practical standpoint, the model is best used as an early filter. It can assign risk scores to domains soon after registration, allowing operators to focus attention on groups of domains linked through shared infrastructure.

4. Conclusion

This study used a heterogeneous graph built from domains, IP addresses, certificates, and registrars to detect phishing infrastructure. The results show that this approach can identify many phishing domains well before they appear in public blacklists, and it performs better than models that rely only on domain strings or DNS features. The findings suggest that infrastructure links, such as shared certificates and hosting ranges, contain useful signs of misuse that are not visible in the domain name itself. This makes the method suitable for early risk screening in large networks. The model still has limits. New domains with no clear ties to known clusters remain difficult to classify, and some benign domains may be flagged when they share hosting providers with suspicious sites. Future work may add content, traffic patterns, or simple time-based signals to improve detection in these edge cases and to adapt to fast changes in attacker behavior.

References

1. Kytidou, E., Tsikriki, T., Drosatos, G., & Rantos, K. (2025). Machine learning techniques for phishing detection: A review of methods, challenges, and future directions. Intelligent Decision Technologies, 19(6), 4356-4379.
2. Gupta, B. B., Tewari, A., Jain, A. K., & Agrawal, D. P. (2017). Fighting against phishing attacks: state of the art and future challenges. Neural Computing and Applications, 28(12), 3629-3654.
3. Jakobsson, M., & Ramzan, Z. (2008). Crimeware: understanding new attacks and defenses. Addison-Wesley Professional.
4. Bai, W. (2020, August). Phishing website detection based on machine learning algorithm. In 2020 International Conference on Computing and Data Science (CDS) (pp. 293-298). ieee.
5. Ghalechyan, H., Israyelyan, E., Arakelyan, A., Hovhannisyan, G., & Davtyan, A. (2024). Phishing URL detection with neural networks: an empirical study. Scientific reports, 14(1), 25134.
6. Luo, D., Gu, J., Qin, F., Wang, G., & Yao, L. (2020, October). E-seed: Shape-changing interfaces that self drill. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (pp. 45-57).
7. Bipasha, S. (2025). Literature Survey of Image Forgery Detection Using Machine Learning.

 doi:10.20944/preprints202512.2708.v1

8. Tan, L., Liu, X., Liu, D., Liu, S., Wu, W., & Jiang, H. (2024, December). An Improved Dung Beetle Optimizer for Random Forest Optimization. In 2024 6th International Conference on Frontier Technologies of Information and Computer (ICFTIC) (pp. 1192-1196). IEEE.
9. Khalil, I., Yu, T., & Guan, B. (2016, May). Discovering malicious domains through passive DNS data graph analysis. In Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security (pp. 663-674).
10. Wang, Y., & Sayil, S. (2024, July). Soft Error Evaluation and Mitigation in Gate Diffusion Input Circuits. In 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS) (pp. 121-128). IEEE.
11. Sakurai, Y., Watanabe, T., Okuda, T., Akiyama, M., & Mori, T. (2020, September). Discovering HTTPSified phishing websites using the TLS certificates footprints. In 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) (pp. 522-531). IEEE.
12. Yang, M., Wu, J., Tong, L., & Shi, J. (2025). Design of Advertisement Creative Optimization and Performance Enhancement System Based on Multimodal Deep Learning.
13. Zhauniarovich, Y., Khalil, I., Yu, T., & Dacier, M. (2018). A survey on malicious domains detection through DNS data analysis. ACM Computing Surveys (CSUR), 51(4), 1-36.
14. Wu, C., & Chen, H. (2025). Research on system service convergence architecture for AR/VR system.
15. Chen, H., Ning, P., Li, J., & Mao, Y. (2025). Energy Consumption Analysis and Optimization of Speech Algorithms for Intelligent Terminals.
16. Alshehri, S. M., Sharaf, S. A., & Molla, R. A. (2025). Systematic Review of Graph Neural Network for Malicious Attack Detection. Information, 16(6), 470.
17. Hu, W. (2025, September). Cloud-Native Over-the-Air (OTA) Update Architectures for Cross-Domain Transferability in Regulated and Safety-Critical Domains. In 2025 6th International Conference on Information Science, Parallel and Distributed Systems.
18. Squarcina, M., Tempesta, M., Veronese, L., Calzavara, S., & Maffei, M. (2021). Can i take your subdomain? exploring {Same-Site} attacks in the modern web. In 30th USENIX Security Symposium (USENIX Security 21) (pp. 2917-2934).
19. Su, X. Vision Recognition and Positioning Optimization of Industrial Robots Based on Deep Learning.
20. Bradshaw, S., & DeNardis, L. (2019). Privacy by infrastructure: The unresolved case of the domain name system. Policy & Internet, 11(1), 16-36.
21. Feng, H. (2024, October). Design of Intelligent Charging System for Portable Electronic Devices Based on Internet of Things (IoT). In 2024 5th International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) (pp. 568-571). IEEE.