J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 9423--9441

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Reinforcement Learning-Based Optimization and Simulation Modeling of International Relations Interaction Strategies

Yanxian Pan[1, ✉]

[1] GUANGXI MINZU UNIVERSITY, Nanning, Guangxi, 530006, China

ABSTRACT

Analyzing the interaction strategies of international relations helps to understand and predict changes in the international landscape, so as to develop and optimize international interaction strategies. Firstly, a single-layer multi-temporal network is modeled for political events, scientific cooperation and international trade in international relations, and added to a multi-layer aggregation network. On this basis, a simulation and analysis method for simulating international relations interaction strategies based on deep learning and multi-intelligence body reinforcement learning methods is proposed. Applying the method of this paper to the arithmetic simulation analysis, it is found that the international relations in the last 10 years have shown the small-world characteristics, and cooperation and conflict coexist. Economic dependence is an influencing factor of conflict between two countries, when the economic and trade links are close, the two countries are less prone to conflict, so the optimization of the international relations interaction strategy should focus on the economic and trade relations.

Keywords: reinforcement learning, international relations interaction, simulation modeling, multilayer aggregation network

1. Introduction

There are three main paradigms in the development of international relations theory, realism, liberalism and constructivism. Before the 1950s, international relations theories heavily emphasized the role of political factors, while with the end of the war, international relations scholars began to pay attention to the role of economic factors in international relations [1-3]. Especially with the end of the Cold War and the development of economic globalization, the role of economic factors in the development of international relations has become increasingly important, and the trend of political economization and economic politicization in the development of international relations has become more and more obvious [4-6]. In reality, the interaction between politics and economy in the

✉ Corresponding author.
E-mail address: panyanxian2024@163.com (Y. Pan).
Received 20 February 2024; Revised 05 April 2024; Accepted 10 November 2024; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-516

development of international relations is real and has been affecting the development of international relations [7-8]. Corresponding to the emergence of this phenomenon, the theory of international relations has also seen a new development, not only the traditional three paradigms of theoretical research to add the economy as an important factor of influence, but also the emergence of the discipline of international political economy specializing in the study of the interaction of political and economic factors in international relations [9-11].

In the early stage of the development of international relations theory, the role of political factors was emphasized, which was closely related to the background of the time. With the end of the war, international relations scholars began to pay attention to the role of economic factors in international relations [12-13]. Especially with the end of the Cold War and the development of economic globalization, the role of economic factors in the development of international relations has become increasingly important, and the development of international relations has shown the trend of political economization and economic politicization. In today's international society, international politics and international economy interact with each other, serve as mutual means, and both sides can realize benign interaction [14-16]. Good political relations can promote the deepening of economic relations and provide a broader platform for interstate cooperation, while the deepening of economic relations is also conducive to deepening communication and understanding, and realizing the identity of different political systems, cultural concepts and ideologies [17-18]. A new world order is quietly taking shape in the interaction between politics and economy.

“At present, China is in the best period of development since modern times, and the world is in a situation of great change not seen in a century, both of which are synchronously intertwined and mutually agitated.” This is a major strategic judgment made by China on today's world [19-20]. Looking at the course of world history since the 20th century, unprecedented changes are taking place in today's world, which are characterized by outstanding instability and uncertainty. As a responsible great power, China strives to promote the development of the international community in the direction of peace and progress [21-23]. China's proposal to build a new type of international relations characterized by mutual respect, fairness and justice, and win-win cooperation is an important diplomatic concept formed by the ruling party of China after in-depth reflection on the proposition of “where are international relations going” at the critical stage of the current profound adjustment of the international pattern and a new round of changes in international rules [24-26]. The construction of a new type of international relations requires good relations among major powers, because historical experience shows that it is the relations among major powers that determine the direction of international relations and the changes in the international situation [27-28].

This study modeled the interaction of international relations based on reinforcement learning techniques. The single-layer multi-temporal phase network is modeled for political events, scientific cooperation and international trade in international relations respectively. To address the one-sidedness of the single-layer multi-temporal phase international relations network in terms of temporal evolution, this paper further constructs a multi-temporal phase aggregation network. The network model adds the information of political events, scientific research cooperation, and trade transactions to the constructed multilayer aggregation network, and based on this, the characterization method of international relations network is proposed. Reinforcement learning-based international relations interaction is transformed into a Markov decision-making process, and the Bellman's optimality equation is introduced to calculate the optimal action values of countries. The characterization of international relations interaction is simulated and modeled and analyzed in arithmetic simulation.

2. Methods of modeling and characterization of interaction strategies in international relations

2.1. Modeling of single-layer multi-temporal networks

2.1.1. Political event network modeling. The political event network established in this paper is a cumulative time network with a quarter as the period of time slice division.

Define an undirected entitled network $G(V,E,W)$ whose nodes are defined as the set of global country/region nodes $v \in V$, and the connecting edges $e \in E$ in the network are defined as the political event links occurring between a pair of nodes in a quarter, and the cumulative value of the Goldstein index of all the political events is used as the weight of the connecting edges between a pair of nodes, which $w \in W$ indicates the positive and negative degree of the relationship between the two nodes.

Starting from the starting date of the selected period, the Goldstein index of all events that appear to be shared between a pair of nodes is accumulated and calculated as shown in Equation (1):

$$w = \begin{cases} \sum \text{GoldsteinScale}, & \sum \text{GoldsteinScale} \geq 0 \\ -\sum \text{GoldsteinScale}, & \sum \text{GoldsteinScale} < 0 \end{cases} \tag{1}$$

The nodes and connecting edges are observed at the event point of the extracted time slice, and if the cumulative value is positive or 0, the corresponding node and connecting edge extractions are added into the positive IRN, and the edge weights are the absolute values of the cumulative values. If it is negative, the corresponding nodes and connected edges are extracted and added into the negative IR network, and the edge weight is the absolute value of the cumulative value.

According to the above network construction method, as of the observation time point of a time slice, a positive IR network and a negative IR network can be output.

2.1.2. Modeling of research collaboration networks. Define an undirected entitled network $G(V,E,W)$ whose nodes are a collection of nodes from countries/regions around the world $v \in V$, and the connecting edges $e \in E$ in the network are defined as the research cooperation behaviors generated by a pair of nodes in a quarter, and the number of all the research cooperation papers is used as the weight of the connecting edges between a pair of nodes, which $w \in W$ indicates the intensity of the research cooperation between two nodes.

Traverse all preprocessed WOS data files stored in the file directory. Set the time-slice observation node sequence $T = \{t_1, t_2, \ldots\ldots t_8\}$, and observe the nodes and connecting edges in the cumulative time network at t_i to extract all the nodes involved in the research cooperation and the research links between the nodes, and output the quarterly research cooperation network by using the intensity of the research cooperation as the weight value of the connecting edges.

2.1.3. International trade network modeling. In this paper, we define an undirected entitlement network $G(V,E,W)$, the nodes in this network are defined as a collection of nodes in countries/regions around the world $v \in V$, and the connecting edges $e \in E$ in the network are defined as the trade transactions generated by a pair of nodes in a quarter, and the total amount of all trade transactions is used as the weight of the connecting edges between a pair of nodes, and this weight $w \in W$ indicates the degree of trade cooperation between two nodes. The value of total trade between a pair of nodes is accumulated from the start date of the selected period.

