World Journal of Information Technology
Print ISSN: 2959-9903
Online ISSN: 2959-9911
DOI: https://doi.org/10.61784/wjit3117

A DATA-DRIVEN POLICY OPTIMIZATION ANALYSIS FRAMEWORK FOR CYBERCRIME GOVERNANCE BASED ON GNN

ZhiFei Xu[1*], PengYu Chen[2], YanJun Fan[3]
[1]*School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, Jiangsu, China.*
[2]*School of Computer and Electronic Information/School of Artificial Intelligence, Nanjing Normal University, Nanjing 210023, Jiangsu, China.*
[3]*School of Intensive Studies, Nanjing Normal University, Nanjing 210023, Jiangsu, China.*
[*]*Corresponding Author: ZhiFei Xu*

Abstract: In today's digital age, the rapid growth of information technology has reshaped global economic, social, and cultural landscapes, yet many countries face challenges in coordinating and implementing effective cybersecurity policies. This paper first applies PCA, SEM, GWR, and LSTM to normalize multidimensional indicators of global cybercrime data, cybersecurity policy metrics, and demographic characteristics. It then builds a data-driven policy optimization framework based on graph neural networks (GNN), incorporating spatiotemporal feature modeling, policy effectiveness evaluation, and cross-border collaborative game analysis, along with multi-source data fusion and dynamic prediction. Experimental results demonstrate high prediction accuracy and policy optimization capability, providing a scientific basis for policymakers to formulate targeted cybersecurity strategies and address evolving cyber threats, thereby improving global cybersecurity levels.
Keywords: Cybercrime governance; Data-driven policy; Graph neural network; Multidimensional analysis; Space-time modeling

1 INTRODUCTION

Modern technology accelerates global connectivity but also expands the threat of cybercrime. Due to its transnational and concealed nature, cybercrime presents complex challenges for investigation and jurisdiction. Moreover, many institutions conceal security vulnerabilities to avoid reputational damage, further complicating governance. In response, countries have developed national cybersecurity policies, and the ITU, as a UN specialized agency, plays a key role in promoting international standards and cooperation.

Cybersecurity policy formulation faces multi-dimensional complexity: policy effects are shaped by individual digital literacy, regional infrastructure disparities, and national legal frameworks. These factors interact dynamically—for example, technological progress alters crime patterns, while demographic shifts affect technology penetration. However, existing linear evaluation methods fail to capture such nonlinear correlations, leading to biased results. Hence, a new analytical framework that integrates spatial and temporal dimensions and quantifies heterogeneous factor interactions is urgently needed to support precise policy design and optimization.

Liu Jianwen leverages the Sparrow Search Algorithm (SSA) to optimize and supplement traditional LSTM, constructing an SSA-LSTM model that provides a technical solution for addressing diverse and complex cybersecurity threats. While its situation prediction demonstrates excellent performance in efficiency and accuracy, it fails to account for the cross-impact of multi-level factors on policy effectiveness [1]. Li Siyuan proposed an anomaly detection method based on the generative adversarial network (GAN) architecture based on the long short-term memory recurrent neural network (LSTM-RNN) as the basic learning model, which solves the problem that the global heterogeneity distribution of cybercrime above is difficult to predict the trend [2]. The literature employs convolutional neural networks for device identification, malicious intrusion prevention, and encrypted traffic detection, thereby solving the difficulty of normalizing multi-dimensional index extraction from global cybercrime data, national cybersecurity policy indicators, and demographic characteristics [3-5]. The literature is a study of the network security evaluation system, which provides a solution to the problem that it is difficult to construct an evaluation system for network security-related policies [6,7]. However, there are still problems that fail to take into account the complex dynamic interaction between other factors [8,9].

Current research lacks a unified framework for jointly analyzing spatio-temporal crime patterns, policy indicators, and demographic factors. This paper applies PCA, SEM, GWR, and LSTM to normalize global cybercrime data, policy indicators, and demographic characteristics, then constructs a GNN-based data-driven policy optimization framework with spatio-temporal modeling, policy evaluation, and cross-border game analysis. Experimental results demonstrate high prediction accuracy and policy optimization capability, offering a scientific basis for targeted cybersecurity strategies and enhanced global cybersecurity.

