

Research on Weighted Directed Dynamic Multiplexing Network of World Grain Trade Based on Improved MLP Framework

Shanyan Zhu, Shicai Gong*

School of Science, Zhejiang University of Science and Technology, Hangzhou, China
Email: Zhushanyan1116@foxmail.com, *scgong@zafu.edu.cn

How to cite this paper: Zhu, S.Y. and Gong, S.C. (2023) Research on Weighted Directed Dynamic Multiplexing Network of World Grain Trade Based on Improved MLP Framework. *Journal of Computer and Communications*, 11, 191-207.
<https://doi.org/10.4236/jcc.2023.117012>

Received: May 25, 2023

Accepted: July 28, 2023

Published: July 31, 2023

Copyright © 2023 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

As the main food source for humans, the global movement of the three major grains significantly impacts human survival and development. To investigate the evolution of the world cereal trade network and its development trend, a weighted directed dynamic multiplexed network was established using historical data on cereal trade, cereal import dependency ratio, and arable land per capita. Inspired by the MLP framework, we redefined the weight determination method for computing layer weights and edge weights of the target layer, modified the CN, RA, AA, and PA indicators, and proposed the node similarity indicator for weighted directed networks. The AUC metric, which measures the accuracy of the algorithm, has also been improved in order to finally obtain the link prediction results for the grain trading network. The prediction results were processed, such as web-based presentation and community partition. It was found that the number of generalized trade agreements does not have a decisive impact on inter-country cereal trade. The former large grain exporters continue to play an important role in this trade network. In the future, the world trade in cereals will develop in the direction of more frequent intercontinental trade and gradually weaken the intracontinental cereal trade.

Keywords

MLP Framework, Food Security, Dynamic Multiplexed Networks, Trade Network Link Forecasting

1. Introduction

With deepening global economic globalization and population growth, the world food market is highly integrated, and the food supply always affects the national

economy. As the staple food of the majority of the world's population, grain is necessary for the survival and development of a country's people, as well as an important strategic material in the face of crises [1] [2]. However, the cultivation of specific grains is often concentrated in some agricultural powers. In the context of the relatively concentrated supply of agricultural products, any small trade behavior of these grain-producing or exporting powers will impact the world grain trade pattern, like the flapping wings of a butterfly [3] [4] [5]. For the sake of the food security of each country, it is a very meaningful topic to forecast the global grain trade flow and its trend by combining various influencing factors.

Considering inter-country grain trade and its influencing factors simultaneously to construct a multilayer network. This multilayer network is called a multiplexed network when it shares the same set of vertices, in this case, the same set of countries, in all layers [6] [7]. After adding the time term, the multiplexed network is regarded as a component at each time snapshot. All components formed through time series constitute the dynamic multiplexed network structure, see **Figure 1**. In this paper, the data on grain trade and its influencing factors over the years will be collected, and the complex network knowledge will be used to represent the data as a dynamic multiplexing network. The link prediction will be made to study the trend of the grain trade in the future.

The research on link prediction of the static network has been quite mature [8] [9] [10] [11] [12], and there have also been some advancements in link prediction in dynamic networks [13] [14] [15] [16]. However, the method that can satisfy the link prediction of a dynamic multiplexed network is sporadic. In dynamic multiplexed networks, a connection in one layer increases the likelihood of connecting the corresponding node pairs in other layers. Therefore, a holistic approach that considers all layers can be superior to one that performs link prediction separately in each layer. Ref. [17] introduced a comprehensive framework called MLP (Multiplex Link Prediction), which simultaneously combines

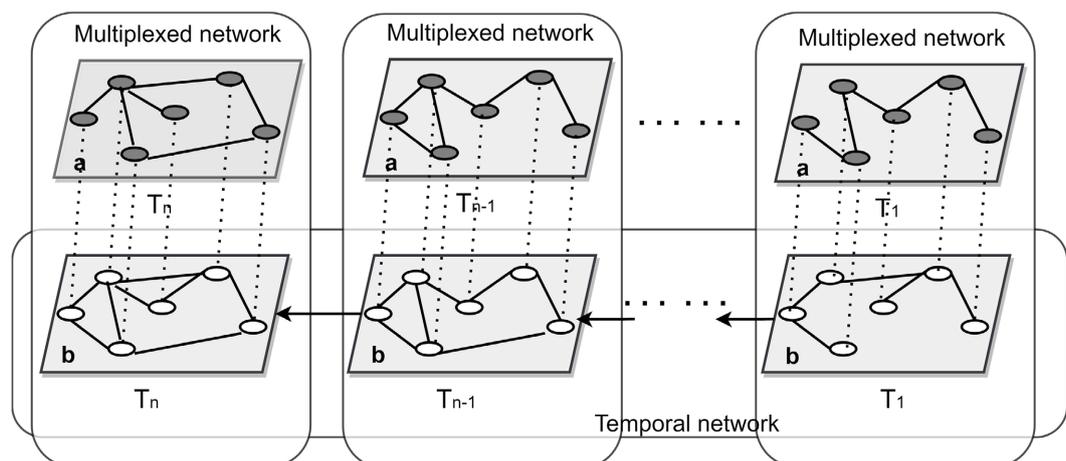


Figure 1. Dynamic multiplexing network diagram. Each vertical box is a multiplexed network component. The horizontal box represents the evolution of the target layer over time.