Traverse all preprocessed UNComtrade data files stored in the file directory. Set the time-slice observation node sequence $T = \{t_1, t_2, \ldots\ldots t_8\}$, and observe the nodes and connecting edges in the cumulative time network at t_i. Extract all nodes and trade links between nodes that have traded, and output the quarterly trade network with the degree of trade cooperation as weights of the connecting edges.

2.2. *Modeling of multilayer multi-temporal aggregation networks*

The single-layer multi-temporal phase international relations network only describes the attributes and dynamic characteristics of international relations in special events in a single thematic dimension, and each layer of the network, as a projection of international relations in a single thematic domain, has the disadvantage of being one-sided in describing the temporal evolution of international relations. Therefore, this paper constructs a multi-temporal phase aggregation network containing multi-dimensional information as the basis of analysis, and the construction method of multi-temporal phase aggregation network is shown in Figure 1.

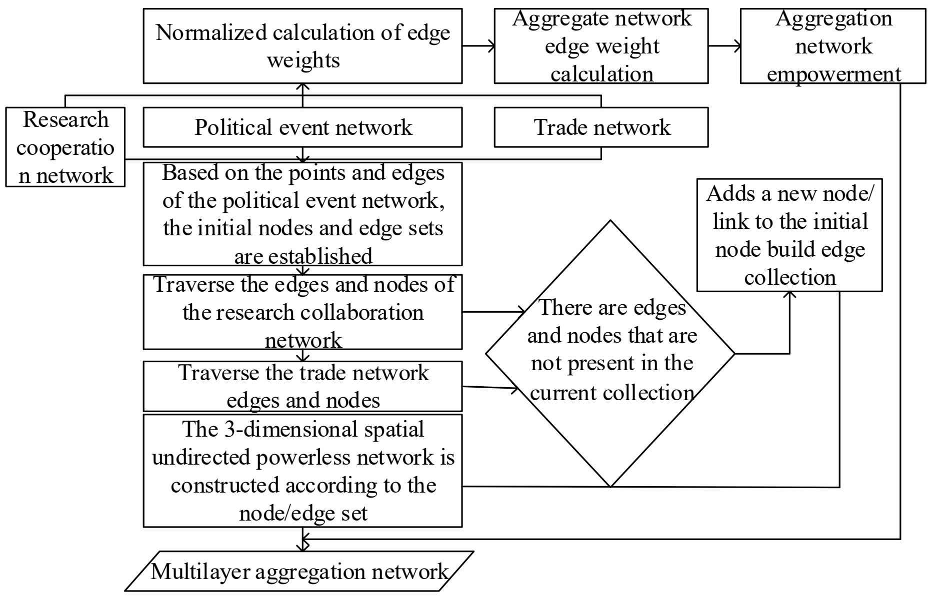


Fig. 1. Construction method of multi-layer aggregation network

2.2.1. Constructing three-dimensional sets of nodes and connected edges. Iterate over all nodes and connecting edges of the 3-layer network and record them to construct a model of the IRN in 3D space. Input the 3 in type single-layer networks output on each time slice. Construct the 3-dimensional network with all nodes and all connected edges using the political event network as the initial network, and the connected edges of this node-constructed network have no weight values. Record all nodes of the political event network as the initial set of nodes, and all its connected edges form the initial set of connected edges. The nodes and connecting edges of the research cooperation network and trade network at the same time stage are traversed, and if the node/connecting edge is not present in the initial set, the node/connecting edge is added to the set. At the end of the traversal the set of nodes and connecting edges of the 3-dimensional network is formed, and a time-stage multilayer aggregation network is constructed on this basis.

2.2.2. Multi-layer aggregation network empowerment. The multilayer aggregation network constructed in the previous step is an undirected and unprivileged network to which the information of three levels, namely, political events, scientific cooperation, and trade transactions, is added.

The nodes and connecting edges of each layer of the network are the projections of the total international network of relations in 3 dimensions, so the relationship between nodes can be represented as a vector u, where $u=[v_1,v_2,v_3]$, v_i denote the components of the relationship in a single dimension, and to reduce the relationships in the three-dimensional network, it is necessary to carry out the operation of the public notice (2):

$$u = v_1 + v_2 + v_3 \tag{2}$$

In this paper, the way to restore the vector values of the principal directions is defined as Equation (3):

$$|u| = |v_1| + |v_2| + |v_3| \tag{3}$$

where $|v_i|$ is the edge weights in a single layer network. $|u|$ is the edge weights in the aggregated network. Its matrix representation is:

$$|u| = [1,1,1][v_1, v_2, v_3]^T \tag{4}$$

Based on Eq. (4) the calculation of edge weights of aggregated networks is started. Firstly, in order to eliminate the bias caused by the order of magnitude difference between different types of data, it is necessary to calculate the normalization of the edge weights of three different types of networks. The normalization is performed according to equation (5):

$$w_{i_n omal} = \frac{w_i - w_{min}}{W_{max} - w_{min}} \times Scale \tag{5}$$

By traversing all the edge weights in the network, the maximum weight $w_{\max}$ and minimum weight $w_{\min}$ are obtained, the original edge weight between a pair of nodes is w_i, and the normalized edge weight is w_{i_nomal}, thus completing the construction of the multilayer aggregation network.

2.3. *Statistical characterization of the international relations network and analysis of inter-layer correlations*

2.3.1. Methods for time-series analysis of statistical characteristics of single-layer IRNs:

1) Statistical Characterization of Single-Layer IRNs

(1) Characteristic path length and diameter. The study of optimal paths is based on the selection of shortest paths. The definition of average path length is as follows: in the same network, there exists a connected shortest distance between any two nodes, and the average value of the shortest distance between all pairs of nodes is the average path length:

$$\mathrm{L} = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \geq j} d_{ij} \tag{6}$$

where d_{ij} is the shortest distance between node i and node j and N is the number of nodes in the network.

(2) Network assortativity and degree correlation. A network is said to be homophobic if nodes of higher degree tend to establish connections with other nodes of higher degree. Conversely, if nodes of higher degree have a tendency to establish connections with other nodes of lower degree, then the network is said to be heterocompatible. In this context, the evaluation of a network is degree-dependent.

2.3.2. Inter-layer correlation analysis of international relations networks:

1) Network Important Node Correlation Analysis. In this paper, we use feature vector centrality to measure the importance of nodes:

$$x_v = \frac{1}{\lambda}\sum_{t\in G} a_v, tx_t \tag{7}$$

The higher the degree of centrality of a node and its partner nodes in a network, the greater will be the centrality of the node's eigenvectors.

Pearson's correlation coefficient is introduced to calculate the inter-layer correlation of nodes:

$$r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2}\sqrt{\sum (Y_i - \bar{Y})^2}} \tag{8}$$

where r denotes the Pearson correlation coefficient and X, Y denotes the elements in the two data sets.

2) Network isomorphism analysis. Define a graph $G_1 = (V_1, E_1)$ and another graph $G_2 = (V_2, E_2)$, there exists a bi-directional mapping relationship between the two graphs, define this bijection as a function π.

