Journal of Electronics and Information Science (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2025.100106
ISSN 2371-9524 Vol. 10 Num. 1

Research on Cybercrime Prevention and Control Strategies Based on K-means++ and PSO Algorithm

Jiaxing Lv[1,*], Yu Ji[2], Jianli Zhang[3]

[1]School of Food and Health, Beijing Technology and Business University, Beijing, China
[2]School of Mathematics and Statistics, Beijing Technology and Business University, Beijing, China
[3]School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing, China
**Corresponding author: 2203023113@st.btbu.edu.cn*

Keywords: Cybercrime; K - means++ clustering algorithm; multiple linear regression; policy effectiveness; particle swarm optimisation algorithm

Abstract: This paper focuses on the field of cybercrime, analysing its global pattern and the effectiveness of policy responses through a multimethod approach. The World Cybercrime Index (WCI) is constructed with the help of experts' experience, which reveals the characteristics of the global cybercrime risk distribution, i.e. Europe and North America have a high prevalence of cybercrime, followed by Asia, and South America and Africa have a low prevalence of cybercrime. The K-means++ clustering algorithm is used to classify the risk of 97 countries/regions, and the results match the actual distribution. In the research of policy effectiveness, we constructed a performance score index and found that there is a mutually reinforcing relationship between cybercrime risk and the level of cybersecurity construction; we defined the security index S, and constructed a regression model by combining the political data and legal density of the ITU, and concluded that the legal measures are the most effective in improving the security index. The particle swarm optimisation algorithm is used to explore the optimal political scenarios, provide decision-making reference for legislators, and analyse the reasons for the differences in legal density, and the research results have important reference value for the formulation of global cybercrime prevention and control policies.

1. Introduction

With the rapid development of network technology, the problem of cybercrime is becoming increasingly serious. It crosses geographical restrictions and takes various forms, ranging from information theft to financial fraud, seriously threatening the security of individuals, enterprises and the State, and causing economic losses that continue to climb, which has become a global problem [1].

In this situation, it is extremely urgent to explore the global distribution pattern of cybercrime and its prevention and control strategies. Defining the risk characteristics of crimes in different regions will help countries to take reasonable precautions; evaluating the effectiveness of existing policies can promote policy optimisation and enhance the level of global cybersecurity protection.

In this paper, we comprehensively use a variety of research methods to analyse the global pattern of cybercrime and the effectiveness of policies, in the hope of providing useful references for cybercrime governance and helping to build a safe network environment.

2. Global Patterns of Cybercrime

2.1 Patterns of distribution of crime risk

To explore patterns in the distribution of crime risk on a global scale, we cite data from a previous survey effort by Miranda Bruced and others. The respondents were leading experts in cybercrime intelligence or investigation from around the world, and the survey asked participants to consider five broad categories of cybercrime:

(1) Technology Products/Services (2) Assault and Ransom (3) Data/Identity Theft (4) Fraud (5) Cashing/Money Laundering

Experts nominated the countries/territories they considered to be the most important sources of each type of cybercrime, and then ranked each nominated country/territory on the basis of the impact, expertise and technical skills of the perpetrators. The result of the survey is the World Cybercrime Index (WCI), which covers a total of 198 countries and territories, but only the top 97 countries/territories have a specific WCI score. Some of the data from the survey results are shown in Table 1:

Table.1. Selected World Cybercrime Indices

rank	country	WCI Score
1	Russia	58.39
2	Ukraine	36.44
3	China	27.86
6	Romania	14.83
25	Canada	1.34
97	Egypt	0.08

Specifically, the WCI is a measure of the location of cybercrime centres [2]. Note that the WCI does not imply that cybercrime occurs, but only measures the risk of cybercrime occurring in that country/region. The source of the WCI is also not real data, but the empirical judgement of a group of experts.