2 METHODS

2.1 GNN-PSGL Algorithm

2.1.1 Definition and derivation of GNN

Graph Neural Networks (GNNs) are deep learning models for graph-structured data. Their core idea is to update node representations by aggregating information from neighboring nodes via message passing. For example, Graph Convolutional Networks (GCNs) explicitly use a normalized adjacency matrix for global information propagation. GNNs feature permutation invariance and local perception, making them naturally suited for graph data. By designing different aggregation functions (e.g., attention, spatio-temporal convolution), GNNs can be extended to frameworks such as GNN-PSGL to integrate complex features like policy indicators and spatio-temporal dynamics.

2.1.2 The role of GNN and the reasons for selection

GNN is a deep learning model for graph-structured data, with core functions including neighbor information aggregation via message passing, learning complex non-Euclidean spatial relationships, and supporting node, edge, or graph-level prediction tasks.

In this study, GNN encodes discrete global crime data, policy indicators, and demographic characteristics into a unified graph structure, overcoming traditional methods' difficulty with heterogeneous data. It captures spatio-temporal crime propagation by coupling GNN with LSTM and transforms SEM-derived policy path coefficients into graph attention weights to deduce causal chains like legal perfection → technology input → crime rate. The graph structure naturally matches the interactive system of national policies and crime flows, and GNN's flexible architecture integrates methodologies such as SEM and game theory for transnational collaboration—a capability not achievable with CNN or RNN.

2.1.3 Definition and joint method of PSGL

PSGL is a joint framework combining Principal Component Analysis (PCA), Structural Equation Model (SEM), Geographically Weighted Regression (GWR), and Long Short-Term Memory network (LSTM). PCA is an unsupervised linear dimensionality reduction method that preserves maximal variance. SEM combines factor analysis and path analysis to test causal relationships among observed and latent variables. GWR extends linear regression to capture spatially varying relationships. LSTM, a special recurrent neural network, learns long-term dependencies in sequential data. Together, these methods normalize and standardize multi-dimensional indicators from global cybercrime data, national cybersecurity policies, and demographic characteristics, enabling GNN-based cybercrime trend prediction and scenario analysis under different policy and socio-economic conditions.

2.2 Cybercrime Governance

Cybercrime, including fraud, data leakage, ransomware, DDoS attacks, and APTs, causes severe economic and security damage. Its transnational nature, legal loopholes, and emerging technologies (AI, blockchain, quantum computing) complicate governance, demanding an intelligent framework that integrates multi-dimensional data and dynamic policy evaluation. Such prediction requires cross-modal fusion of heterogeneous data, spatio-temporal hotspot and trend capture, causal policy reasoning, game-theoretic transnational coordination, and robustness to missing data in developing countries.

2.3 Construction Of Multidimensional Index System

2.3.1 Multidimensional index selection

In the dimension of crime characteristics, the article select the core indicators such as attack frequency, economic loss amount and attack type distribution. These data objectively reflect cybercrime scale and harm. The policy dimension uses ITU's GCI, focusing on legal framework, technical protection, and organizational collaboration—internationally comparable indicators covering key cybersecurity policy elements. The socio-economic dimension includes internet penetration, higher education proportion, and per capita GDP. Together, they form a causal chain of crime performance, policy intervention, and social environment, laying a foundation for SEM path analysis and GNN feature fusion.

From a research design perspective, this indicator selection is highly targeted. Crime characteristics focus on spatio-temporal attack patterns as basic prediction inputs. Policy indicators align with governance system links to directly support policy evaluation. Socio-economic indicators reveal crime-breeding conditions for source control. This strategy enables the GNN-PSGL framework to fully capture the complex system of cybercrime governance while maintaining computational efficiency and interpretability, providing reliable data support for policy optimization.