Since the size scale and number of nodes of the 3 types of networks are not necessarily the same, a projection of the 3 types of networks on the same time slice is required.

The 3 projected networks of each time slice are analyzed two by two for isomorphism, and it is determined whether there is an isomorphic relationship between these networks. In general, there is a connecting edge between two nodes in one dimension, and there is not necessarily a connection in the other dimension. At this point, this paper defines a pair of data to evaluate the degree of isomorphism between networks:

$$I_i = \frac{k_{ij}}{n_i}, I_j = \frac{k_{ij}}{n_j} \tag{9}$$

where k denotes the number of connected edges that both networks have, n represents the total number of connected edges of a network, and the ratio of the number of shared connected edges to the total number of connected edges of that network I is used to denote the degree of isomorphism between the two networks.

3. Reinforcement Learning Based Simulation Method for International Relations Interaction Strategies

3.1. Enhanced Learning Theory

Reinforcement learning is a machine learning paradigm that consists of two core components: the intelligent body and the environment. During reinforcement learning, an intelligent body acts based on the current state of the environment and receives rewards from it as a way to continuously optimize its behavioral strategies [29].

3.1.1. Markov decision-making process. Unlike supervised learning, reinforcement learning does not require pre-labeled training data, but learns by trial and error, i.e., the intelligence analyzes the environmental feedback during training and uses it to dynamically adjust the execution strategy to maximize the cumulative gain, which can be modeled as a Markov process (MDP).The MDP can be represented as a quintuple $MDP = \{S, A, R, P, \lambda\}$, where R denotes the reward, S denotes the state, A denotes the action space, P and $\lambda \in [0,1)$ denote the state transfer function and decay factor, respectively.

A Markov decision process is used to describe a decision problem faced by a decision maker in a changing environment, which consists of a set of states, actions, and rewards, with the goal of enabling the decision maker to maximize the expected reward at each future step [30]. For moment τ, the intelligent body perceives the current state of the environment $s_t \in S$ and chooses to perform an action based on this state $a_t \in A$. Under the action of the state transfer function $P: S \times A \to S$ of the environment, the environment shifts to a new state s_{t+1}. At the same time, the intelligent body receives a reward $r_t = R(s_t, a_t)$, where $R: S \times A \to R$ is the reward function, which denotes the mapping of the action and state to the reward. In the Markov decision process, the goal is to learn an optimal policy $\pi: S \to A$ that allows the intelligent body to maximize the expected value of the cumulative rewards while interacting with the environment $E(G)$. The cumulative rewards are the sum of the rewards received from the initial state s_0 to the termination state s_T, which can be denoted as $G = \sum_{t=0}^{T} r_t$. Taking into account that the influence of the future rewards on the current decision may decrease over time, a decay factor $\lambda \in [0,1)$ is introduced to Adjusting the weight of future rewards, the cumulative reward with attenuation factor is calculated as follows:

$$G_t = r_t + \lambda r_{t+1} + \lambda^2 r_{t+2} + \ldots + \lambda^{T-t} r_T \tag{10}$$

In the above equation, G_t is the cumulative reward computed starting at time t, r_i is the reward received at time i, and λ^{T-t} is the decay weight corresponding to the future reward r_T. The optimal strategy π^* is determined by maximizing the expected cumulative reward starting from any state $s \in S$:

$$\pi^*(s) = \arg\max_{\pi} \left\{ E\left[\sum_{t=0}^{\infty} \lambda^t \cdot r_t \mid s_0 = s, \pi \right] \right\} \tag{11}$$

where $\pi^*(s)$ is the optimal action choice in state s, E denotes the expectation, $s_0 = s$ indicates that the intelligence starts from state s, and π is the strategy.

3.1.2. Bellman's equation. In order to formally describe the iterative relationship between state value and action value in reinforcement learning, the Bellman optimality equation is introduced. This equation describes the optimal action value function and the optimal policy under a Markov decision process.Bellman's optimality equation solves the optimal control problem in reinforcement learning by stating that the optimal value function can be recursively expressed as a relational equation consisting of the desired reward value and the optimal policy obtained by converting all possible next states and finding the best policy in a deterministic environment [31]. In practice, since the state space can be very large and it is difficult to enumerate all states, it is often necessary to approximate the Q-value function with the help of learning methods and select the optimal policy accordingly. According to Bellman's optimality equation, the optimal value of the action taken by an intelligent in a certain state can be calculated, and an effective task execution strategy can be formulated accordingly to realize the goal of reinforcement learning.Bellman's optimality equation is as follows:

$$Q^*(s,a) = \sum_{s'} P(s,a,s')\left[R(s,a,s') + \gamma \max_{a'} Q^*(s',a')\right] \tag{12}$$

In the above equation, $P(s,a,s')$ is the probability of transferring the action a taken from state s to state s', $R(s,a,s')$ is the reward gained by taking the action in the current state, and γ is the discount factor, which indicates the discounting factor for future gains at the current moment. $Q^*(s,a)$ is the optimal Q value, which can be used to train the neural network, where the loss function is defined as the difference between the optimal Q value $Q(s,a)$ and the value Q predicted by the network $Q(s,a)$ and back propagated through the network. The loss function is defined as follows:

$$\overbrace{E\left[R_{t+1}+\gamma\max_{a'}Q^{*}\left(s',a'\right)\right]}^{\text{Target}}-\overbrace{E\left[\sum_{k=0}^{\infty}\gamma^{k}R_{t+k+1}\right]}^{\text{Predicted}} \tag{13}$$

3.2. *Multi-intelligent Body Reinforcement Learning Theory*

Multi-intelligent Reinforcement Learning (MARL) typically uses a Partially Observable Markov Process (POMDP) as a mathematical model, denoted as $Dec-POMDP(S,T,A,R,\gamma)$. Assume that there are N intelligences, each of which has its own local state, action space, and receives its own reward. Thus, the MDP in a single-intelligent system can be extended to a MAS, where the state space can be decomposed as $S=S_1\times S_2\times\cdots\times S_N$, the action space can be represented as $A=A_1\times A_2\times\cdots\times A_N$, the reward function can be represented as $R=R_1\times R_2\times\cdots\times R_N$, and the transfer function becomes $T=S\times A_1\times\cdots\times A_N$. At each time step t, each intelligent k chooses to perform the corresponding action a_t^k based on the observed state o_t^k, and S and A are the global state and joint action. In this scenario, since the intelligences are not visible to each other, they compensate for the information that cannot be directly observed through an information sharing mechanism. With the continuous advancement of Deep Reinforcement Learning (DRL) techniques, researchers have begun to explore the application of DRL to multi-intelligent systems and have proposed a variety of Multi-Intelligent Deep Reinforcement Learning (MADRL) algorithms. These algorithms have been validated in several classical MAS experimental environments, proving their effectiveness.

Generally speaking, the training process of multi-intelligence reinforcement learning can be divided into centralized training and distributed training according to whether or not the intelligences need information from other intelligences when updating their own strategies: correspondingly, it is divided into centralized execution and distributed execution according to whether or not external information is needed in the execution phase. Combining the training and execution phases, multi-intelligence reinforcement learning can be divided into three paradigms: distributed training distributed execution (DTDE), centralized training centralized execution (CTCE), and centralized training distributed execution (CTDE).