2.2 K-means++ Clustering Algorithm

In order to classify different countries based on their cybercrime risk, we used the K-means++ clustering algorithm. K-means++ is an improved K-means clustering algorithm that improves clustering by optimising the selection of the initial centre of mass. To classify different countries based on cybercrime risk, we used K-means++ clustering algorithm. K-means++ is an improved K-means clustering algorithm that improves clustering by optimizing the selection of the initial center of mass. K-means++ selects initial centers of mass by a specific probabilistic method that makes the mutual distance between these centers as large as possible. This reduces the likelihood of finding poorer local minima and thus improves the quality of the final clustering results. In addition, K-means++ has a faster convergence rate; due to a better choice of the initial center of mass, K-means++ typically requires fewer iterations than the standard K-means algorithm to reach a steady state, and can therefore converge faster [3,4].

We have categorized different countries into five clusters according to how easy it is to commit cybercrime: very high risk of crime, high risk of crime, medium risk of crime, low risk of crime, and

very low risk of crime. We scored the crime risk on a star scale, with five stars (★★★★★) indicating a very high crime risk and one star (★) indicating a very low crime risk.

The initial cluster centres were obtained by setting up a random set of WCI scores using SPSS as shown in Table 2:

Table.2. Initial clustering center

clustering	1	2	3	4	5
WCI Score	58.39	36.44	27.86	21.28	0.08

We artificially set up 10 iterations, and convergence was achieved at the 5th iteration as there were no changes or only small changes in the clustering centers. The minimum distance between the initial centers was 6.580.

The final clustering centres obtained are shown in Table 3:

Table.3. Final clustering center

clustering	1	2	3	4	5
WCI Score	58.39	36.44	24.72	9.90	0.75

In the end, we achieved a risk classification of 97 countries/regions, and the number of countries/regions corresponding to each risk level is shown in Table 4 below:

Table.4. Results of K-means++ cluster analysis

Risk Level	★★★★★	★★★★	★★★	★★	★
Country number	1	1	3	5	87
Representative countries	Russia	Ukraine	US etc.	UK etc.	Vietnam etc.

According to the table, we can see that the distribution of crime risk levels is very uneven, showing a sharp teardrop shape.

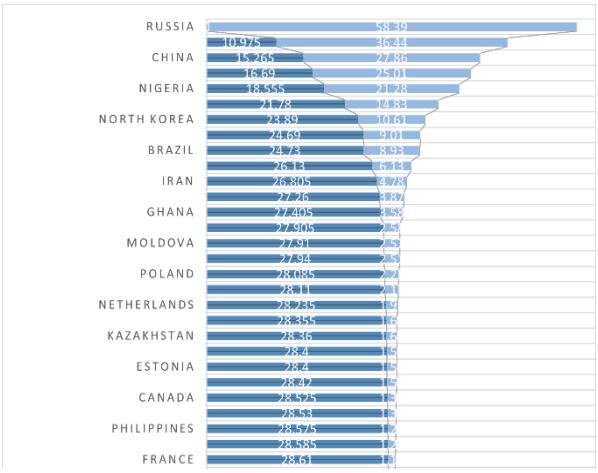

Figure 1: Distribution of crime risk levels

The country in risk level 1 (very high crime risk) is Russia, which is in a separate level. The country in the second risk level (high crime risk) is Ukraine, which is in a class by itself. The countries with medium crime risk are China, the United States and Nigeria. The five countries with low crime risk are Romania, North Korea, the United Kingdom, Brazil and India. The 87 countries with very low

crime risk are represented by Iran. We compare the classification results from the cluster analysis with Figure 1.

When compared, the two distributions are found to be very similar, indicating that our clustering results can represent the actual risk distribution with strong realistic relevance.

3. Policy Effectiveness Model

3.1 Performance Scores and Offence Risk

Cybercrime can virtually ignore the limitations of time and space, a feature that makes it highly mobile and covert. In recent years, cybercrime has been occurring frequently around the world and has shown a trend towards globalisation. To cope with the invasion of networks by lawless elements, the United Nations promulgated the United Nations Convention against Cybercrime on 24 December 2024, which was signed by a total of 193 Member States, demonstrating the determination of the people of the world to jointly combat cybercrime. Countries have formulated policies and regulations to respond to cybercrime with legal procedures, and to measure the effectiveness of the implementation of laws, we firstly plotted Figure 2.