2.3.2 Multidimensional index calculation

In this study, a systematic data processing process was used to standardize the multi-dimensional indicators. For cybercrime data, logarithmic transformation is first performed on continuous variables such as attack frequency and economic loss to eliminate the long tail distribution, and then Z-score standardization is used to make it conform to the standard normal distribution. The formula of Z-score standardization is:

$$(X-\mu)/\sigma \tag{1}$$

For policy evaluation indicators, based on the original scores of the six dimensions of ITU-GCI, the range method is normalized to the [0, 1] interval ,whose formula is:

$$(X-X_{min})/(X_{max}-X_{min}) \tag{2}$$

And the semantic features of the policy text are retained and converted into a 300-dimensional vector by Word2Vec. In

the demographic characteristics, the inverse sine transform is used for the proportional data of Internet penetration rate, and the Box-Cox transform is used for economic indicators such as GDP. In particular, considering the heterogeneity of cross-country data, the article calculate standardized parameters by regional grouping (such as μ and σ by continent) to avoid the data of developed countries dominating the global distribution.

3 METHODOLOGY

To deeply explore the global distribution of cybercrime and identify high-risk countries, multi-source heterogeneous data are integrated. Ten years of cybercrime data are collected from INTERPOL and Kaspersky, covering phishing, malware, and data leakage with details on origin, target, timestamp, and economic losses. These are enriched with World Bank geospatial data and UN population distribution data.

The article use a linear model to explore the relationship between IP addresses and cybercrime rates, including the number of IP addresses, the IP address cybercrime threat coefficient score, and the IP address involves the scope of attack.

As a basic and practical data analysis tool, the linear model provides a unique perspective and method for this research. Based on the assumption that there is a linear relationship between variables, the linear model describes the change law between independent variables and dependent variables by establishing mathematical equations. The calculation formula is:

$$CCCR=\beta_0+\beta_1 \cdot IPN+\beta_2 \cdot IPCTCS+\beta_3 \cdot IPRAT+\epsilon \tag{3}$$

Among them, β0 means intercept term, β1, β2 and β3 means the coefficient of each variable, ε means random error term, indicating the influence of other unconsidered factors.

Through the above comprehensive formula, the article can quantify the relationship between education level and network crime rate, so as to better analyze the probability of network crime through IP address.

This study refines five dimensions (law, technology, organization, capacity building, international cooperation) from GCI and OECD into over 20 indicators including data protection and infrastructure investment. Using government and institutional data, SEM constructs a causal model of policy factors and cybercrime via path analysis, while PCA examines influences of legal severity, perfection, enforcement, population, and income on cybercrime.

The article construct the normalized importance of features to evaluate the impact of one of these factors on cybercrime. The calculation formula is:

$$\mathrm{IOFN}(f_j)=\frac{\sum_{k=1}^{K} |v_{kj}| \cdot \lambda_k}{\sum_{i=1}^{7} \sum_{k=1}^{K} |v_{ki}| \cdot \lambda_k} \tag{4}$$

Through the above comprehensive formula, the article can quantify the impact of the factors to be explored on cybercrime. This method combines the dimensionality reduction ability and feature importance analysis of PCA, which can effectively reveal the role of various factors, so as to provide data support for policy formulation and resource allocation.

After that, the article made a more detailed analysis of the three aspects of education level, Internet penetration rate and population.

The article use a linear model to explore the relationship between education level and cybercrime rate, in which the education level is replaced by the number of college students.

As a basic data analysis tool, the linear model assumes a linear relationship between variables and describes it through mathematical equations. For example, when examining education level and cybercrime rate, the model assumes a stable linear influence, which simplifies analysis for preliminary insights despite real-world complexity. Its calculation formula is:

$$C=\beta_0+\beta_1 P+\epsilon \tag{5}$$

Among them, C : network crime rate (dependent variable), β0 : intercept term, which indicates the network crime rate when P = 0, β1 : slope, which indicates the change of network crime rate for each unit increase in the number of college students, ε : random error term, which indicates the influence of other factors not considered.

Through the above comprehensive formula, the relationship between education level and cybercrime rate can be quantified, so as to better analyze the probability of cybercrime through education level.

The article also construct a mathematical model to describe the relationship between Internet penetration and cybercrime. This model can be used to analyze and predict the trend of network crime rate under different Internet penetration rates. It expresses this relationship through a function :

$$C(P)=\alpha \cdot P^{\beta} \cdot e^{-\gamma P}+\delta \tag{6}$$

Explanation:

a. : Indicates the power-law effect of Internet penetration rate on cybercrime rate. When β > 1, the cybercrime rate increases with the penetration rate. When β < 1, the growth slows down.

b. : Indicates that when penetration is high, the growth of cybercrime rates is inhibited (for example, due to increased awareness of cybersecurity).

c. α : Proportional coefficient, control the overall size.

d. δ : The underlying crime rate, which may exist even when Internet penetration is zero.