In the DTDE framework, each intelligent body independently performs policy updating and policy realization using only its own local information, which does not involve information exchange.IQL is a representative algorithm built on the DTDE framework, which has the advantage of focusing on only a single intelligent body, thus providing good scalability. However, since the intelligences ignore the information or states of other intelligences, they are often in a non-stationary process, which hinders the learning of optimal policies. To overcome the inefficiency of distributed learning, optimistic optimization and lagging optimization can mitigate this problem to some extent by giving a larger update weight to the current better Q-value. Meanwhile, when using DQN for policy updating, directly utilizing the data in the experience playback pool can exacerbate the non-stationarity problem since the policies of other intelligences are also being updated. In this regard, importance sampling can be used to mitigate this effect:

$$\mathrm{L}\left(\theta_i\right)=\frac{\pi_{-i}^{t_c}\left(a^{-i}\mid o^{-i}\right)}{\pi_{-i}^{t_i}\left(a^{-i}\mid o^{-i}\right)}\left[\left(y_i^0-Q_i\left(o^i,a^i;\theta_i\right)\right)^2\right] \tag{14}$$

The θ_i in the above equation is the Q network parameter of the intelligent body i, t_c is the current moment, t_i is the collection time of the sample, and y_i^Q is the temporal difference target, and the computation is similar to that presented in the section on single intelligent bodies. Meanwhile, the use of techniques such as recurrent neural networks can also alleviate the above problems to some extent. Compared with distributed training, the centralized training framework allows the intelligences to exchange information during the training process and even get all the information of other intelligences.

Under the CTCE framework, the intelligences learn a centralized joint strategy. In the CTCE framework, a multi-intelligent system can be trained using any of the single-intelligent reinforcement learning algorithms. In this class of frameworks, the complexity of the algorithms exhibits dimensionality explosion as the dimensionality of states and actions grows, a problem that can generally be solved by policy decomposition or value decomposition (e.g., $Q_{tot}(o^1,\cdots,o^n,a^1,\cdots,a^n)=\sum_i Q_i(o^i,a^i)$). Although decomposition class methods can alleviate the dimension explosion problem, CTCE cannot evaluate the interactions between each intelligence, when the confidence allocation problem will bring more serious impact on the learning efficiency. To cope with these challenges. In this paper, we adopt a skill learning approach, or grouping the intelligences first, and then train them using a local CTDE strategy.

4. Arithmetic simulation

4.1. Characterization of international relations interactions

4.1.1. Scale-free characterization. The scale-free nature of the network is manifested in the fact that a few nodes possess a high degree or strength, and a few nodes have strong control and influence in the network. In aviation networks, the paralysis of a few key nodes may pose a serious threat to the entire aviation network. The development of conflict in a few key nodes in an international relations network has a significant impact on global international relations. The degree distribution of nodes in scale-free networks has a power law distribution.

The intensity distribution of network nodes is analyzed by fitting a power law distribution. First, the node intensity distribution of the international aviation network, the international relations conflict network, and the international relations cooperation network are statistically analyzed using OAG data and GDELT data for the 10-year period from 2014-2023, respectively. The node strength of the international aviation network reflects the sum of flight traffic between the node and other nodes in 10 years. The node intensity of international relations network refers to the sum of the impact level of cooperation or conflict events between nodes and other nodes. Figure 2 shows the probability distribution of node intensity for the three networks.

All three networks in Fig. 2 exhibit scale-free characteristics. The power-law distribution is analyzed based on the characteristics of the probability density function of the node intensities, Eq. $Y=kX^{-a}$, where Y is a probability distribution, X is a random variable referring to the value of the node intensities, k is a constant, and a is a power-law exponent. The power-law distribution is a linear function in the double logarithmic coordinate system. Commonly used power law distribution fitting models include Zipf distribution, Pareto distribution, Yule-Simon distribution and so on. Power law distribution fitting of node intensities is performed using a modified Zipf distribution model, Zipf-Mandelbrot.Zipf-Mandelbrot introduces a translation parameter, ρ, to characterize the saturation effect that results in sampling of the data.The Zipf-Mandelbrot distribution uses complementary cumulative probability distribution functions to fit the data. The advantage of the complementary cumulative probability distribution is that it better characterizes the distribution of the tails, Eq. $A_c(M>X)=k(X+\rho)^{-a}$, where $A_c(M>X)$ is the complementary cumulative probability distribution function of X denoting the probability that the random variable M is outside the agreed-upon range X and ρ is the panning parameter.

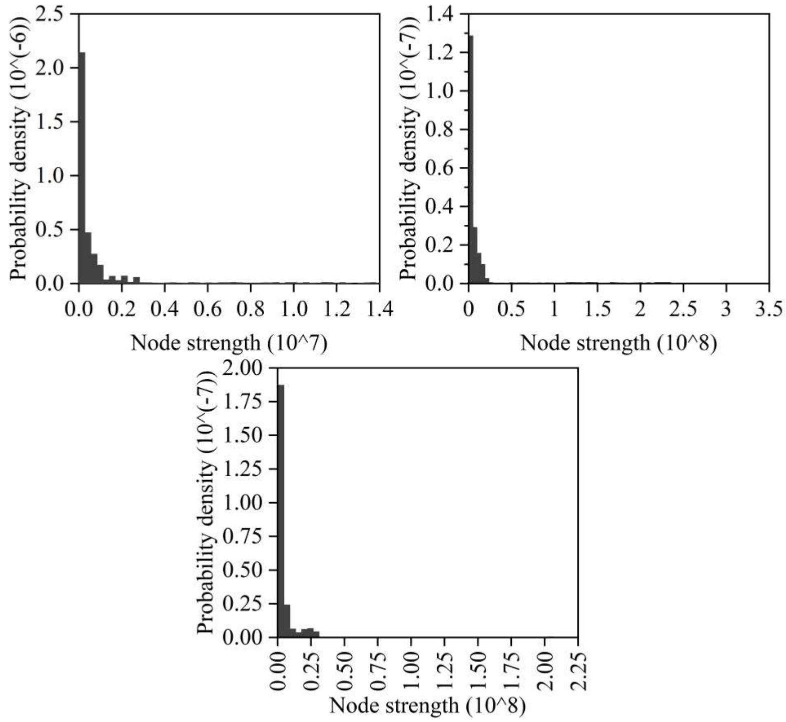


Fig. 2. Probability density distribution diagram of network node strength

4.1.2. Small world characterization. Statistical and evolutionary analyses of the overall measurements of the aviation network and the international relations network over a 10-year period using OAG data and GDELT data for the 10-year period from 2014 to 2023 are conducted to compare the overall nature of the networks of different topics and to reveal the similarities and differences in the global characteristics and evolutionary laws of the networks. The overall measure of the international aviation network is calculated, as shown in Table 1. N in the table is the number of nodes, E is the number of connected edges of the network, G is the average degree, D is the graph density, C is the average clustering coefficient, R is the average shortest path length, and DA is the degree assortment coefficient.