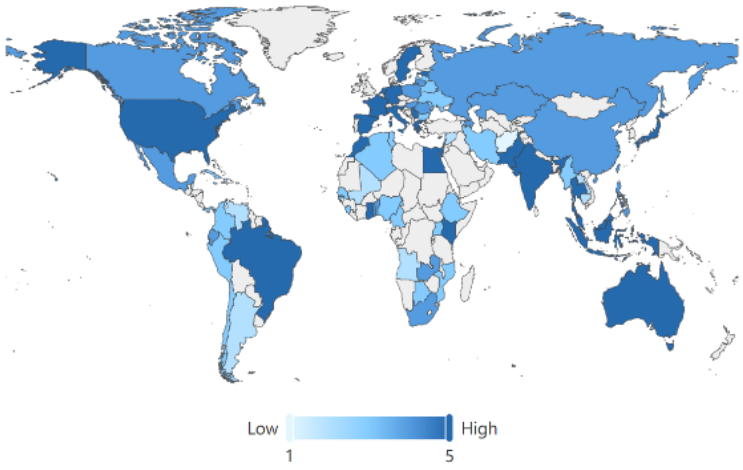

Figure 2: 2024 State Performance Score Map

The “Performance Score” is a ranking of each country based on the International Telecommunication Union's five dimensions: Tier 1 means the country has exemplary cybersecurity; Tier 2 means the country is in an advanced state of cybersecurity; Tier 3 means the country is under construction; Tier 4 means the country is still gradually developing; Tier 5 means the country is in the construction stage of cybersecurity. As mentioned above, the smaller the number, the higher the ranking, which is not conducive to our statistical work, so we constructed an index that can be used to positively express a country's cybersecurity level: Performance Scores (abbreviated as P). P is the result of Tier's positive normalization, and the value ranges from 1 to 5, with the higher scores representing the country's higher cybersecurity construction level.

Comparison with 2024 Global distribution of cybercrime risk making comparisons. It can be seen that both images have roughly the same distribution of color shades, and both are mainly concentrated in economically and scientifically developed regions. These countries have a high degree of cybersecurity, they are the United States, Europe, and Japan. Countries in the Middle East and Africa, which are less technologically advanced, are generally lighter in color and even have gaps in some areas.

As we can see here, this suggests that countries with a high risk of cybercrime have a higher level of cybersecurity building and that the risk of cybercrime and cybersecurity building seem to be reinforcing each other.

3.2 Evaluation of the Effectiveness of Laws

We have discussed global cybercrime in terms of WCI, reporting rate α, and prosecution rate β, and have come up with various conjectures about cybercrime patterns. If we want to go further to obtain the implied mathematical relationships between them in a more scientific way, we need to find a composite parameter to measure the crime indicators. It has to include the impacts generated by the risk of crime (WCI), and (α, β), while, in order to characterise its relationship with active measures, we define the security index S.

$$S_i = \frac{\frac{1}{WCI_i} + \alpha_i + \beta_i}{\sum_{i=1}^{n}(\frac{1}{WCI_i} + \alpha_i + \beta_i)} \tag{1}$$

According to the five political data in the Global Cybersecurity Index GCI given by the International Telecommunication Union (ITU): Legal Measures, Technical Measures, Organization Measures, Capacity Development, and Cooperation Measures, we denote them as L, T, O, A, CO respectively.

At the same time, we retrieved the number of legal documents published by different countries in a certain period as “legal density” (LD) from the Web of Science with the keywords “cybercrime” and “law”(LD), which takes into account the influence of time factor on the security index.

$$LD = \frac{\sum_{i=1}^{n} N_i}{n} (i = 1, ..., n) \tag{2}$$

Where N_i refers to the number of legal documents enacted in years and n denotes the number of years in which laws were enacted.