information from all layers by building the target layer. The MLP framework uses a likelihood-based approach to learn cross-layer dependencies and a temporal decay function to model network dynamics. The ranking aggregation is then used to gather information from multiple topology measures into a scoring matrix that is used to rank potential links. This framework applies to undirected multilayer networks, and we will extend it to weighted directed networks to adapt to the research topic. So, in this paper, the research idea of the MLP framework will be retained in general, and corresponding adjustments will be made in the selection of similarity measures, weight determination methods, and target layer structure to predict the world grain trade network.

In the next section, the source of the data and the construction of a dynamic multiplexing network are described. The third section details the application of the improved MLP framework in grain trade and puts forward a new index to measure prediction accuracy. Sections 4 to 6 provide a description and perspective of the algorithm prediction results and summarize the paper.

2. Data and Network Construction

The three staple grains in the global grain trade are rice, wheat, and corn, which are this study's subjects. The grain trade volume, also known as the grain trade flow, is the total trade quantities of the three different types of grains moving simultaneously in the same direction (unit: kg). The data on commodity trade flows are derived from the World Food and Agriculture Organization (FAO) [18].

Look for indicators that capture grain supply and demand. It was found that the area per capita cropland (ha/per) can partially reflect the extent to which a country's grain supply is satisfied. Generally, a nation with less cropland per person typically has fewer grain supplies and must import grain to meet domestic demand. In addition, the percentage of imported grain in a nation's grain demand can be calculated using the cereal import dependency ratio (%). That is, how reliant a nation's food security is on the world grain market. The cereal import dependency ratio is capped at 100%, meaning the country depends entirely on imports for its cereal demand. It could also be negative, suggesting that the nation's cereals could be exported after domestic demand is met. All indicator data are from the indicator database of FAO. Also, the existence of a trade agreement between nations may have an impact on how frequently commerce occurs. The WTO's Regional Trade Agreement database provides information on trade agreements [18]. The above four metrics that we found to directly or indirectly reflect cereal trade volume, arable land area per capita, cereal import dependency ratio, and the number of trade agreements. For the consistency of data caliber, all data are taken from 2000 to 2020. In the following, the data of these metrics will be used as the basis for constructing the multiplexed network.

Data cleansing was placed before the network creation. Countries with completely missing data were removed, and the remaining 176 trade-participating countries were used as the study population. The records containing a small

amount of missing data were filled in. For instance, cereal import dependency ratio data were taken from the average value of three years, and the missing records for 2019 and 2020 were defined as the same in 2018. Also, the grain trade flow data of less than 100 kg were eliminated for simplicity. After the data are collated, it is ready for network construction.

Different from single network modeling, a dynamic multiplexed network can be modeled as $G = \{G_0, G_1, \dots, G_T\}$, where, $G_t, t = 0, \dots, T$, represent the states of multiplexed network components at different time snapshots. Each component is defined as $G_t = \langle V, E_t^1, \dots, E_t^M \rangle$. The node sets of all graphs are equally represented as V . E_t^m represents the connected edge set of layer m in the time t .

On this basis, the multiplexed network we build consists of four layers: 1) Grain Trade Network. A network of trade flows consisting of the volume of grain imports and exports. This layer is a weighted directed network, which is also used to be the basis for constructing the target layer in the dynamic multiplex network prediction process. 2) Network of Trade Agreement Relationships. If there is a trade agreement between two countries, then there is a two-way connection between the pair of national nodes, and the edge weight is defined as the number of trade agreements between countries. 3) Cultivated Land Area Relationship Network. FAO set the warning line of arable land per person at 0.975 mus (converted to 0.053 hectares per person). According to the basic principles of economics, add the directed edge from the countries with higher per capita cultivated land area to those with lower, and the country with a value less than or equal to 0.053 does not have outdegrees. 4) Network of Dependency Rates on Cereals Imports. Similar to the Cultivated Land Area Network, with an import dependency rate of 5% as the basic self-sufficiency standard. All countries with a rate greater than 5% are regarded as grain-importing countries, and the probability of exporting large quantities of grain is considered as non-existent. Add the directed edge from low dependency to high to simplify the network. The points in the above four network layers are all represented as a country and correspond one-to-one. The schematic diagram of a multiplexed network component in a time snapshot is shown in **Figure 2**.

As for the division of the training set and test set, due to the unique status of the grain trade network, only the grain trade data for each time snapshot are divided into the training set and test set (the reasons stated in Section 3.1). The 10% of the existing trade edges are taken as the test set E_t^P , and the remaining edges are used as the training set E_t^T . The edges that do not exist in the original trade network are classified as “non-existent edges”, denoted as E_t^N . The Grain Trade Network in the dynamic multiplexed network is composed of the training set. The test set is used to test the accuracy of the prediction algorithm.