As can be seen in Table 1, the number of nodes in the aviation network remains stable, and the number of connected edges of the network changes significantly. The number of contiguous edges has continued to increase since 2014, with a large decrease by 2022, and regresses to the 2018 level by 2023. This indicates that the number of routes in the aviation network has continued to increase, but some routes between countries or regions have been forced to be canceled under the impact of the global New Crown epidemic and multiple localized conflicts in 2023. The graph density, average clustering coefficient, and average route length of the network also show an increasing and then decreasing trend. The graph density of the aviation network stays around 0.13, which is at a low level, indicating that the global aviation links have become more sparse. The average clustering coefficient is 0.645, indicating that the overall aggregation level of the aviation network is high. The average path length of the network is 2.23, reflecting that the aviation network has better overall connectivity. Taken together, the aviation network has a lower graph density, a higher average clustering coefficient, and a shorter average shortest path length, exhibiting small-world characteristics.

Table 1. 10-year evolution of international aviation network

Year	N	E	G	D	C	R	D_A
2014	234	2731	24.39	0.1	0.62	2.24	-0.07
2015	234	2759	24.61	0.12	0.69	2.23	-0.03
2016	234	2787	24.86	0.1	0.62	2.26	-0.08
2017	234	2878	25.69	0.11	0.62	2.25	-0.07
2018	235	2941	26.19	0.14	0.66	2.23	-0.05
2019	235	3038	26.97	0.15	0.63	2.19	-0.05
2020	235	3031	27.02	0.15	0.65	2.22	-0.05
2021	235	3228	28.66	0.15	0.67	2.18	-0.05
2022	233	2864	25.72	0.13	0.64	2.24	-0.07
2023	233	2970	26.66	0.12	0.65	2.19	-0.07

The 10-year measurement evolution of the international conflict network is shown in Table 2.The change in the number of connecting edges is similar to that of the air traffic network, which has gone through an upward and then a downward phase. The number of connected edges of the conflict network reached an extreme value in 2017, and then maintained a decreasing trend. The graph density, average clustering coefficient and average shortest path length of the conflict network all show more obvious small-world characteristics than the air network. The graph density ranges from 0.25 to 0.4, indicating denser conflict links between countries. The average clustering coefficient is 0.726, showing a tendency of clustering of international conflict ties. The average shortest path is about 1.65, indicating that countries or regions are more prone to conflict. These characteristics indicate that the nodes in the conflict network of international relations are more closely linked to each other than in the aviation network and that the small-world characteristics are more pronounced.

Table 2. 10-year evolution of international conflict network

Year	N	E	G	D	C	R	D_A
2014	221	7492	68.41	0.27	0.69	1.72	-0.23
2015	217	8218	75.44	0.38	0.71	1.66	-0.22
2016	216	9172	83.70	0.36	0.75	1.66	-0.23
2017	220	9449	86.30	0.40	0.74	1.63	-0.25
2018	214	9218	84.97	0.38	0.75	1.62	-0.20
2019	215	8737	80.50	0.35	0.74	1.65	-0.21
2020	220	8626	79.48	0.37	0.71	1.66	-0.20
2021	217	7861	72.13	0.31	0.74	1.69	-0.25
2022	209	7754	71.18	0.33	0.70	1.71	-0.24
2023	218	7322	67.18	0.32	0.73	1.70	-0.18

4.1.3. Analysis of the overall relevance of the interactive network of international relations. The similarity of network structure can be utilized to evaluate the degree of network association. The node and overall structural similarity of international aviation networks and international relationship networks are analyzed to identify the time nodes where the similarity changes significantly, which lays the foundation for the local network association analysis.Jaccard similarity (J) is a measure of the degree of similarity of a set. The Jaccard similarity formula for the 2 sets A and B is $J(A, B)=(|A\cap B|)/(|A\cup B|)$, which can be used to represent the overlap between countries or regions involved in aviation interactions (A) and countries or regions involved in international relations interactions (B) in the same time scale. The range of values is [0, 1], and the closer to 1 indicates that the 2 sets are more similar.

Figure 3 shows the network node similarity evolution. The overlap between the international cooperation network and the international conflict network is about 97%, indicating that cooperation accompanies the emergence of conflict and that the international relations game between countries or regions is complex. The node overlap between the international aviation network and the international relations network is about 77%, indicating that the nodes of the countries involved in the complex interaction between aviation cooperation and multiple events have a high similarity, although they are not identical.

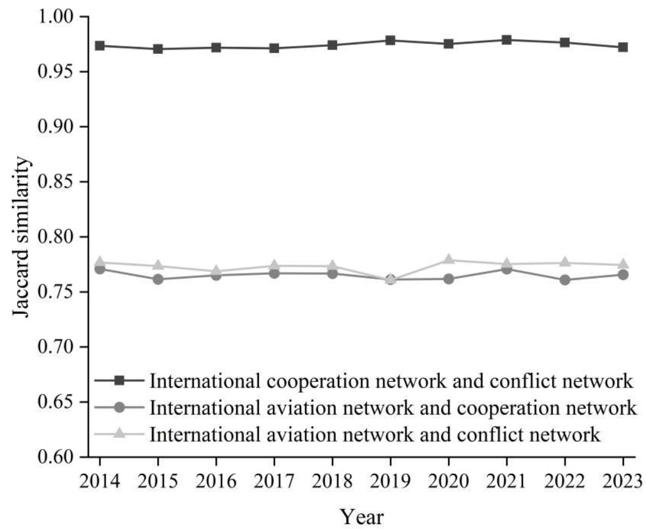


Fig. 3. Evolution map of network node similarity

The overlapping nodes in the 2 different thematic networks were extracted to analyze the evolution pattern of the correlation coefficients of the overlapping node strengths of the aviation network and the international relations network for the 10-year period from 2014-2023, as shown in Figure 4.

As can be seen in Figure 4, the correlation coefficient of the node strength of the aviation network with the international relations network is about 0.6, which shows a weak correlation. The node strength correlation between the aviation network and the international cooperation network is slightly higher than its correlation with the international conflict network. The higher correlation coefficient between the international cooperation network and the conflict network indicates that international cooperation and conflict exist along with each other, which is in line with the international interaction characteristics of the variable-sum game. The correlation coefficient between the aviation network and the international relations network is relatively stable until 2022, but decreases more by 2022. This indicates that the global epidemic did not have a large impact on the node similarity of the two networks before, but the Russian-Ukrainian conflict caused the aviation network and the international relations network to evolve more heterogeneously.