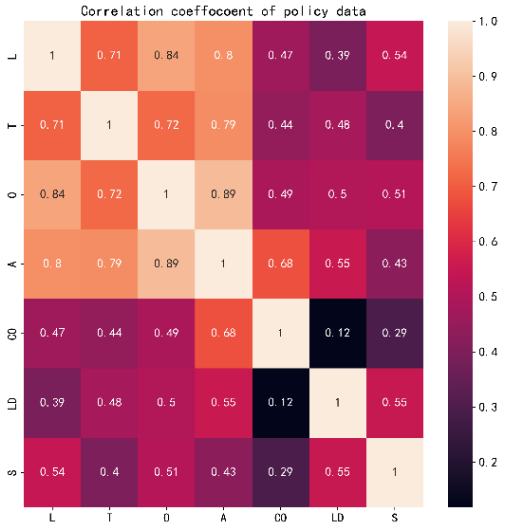

Figure 3: Heat maps of political data and S

Based on the Spearman correlation coefficients of the scores of the six political data and S, we obtained, as shown in the figure 3, that most of the Spearman correlation coefficients are distributed around 0.3 to 0.7, indicating that there is a certain correlation between the political data and S [5].

Therefore, we use political data as an indicator for evaluating the security index to build a multiple linear regression model [6], where C is a constant term:

$$S = \xi_1 L + \xi_2 T + \xi_3 O + \xi_4 A + \xi_5 CO + \xi_6 LD + C \tag{3}$$

To ensure that the selected data are representative, countries were selected from each of the five cybercrime risk classes, containing 32 countries including the Russian Federation, which has the highest WCI score, the United Kingdom, which has a medium score, and Angola, which has the lowest score. A linear regression model was used to obtain the regression coefficients for each indicator, as shown in table 5.

Table.5. Coefficient of S

ξ_1	Legal Measures	2.2623	ξ_2	Technical Measures	0.0188269
ξ_3	Organization Measures	0.561625	ξ_4	Capacity Development	-1.15589
ξ_5	Cooperation Measure	0.598365	ξ_6	Law Density	0.462587

The test value of the F-test at a significance level of 0.05 is 10.29 which is greater than the critical value of 2.49 at that significance level and the test result is shown as True. Meanwhile, the correlation coefficient of the R-test is 0.84, which is close to 1, which shows that the independent variable has a strong linear correlation with the dependent variable.

Through the Coefficient of S, we can see that, $\xi_1, \xi_2, \xi_3, \xi_5, \xi_6$ are all positive, which indicates that improving laws, upgrading science and technology, organization building, enhancing cooperation, and increasing legal literature published per unit of time are all conducive to improving S. ξ_1 is the highest of them, which indicates that, under the condition of equal intensity of improvement, publishing more laws and regulations on cybercrime and cybersecurity is the most effective for the enhancement of S.

3.3 Particle Swarm Optimization (PSO)

The Particle Swarm Optimization (PSO) algorithm originates from the study of bird flock behavior and simulates the process of searching for food during flight [7]. The algorithm searches for an optimal solution through cooperation and competition between individuals (i.e., “particles”) in the flock. Each particle has a position vector and a velocity vector, moves through the solution space to find the optimal solution, and adjusts its direction and speed of movement based on its own experience and the experience of the flock.

The PSO algorithm is relatively insensitive to the selection of the initial solution and the setting of parameters, and even in the case where the distribution of the initial solution is more scattered or the parameter setting is not too reasonable, the algorithm can gradually converge to a better solution through iteration, and it has better robustness. Meanwhile, the PSO algorithm performs well in dealing with continuous optimization problems, can effectively search for the optimal solution in the continuous space, and has a wide range of applications in the fields of function optimization and parameter optimization.

We use this intelligent algorithm to explore the optimal political scenario, i.e., how to most rationally distribute the 6 political data.

We use WCI, reporting rate α, and prosecution rate β as Target Variable and the six political data as Decision Variables. We construct the objective functions of WCI, α and β, respectively, where θ, δ, εare the weights, C_1, C_2, C_3 are constant terms:

$$\min \ln(\mathrm{WCI}) = C_1 + \sum_{i=1}^{6} \theta_i X_i \tag{4}$$

$$\max \alpha = C_2 + \sum_{i=1}^{6} \delta_i X_i \tag{5}$$

$$\max \beta^{\frac{1}{20}} = C_3 + \sum_{i=1}^{6} \varepsilon_i X_i \tag{6}$$

$$\text{s.t.} \begin{cases} x_1, x_2, x_3, x_4, x_5 \le 20 \\ x_6 \le 44 \\ 0 \le x_i, i = 1,...,6 \end{cases} \tag{7}$$

According to our survey, the five political data given by GCI are all between 0 and 20. Meanwhile, we searched a total of 1,147 papers on the Web of Science with the keywords “cybercrime” and “law” in 25 years, so we set the upper limit of the density of laws in each country to 44, giving the above constraints.