Through the above comprehensive model, the relationship between Internet penetration rate and cybercrime rate can be quantified, so that the probability of cybercrime can be analyzed through Internet penetration rate.
The article use the logistic model to explore the relationship between population and cybercrime rate.
The logistic model can effectively deal with the nonlinear relationship, which is very important for studying the complex relationship between population and cybercrime rate. With the growth of population, the change of population structure and the difference of population distribution, the network crime rate is not a simple linear change. For example, in densely populated areas with a high degree of network penetration, the number of potential opportunities and targets for cybercrime is increasing, which may lead to a non-linear increase in crime rates. Its calculation formula is :

$$P(C=1|P)=\frac{1}{1+e^{-(\beta_0+\beta_1 P)}} \tag{7}$$

Among them, P (C = 1 | P) : under the condition of given population size P, the probability of cybercrime, β0 : intercept term, represents the logarithmic probability (log-odds) when P = 0, β1 : regression coefficient, represents the influence of population size P on the probability of cybercrime.
After completing the above work, the article construct a graph neural network (ST-GNN) model. This model combines graph neural network (GNN) and time series models (such as LSTM) to capture spatial and temporal dependencies in cybercrime prediction tasks. The model is divided into four modules : space module, time module, hybrid module and prediction module, and also includes attention mechanism.
The article input the node characteristics of the graph (such as crime rate, IP address traffic, etc.), the edge weights of the graph (such as geographic distance, communication frequency, etc.) and time series data to output the future network crime rate prediction value. The calculation formula is :

$$\mathrm{FCCR}_t^{t+1}=\mathbf{W}_c \cdot \sigma\left(\mathbf{W}_f \cdot \left[\sigma\left(\sum_{t' \in N(t_t)} w_{tt'}^t \cdot \mathbf{W}_s \cdot \mathbf{x}_{t'}^t\right);\mathrm{TCN}(\mathbf{h}_t^{t-Tt})\right]\right)+b_c \tag{8}$$

This formula incorporates spatial dependency modeling to capture crime spread patterns across geospatial space and temporal modeling to detect crime trends (e.g., upward/downward movements or seasonal fluctuations). By preprocessing historical cybercrime data, extracting key features, and optimizing model parameters with cross-validation, prediction accuracy is improved. Simulation results then inform practical cybersecurity policy recommendations—such as strengthening supervision and technology investment in predicted high-incidence areas and adjusting policy priorities under different scenarios—covering policy direction, measure optimization, and resource allocation.

4 EXPERIMENTAL VERIFICITION

This paper collects ten years of cybercrime data from INTERPOL and Kaspersky, covering phishing, malware, and data leakage with details on origin, target, timestamp, and economic losses. These are enriched with World Bank geospatial data and UN population distribution. Referring to the GCI and OECD frameworks, over 20 specific indicators are refined from five dimensions (law, technology, organization, capacity building, international cooperation), including data protection laws, critical infrastructure investment, and emergency response structures, and are applied to the analysis framework.The specific results are as follows:

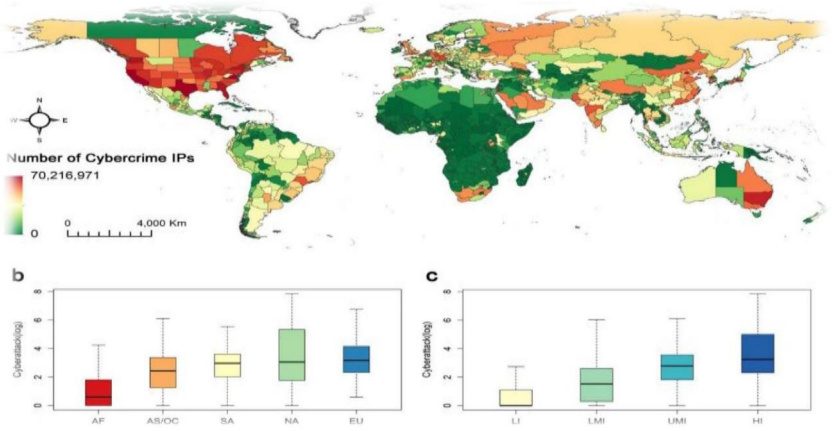

Figure 1 Number of Cybercrime IPs
Source: http://bzdt.ch.mnr.gov.cn/

Figure 1 is a global network crime event heat map, which can clearly see that the United States and Canada in North America, the United Kingdom, Germany, France and other countries in Europe, as well as Japan and South Korea in Asia, show a very significant high network crime density. Taking the United States as an example, California, as the core area of the technology industry, has brought together many Internet giants and innovative enterprises, such as Apple and Google in Silicon Valley. This is not only the forefront of global scientific and technological innovation, but

also the goal of cyber criminals. Using geographic information system (GIS) technology and spatial autocorrelation analysis,jsut like figure 2, it can be clear that there is a strong spatial aggregation of cybercrime in these areas. In New York, due to its important position as an international financial center, e-commerce, online finance and other businesses are large in scale, attracting a large number of cyber criminals to covet the economic interests.