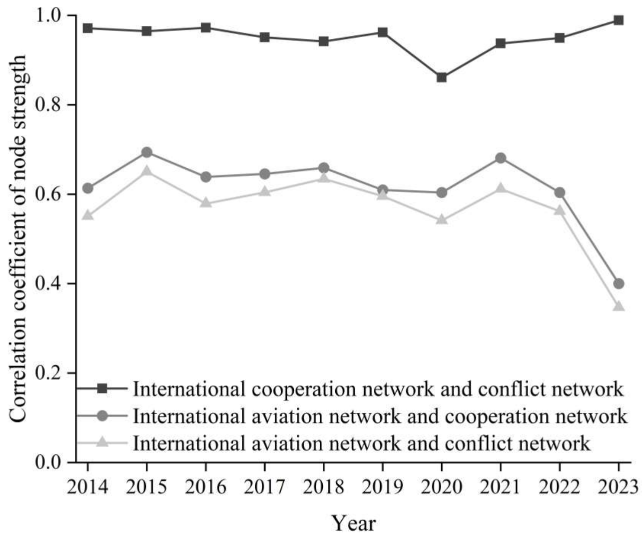


Fig. 4. Evolution map of network node strength

4.2. Simulation modeling analysis of international relations interactions

4.2.1. Simulation execution process. After the international relations interaction strategy modeling, the simulation is executed as follows:

1) Initialize the distribution of national Agent strategies of both populations.
2) In each round of the game, randomly select two Agents of both sides to play the game two by two.
3) Calculate the gains of both parties after one round of the game.
4) Update the strategies and adjust the countermeasures for the next round of the game.
5) Repeat this process many times until the system reaches the maximum number of iterations or reaches a stable state.
6) Output the game results under the influence of different game strategies and analyze them comparatively.

4.2.2. Simulation data. The parameters of the model are set as follows: gene chain length=19, population size of both sides=28, gene crossover probability=0.8, and gene variation probability=0.08. In this paper's research, the international strategic gains of both sides are selected as the gain indicator at the border demarcation level, such as the degree of conformity of the strategy of country A to that of country B to the realization of its overall international strategy, and if the degree of conformity is higher, then the corresponding gain is greater. At the economic and trade level, the domestic economic gains of both countries are taken as the gain indicator, i.e., the proportion of total trade and investment of both countries in the overall trade and investment, if the proportion is larger, the gain of choosing cooperation strategy is larger.

4.2.3. Simulation experiments and analysis:

1) Effect of initial strategy distribution. The effect of initial state change on the evolutionary outcome under the basic data setup is shown in Fig. 5. Let the initial cooperation probabilities be (0.1, 0.1), (0.3, 0.7), (0.5, 0.5), (0.7, 0.3), and (0.9, 0.9). When the initial cooperation probability is (0.1, 0.1), both parties are under tension and both have a 0.9 likelihood of intense confrontational behavior. When the initial cooperation probability is (0.3, 0.7), country B has a stronger initial willingness to adopt a cooperative strategy relative to country A. The opposite is true when the initial cooperation probability is (0.7, 0.3). An initial cooperation probability of (0.5, 0.5) implies that both parties are more

ambiguous in their attitudes. When the initial cooperation probability is (0.9, 0.9), it means that both sides are more restrained and have a stronger initial willingness to adopt cooperative strategies.

As can be seen from Figure 5, regardless of the distribution of the initial strategies of the two countries, the final outcome evolves towards adopting intense conflict. Despite some differences in the evolutionary paths, none of them affects the evolutionary steady state equilibrium. This suggests that the evolutionary outcome is mainly influenced by the revenue parameter, independent of the initial strategy distribution. It also shows that the evolutionary steady-state equilibrium is globally stable.

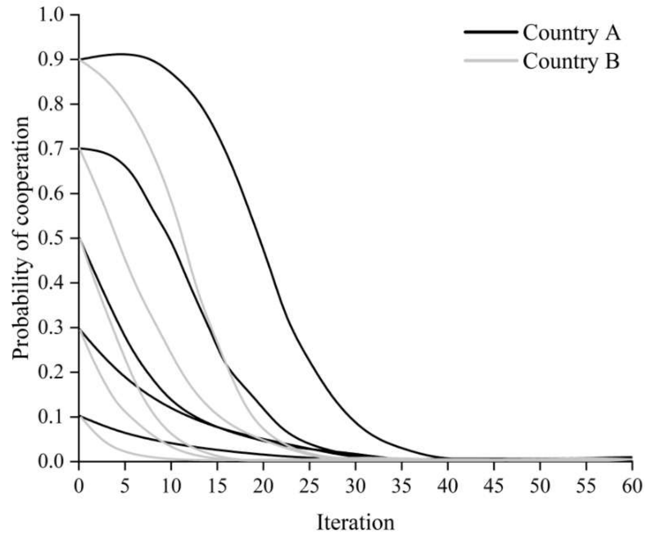


Fig. 5. The evolutionary results of the initial strategy distribution

2) Influence of the degree of economic dependence. From the above analysis, the evolutionary steady state equilibrium is not affected by the initial value, implying that the evolutionary process is affected by the parameters in the basic parameter settings. Firstly, we analyze the effect of the change of economic dependence weights by modifying the values of economic dependence weights kA and kB under the basic parameter settings. When the economic dependence weights of both countries are (0.9, 0.9), (0.9, 0.7), (0.7, 0.5) and (0.1, 0.1), the evolution results are shown in Figure 6.

When both countries have high economic dependence, the demarcation attitudes of both countries evolve towards cooperation. At this time, both countries can obtain higher economic gains while demarcating the border and cooperating, so they both prefer to choose the cooperative strategy.

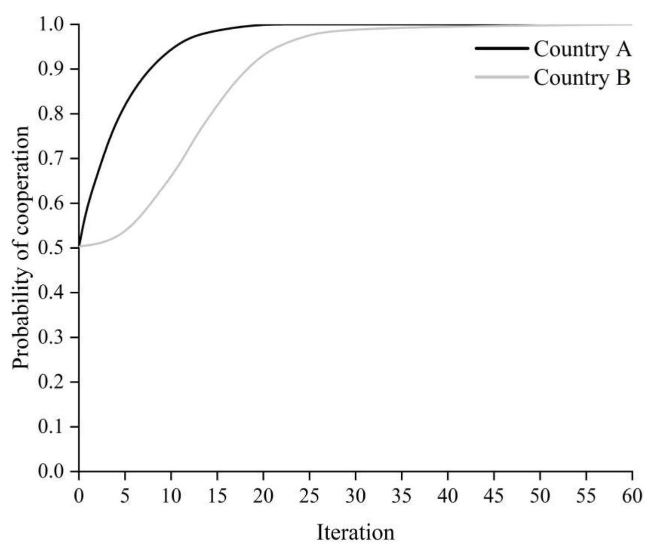
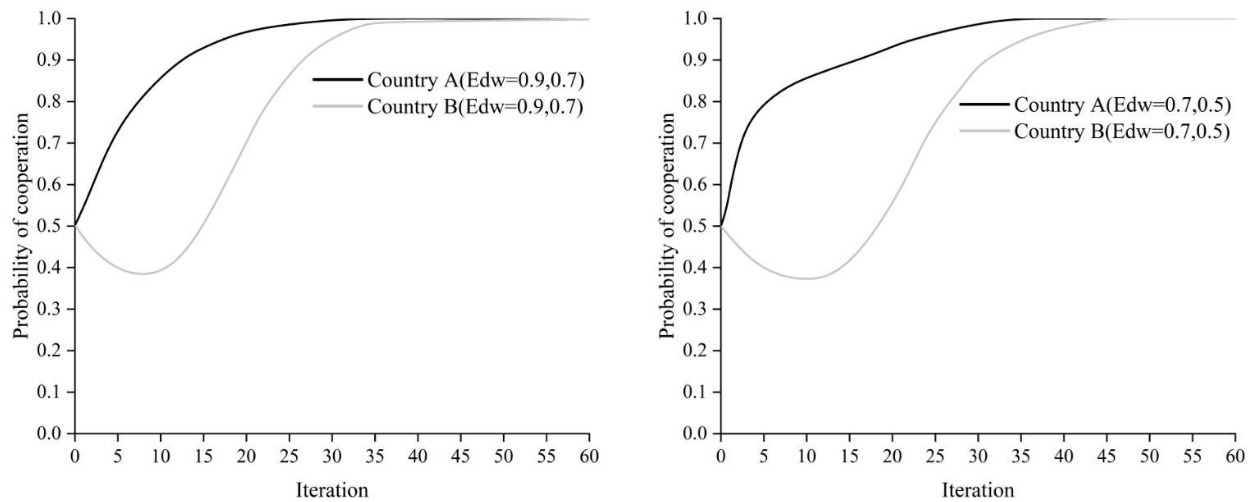