At the significance level of 0.05, the test values of F-test are 4.44, 3.13 and 3.88 respectively, which are all greater than the critical value of 2.49 at this significance level, and the test results show True with strong linear correlation. The optimal solutions obtained after optimisation by PSO algorithm are shown in Table 6:

Table.6. Optimal Solution

L	T	O	A	CO	LD
10.2750	20.0000	7.4906	10.2291	20.0000	0.0261

The maximum values corresponding to WCI, α, and β are shown in Table 7:

Table.7. Maximum Values（WCI, α, β）

WCI	α	β
1.100223	0.001007	0.674296

Based on the optimization results, we can confidently advise the lawmakers that the scores of Legal Measures, Technical Measures, Organization Measures, Capacity Development and Cooperation Measures are 10.28, 7.49, 10.23, 20.00, and the Legal Density is 0.0261, and the WCI can get a minimum value of 1.10 and higher values of α and β under this scenario. , 20.00, 7.49, 10.23, 20.00, and the legal density is 0.0261, the WCI can get a minimum value of 1.10 under this scheme, and α and β can get higher values of 0.001 and 0.067. (Two decimals are reserved according to the real situation.)

There is a wide variation in the publication of legal literature in different countries, with the United States at the top of the list publishing 157 documents in 25 years (a legal density of 6.28), while many countries in the African region do not even have a record of legal literature published in 25 years, the

result is a very large extreme difference in the data, and a large dispersion in the data. We use this as a reason to speculate on the reasons for the low legal density.

4. Conclusions

This study explores various aspects of global cybercrime. By using the existing Workload Characteristic Index (WCI) from prior research, the distribution of global cybercrime risk is clarified. The K-means++ clustering algorithm's classification results match the actual situation, aiding countries in formulating prevention strategies. In the research on policy effectiveness, a mutual reinforcement between cybercrime risk and cybersecurity construction is found. A regression model shows legal measures are most effective in boosting the security index. The optimal political scenario from the PSO algorithm offers decision-making references for legislators. Yet, the study has limitations. It didn't consider more indicators for evaluating the security index S, and a more comprehensive view could be adopted for optimizing cybersecurity policies, like factoring in the employment rate. In the future, improving data acquisition methods and the research model can provide more accurate support for global cybersecurity governance.

References

[1] Wang Chao. Empirical Study on Judicial Application Characteristics of the Offence of Helping Information Network Criminal Activities[J/OL]. Journal of Henan Police Academy,1-14[2025-02-15]. https://doi.org/10.16231/j.cnki.jhpc.20250205.004.

[2] Bruce, Miranda, et al. "Mapping the global geography of cybercrime with the World Cybercrime Index." PLoS ONE 19.4(2024).

[3] KONG Weijiang, WANG Yu, CHEN Lu, et al. Design and implementation of massive target rendering system based on K-Means++ and GeoHash[J/OL]. Computer Measurement and Control,1-12[2025-02-15]. http://kns.cnki.net/kcms/detail/11.4762.TP.20241203.1452.006.html.

[4] Liu Chunyu. Distributed energy storage cluster division method based on improved K-means clustering algorithm[J]. Northeast Electric Power Technology,2025,46(01):1-5.

[5] LIU Juan, PENG Jianhong, CAO Wei, et al. Study on the correlation between quality and efficiency index of county cities and comprehensive competitiveness of cities--an empirical analysis based on Spearman's rank correlation coefficient[J]. Quality and Certification,2024,(12):63-65.DOI:10.16691/j.cnki.10-1214/t.2024.12.013.

[6] I International Telecommunication Union. (2024). Global Cybersecurity Index 2024 5th Edition.

[7] ZHANG Guohao, WANG Cailing, WANG Hongwei, et al. Improved particle swarm optimisation algorithm combined with BP neural network model for the prediction of total phosphorus concentration in water transmission spectra[J]. Spectroscopy and Spectral Analysis,2025,45(02):394-402.