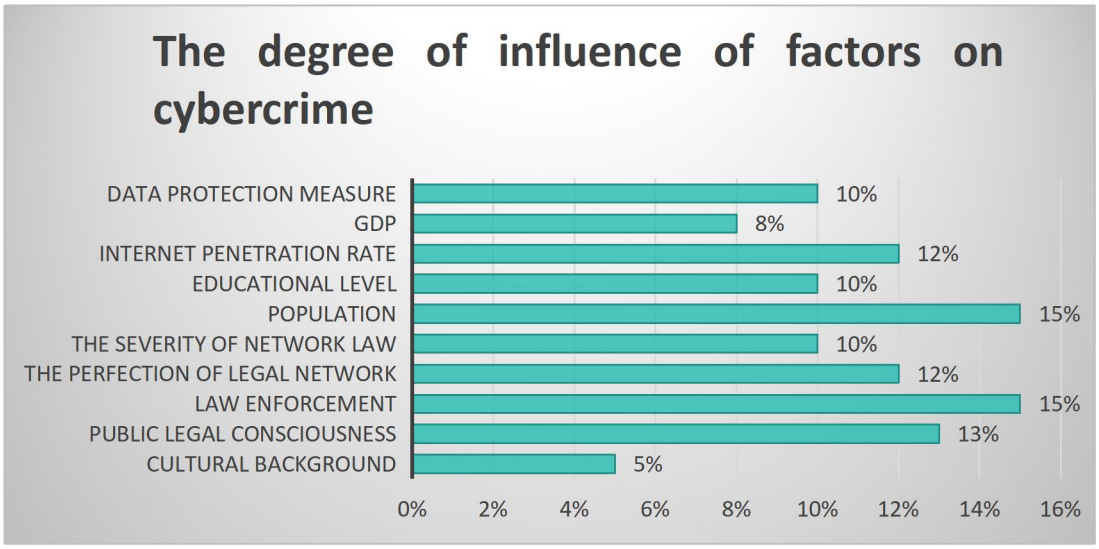

Figure 2 The Degree of Influence of Factors on Cybercrime

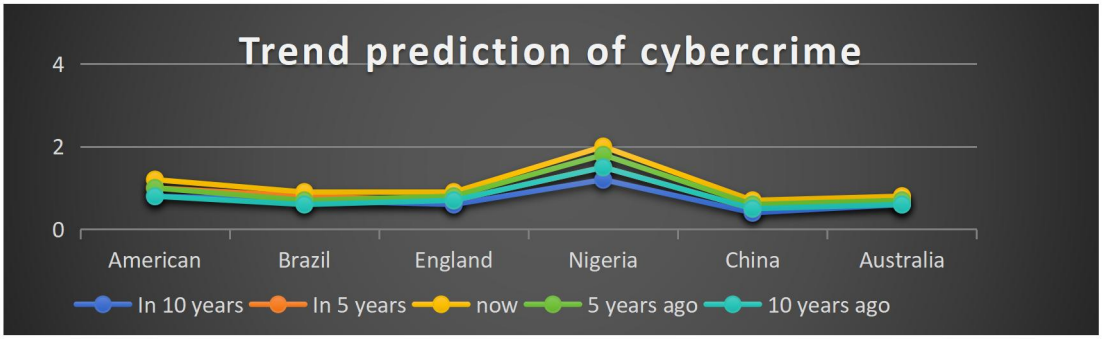

Figure 3 Trend Prediction of Cybercrime

The US cybercrime rate has risen over the past decade but is expected to decline in five to ten years, like the figure 3, indicating early cybersecurity measures lagged. Brazil's rate remains high with past increases; despite a projected decline, stronger cybersecurity construction is needed. The UK shows a stable rate with a future downward trend, allowing policy consolidation and enhanced international cooperation. Nigeria's rate is high with significant past increases; though a decline is expected, the situation remains grim, requiring comprehensive governance improvements including law enforcement and public awareness. China's rate has been stable and low for a decade and continues to decline, calling for reinforcement of existing policies and application of emerging technologies. Australia's rate has fluctuated but will decline in the future, necessitating optimized strategies focused on long-term prevention.

This innovative framework integrates policy science with graph neural networks, overcoming traditional limits in data integration, policy simulation, and cross-border optimization to shift from single-point prediction to systemic governance. As the first multi-modal system handling policy text, spatio-temporal crime patterns, and socio-economic indicators, it improves prediction accuracy and timeliness while offering quantifiable, verifiable decision tools that resolve policy lag and fragmentation, moving global cybercrime governance toward data-driven, intelligent intervention.

5 CONCLUSION

In the digital age, many countries face difficulties in coordinating and implementing effective cybersecurity policies. This paper first applies PCA, SEM, GWR, and LSTM to normalize multi-dimensional indicators from global cybercrime data, national cybersecurity policies, and demographic characteristics. It then builds a GNN-based data-driven policy optimization framework integrating spatio-temporal feature modeling, policy effectiveness evaluation, and transnational collaborative game analysis with multi-source data fusion and dynamic prediction. Experimental results show high prediction accuracy and policy optimization capability. Using the GCI evaluation system and official data, the study finds that cybercrime exhibits significant spatial agglomeration linked to digital economy development, network infrastructure, and cybersecurity maturity. GWR reveals spatial heterogeneity in demographic impacts. SEM and PSM show that legal, technological, organizational, capacity-building, and

international cooperation dimensions intertwine to affect cybercrime control, with technological innovation and international cooperation playing key roles. A dynamic game model reveals that emerging technologies both enable cybercrime and provide protective tools, while imbalances in development and protection lag worsen risks. Random forest and social network analysis confirm complex nonlinear relationships between demographic characteristics and cybercrime. Based on complex network and game theory, the current international cooperation network is found to be structurally unbalanced and insufficiently deep. Incentive-compatible mechanisms and blockchain-based information sharing platforms are expected to enhance transnational cybercrime governance. This study provides a comprehensive theoretical framework and practical guidance for policymakers, practitioners, and researchers to improve global cybersecurity governance.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Liu Jianwen. Research and Application of Network Security Situation Awareness Technology Based on Big Data. Nanchang University, 2024.
[2] Li Siyuan. Research on abnormal behavior detection method of network traffic in railway signal system. Beijing Jiaotong University, 2024.
[3] Peng Kaikang. Malicious encrypted traffic identification based on convolutional network and attention mechanism. Nanhua University, 2024.
[4] Source. Research on network intrusion detection method based on feature selection and multi-scale fusion. Xi 'an University of Science and Technology, 2023.
[5] Gao Zihan, Cheng Lu, Zhou Aiping. Internet of Things device identification method based on convolutional neural network. Computer and network, 2025, 51(04): 363-368.
[6] Cao Shunshun, Wang Shiyi, Hu Xiaoming. Research on blockchain security evaluation under network security level protection. Network security technology and application, 2025(07): 1-4.
[7] Yang Chunxia, Tao Kerui, Song Yongsheng. Ship network security state recognition based on deep learning algorithm. ship science and technology, 2023, 45(21): 193-196.
[8] Yang Jiale. A blockchain eclipse attack detection and defense method based on CNN. Nanjing University of Posts and Telecommunications, 2023.
[9] Meng Weihang. Application of Neural Network in Computer Network Security Evaluation. Metallurgy and Materials, 2023, 43(07): 157-159.