Fig. 6. The economic dependence evolution results on the weight of (0.9, 0.9)

The evolution results when the economic dependence weights are (0.9, 0.7) and (0.7, 0.5) are shown in Fig. 7 (a) and (b), when the economic dependence index of one country or two countries is reduced to a certain extent, the two sides of the game, especially the side with the reduced dependence index, no longer choose the cooperative strategy firmly as in Fig. 7, and even start to try the conflict strategy, but the final evolution results in the two sides still adopting the cooperative strategy.



(a) Economic dependence weight=(0.9,0.7) (b) Economic dependence weight=(0.7,0.5)

Fig. 7. The economic dependence evolution results on the weight of (0.9,0.7), (0.7,0.5)

The results of the evolution of economic dependence with weights of (0.1, 0.1) are shown in Figure 8, where both countries' demarcation attitudes evolve towards conflict when both countries have low economic dependence. At this time, the gains at the economic and trade level are less binding on both countries, and both countries prefer to adopt the strategy of conflict.

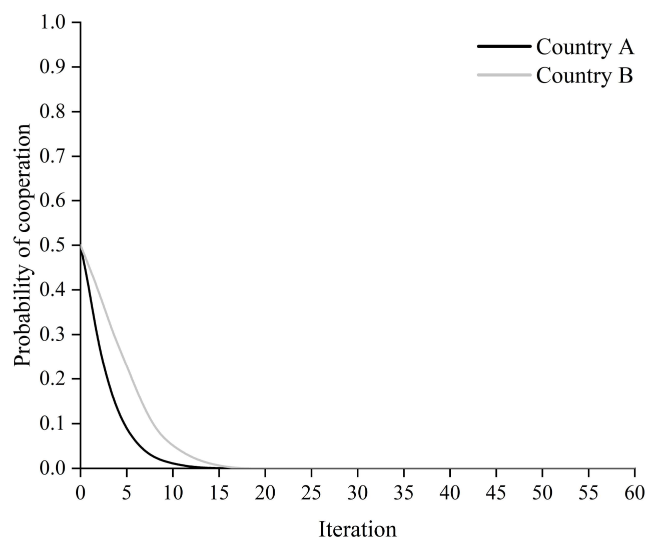


Fig. 8. The economic dependence evolution results on the weight of (0.1, 0.1)

3) Impact of Economic Gain Parameters. In terms of the current relationship between most countries around the globe, the aspect where the cooperation between the two sides can bring significant gains is still in the field of economic and trade. Therefore, in this subsection, the impact of economic and

trade gains on the evolutionary process will be analyzed by modifying the gain parameters of economic and trade cooperation of the game countries. Under the basic parameter settings, increase the value of gains from economic and trade cooperation between the two sides, and set the incremental gain Δ. When Δ takes the values of 3, 5, 10, and 15, respectively, the obtained evolutionary trajectory diagrams of the two sides are shown in Figs. 9 to 12.

As can be seen from the figure, when the increment of economic gains is small, the demarcation attitudes of both countries evolve in the direction of conflict. At this time, the economic gains cause no significant change in the strategy choices of both sides of the game.

As the economic gains from cooperation between the two sides increase, country A begins to seek cooperation, but is forced to adopt non-cooperative behavior in the end due to the other side's non-cooperation. As can be seen from the evolutionary trajectory in the figure, although the final evolutionary result is still both parties adopt conflicting strategies, it takes a longer evolutionary time to finally reach a stable state.

When Δ=10, country A adopts cooperative strategy from the beginning with a higher probability, and country B still tends to non-cooperative strategy at the beginning, but due to the attraction of economic gains and the other party's insistence on cooperative strategy, it eventually moves towards cooperation as well.

When Δ = 15, then the incremental economic gain is larger, and both sides of the game tend to adopt the cooperative strategy. Compared with Fig. 11, at this time both sides of the game are more determined to choose the cooperative strategy, and quickly move towards full cooperation.

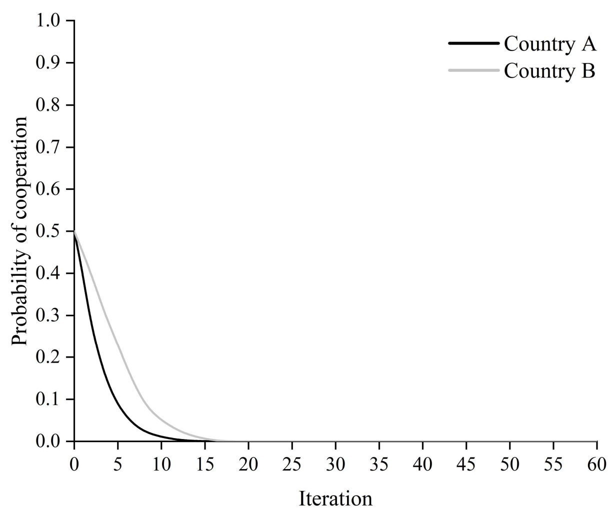


Fig. 9. The evolution results when Δ=3

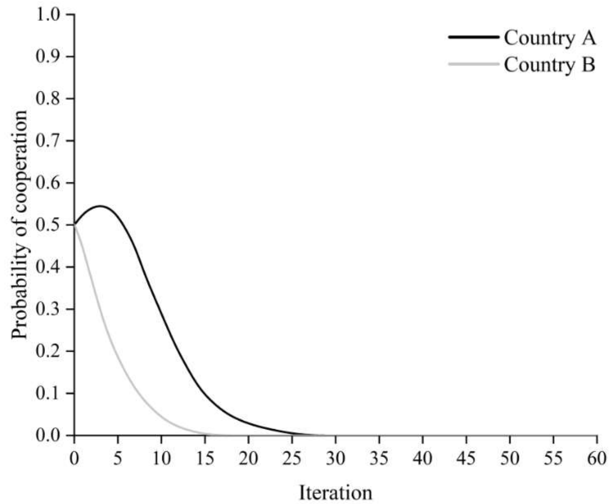


Fig. 10. The evolution results when Δ=5

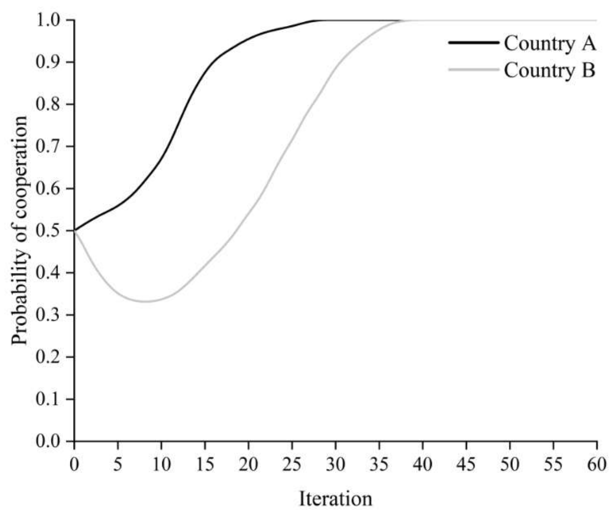


Fig. 11. The evolution results when Δ=10

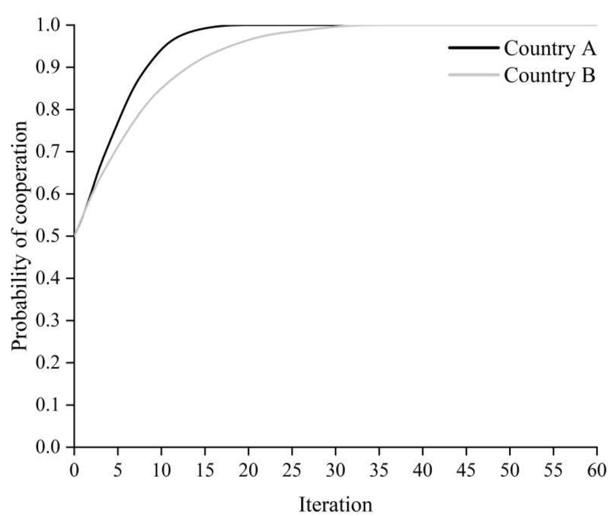


Fig. 12. The evolution results when Δ=15

5. Conclusion

In this paper, a simulation modeling analysis of international relations interaction strategies is carried out based on the reinforcement learning method.

The three kinds of networks constructed in this paper exhibit scale-free properties. The average clustering coefficient of the aviation network is 0.645, and the average path length is 2.23. The graph density is between 0.25 and 0.4, and the average clustering coefficient of the conflict network is 0.726, and the average shortest path is about 1.65. It shows that in recent 10 years, the international relations in general have shown the characteristics of a small world, and the international cooperation exists along with the conflict. Simulation analysis finds that when both countries have high economic dependence, the demarcation attitudes of both countries evolve in the direction of cooperation. Whereas the gains in the economic and trade dimensions are less binding for both countries, both countries are more inclined to adopt the strategy of conflict.

This result reveals that in order to optimize the interaction strategy of international relations, the expansion of economic ties and the rational control of differences should be considered.

References

[1] Rosenau, J. N. (2018). Turbulence in world politics: A theory of change and continuity. *Princeton University Press*.

[2] O'brien, R., & Williams, M. (2024). Global political economy: Evolution and dynamics. *Bloomsbury Publishing*.

[3] Jervis, R. (2017). Perception and misperception in international politics: New edition. *Princeton University Press*.

[4] Kobrin, S. J. (2022). Managing political risk assessment: Strategic response to environmental change. *Univ of California Press*.

[5] Eckhard, S., & Ege, J. (2018). International bureaucracies and their influence on policy-making: A review of empirical evidence. *Governance by International Public Administrations*, 12-30.

[6] Uslaner, E. M. (Ed.). (2018). The Oxford handbook of social and political trust. *Oxford University Press*.

[7] Brecher, M., & Wilkenfeld, J. (2022). A study of crisis. *University of Michigan Press*.

[8] Mabon, S. (2018). Saudi Arabia and Iran: power and rivalry in the Middle East. *Bloomsbury Publishing*.

[9] Starr, H., & Starr, H. (2021). Democratic dominoes: Diffusion approaches to the spread of democracy in the international system. *Harvey Starr: Pioneer in the Study of Conflict Processes and International Relations*, 123-145.

[10] Gorwa, R. (2019). What is platform governance?. *Information, communication & society*, 22(6), 854-871.

[11] Klotz, A. (2018). Norms in international relations: The struggle against apartheid. *Cornell University Press*.

[12] Haas, E. B. (2024). Beyond the nation-state: Functionalism and international organization. *ECPR Press*.

[13] Joseph, G. M., LeGrand, C. C., & Salvatore, R. D. (Eds.). (2020). Close encounters of empire: writing the cultural history of US-Latin American relations. *Duke University Press*.

[14] Dunne, T., & Reus-Smit, C. (Eds.). (2017). The globalization of international society. *Oxford University Press*.

[15] Iklé, F. C. (2024). How nations negotiate. *Plunkett Lake Press*.

[16] Szkudlarek, B., Osland, J. S., Nardon, L., & Zander, L. (2020). Communication and culture in international business–Moving the field forward. *Journal of World Business*, 55(6), 101126.

[17] Acharya, A., & Buzan, B. (2019). The making of global international relations. *Cambridge University Press*.

[18] Halperin, S., & Heath, O. (2020). Political research: methods and practical skills. *Oxford University Press, USA*.

[19] Holmes, M. (2018). Face-to-face diplomacy: Social neuroscience and international relations. *Cambridge University Press*.

[20] Dunne, T. (2024). International relations theories: Discipline and diversity. *Oxford University Press*.

[21] Legro, J. W. (2019). Rethinking the world: Great power strategies and international order. *Cornell University Press*.

[22] Rosenberg, J. (2024). The empire of civil society: a critique of the realist theory of international relations. *Verso Books*.

[23] Witt, M. A. (2019). De-globalization: Theories, predictions, and opportunities for international business research. *Journal of International Business Studies*, 50(7), 1053-1077.

[24] Young, O. R. (2018). International governance: Protecting the environment in a stateless society. *Cornell University Press*.

[25] Egloff, F. J. (2022). Semi-State Actors in Cybersecurity. Oxford University Press.

[26] Sørensen, G., Møller, J., & Jackson, R. H. (2022). Introduction to international relations: theories and approaches. *Oxford university press*.

[27] White, B. (2017). Understanding European foreign policy. Bloomsbury Publishing.

[28] Baylis, J., Smith, S., & Owens, P. (Eds.). (2020). The globalization of world politics: An introduction to international relations. *Oxford university press*, USA.

[29] Beaudoin Marc-Antoine & Boulet Benoit. (2022). Improving gearshift controllers for electric vehicles with reinforcement learning. Mechanism and Machine Theory.

[30] Vivek S. Borkar. (2025). Some big issues with small noise limits in Markov decision processes. Systems & Control Letters105972-105972.

[31] YuGuo & ZhenyiDai. (2024). Hamilton–Jacobi–Bellman equation based on fractional random impulses system. Optimal Control Applications and Methods(4),1716-1735.