3. Method Introduction

3.1. Weight Assignment for Layers and Edges

In a dynamic multiplexing network, each network layer is coevolutionary. That

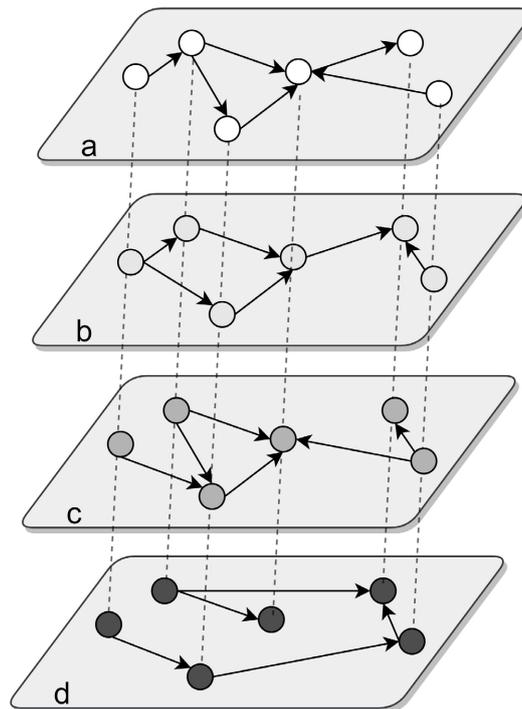


Figure 2. Diagram of a multiplexed network on a snapshot in time.

is, one layer may produce or delete the edge due to the structural changes of other layers [19]. This adds to the difficulty of link prediction for such networks. A target layer can be created, and the network information of each layer on each time snapshot is compressed into the target layer. The link prediction for the target layer uses the information from all layers and achieves the goal of predicting the future simultaneously.

Our objective is to explore the impact of three indicators—trade agreements, arable land per capita and Import dependency ratio—on cereal trade, and to use the level of impact as a basis for calculating and predicting a comprehensive network of relationships for future world cereal trade. This integrated relational network as the target layer to be constructed and it is logical to use layer (a) as the basis of the target layer. The initial state of the target layer inherits exactly the way the points and edges are connected in layer (a), while the weights of the edges need to be redefined. Since there are different levels of correlations among the various network layers, we can quantify such correlations as layer weights that play a role in the calculation of the weights of the target layer edges. L^i and w^i represent layer i and its layer weight respectively. L^* is the target layer. The **Algorithm 1** is the precise building procedure for layer weights and edge weights.

$\Gamma(i)$ is the set of neighbors of i . $w(i, j)$ is the weight of the link between i and j . $s_{out}(i) = \sum_{j \in \Gamma(i)} w(i, j)$ denotes the sum of the out-weights of i , and $s_{in}(j) = \sum_{i \in \Gamma(j)} w(i, j)$ denotes the sum of the in-weights of j . In the equation for the edge weights w_e of the target layer, the first half rate represents the contribution

Algorithm 1. Construction algorithm of layer weight and edge weight.

Input: $\{E^1, \dots, E^M\}$.

Output: The edge weight matrix of the target layer.

Calculate the weights of each layer:

For $m \in \{b, c, d\}$.

The ratio of overlapping edges between each layer and the initial state of the target layer is taken as the weight of each layer.

$$w_m = \text{likelihood}(\text{Link in } L^a \mid \text{Link in } L^m)$$

Calculate the edge weights of the target layer:

For $e \in E^a$, $i \xrightarrow{e} j$, indicates e is the edge from i to j .

$$w_e = \text{rate} + \sum_{m \in \{b, c, d\}} w_m \times \text{linkexist}(e) \quad \text{linkexist}(e) = \begin{cases} 1, & e \text{ exists in layer } m, \\ 0, & e \text{ does not exist in layer } m. \end{cases}$$

The rate part indicates the impact of layer (a), the formula is

$$\text{rate} = \log_{100}(s_{out}(i)) + \log_{100}(s_{in}(j))$$

of layer (a), and the second half indicates the effect of layers except (a) on the edge weights.

3.2. Similarity Indices

The most popular method of link prediction is based on similarity calculation. The similarity index represents the similarity or proximity between nodes. A score S_{xy}^{index} is assigned to each pair of nodes x and y , called the similarity score [20] [21]. The higher the similarity score, the more likely the link exists. Indices are defined in various ways, which mainly depend on calculating the basic topology features of nodes. In this paper, the target network for link prediction is a directed weighted network, and we have not found some similarity indices applicable to this type of network. We make reasonable modifications to some traditional similarity indices for adapting this network. Four similarity indices based on local information are collected in this paper. The new definition and its modification process are shown below.

- **Common Neighbor Index (CN)**

$$S_{xy}^{CN} = |\Gamma(x) \cap \Gamma(y)| \quad (1)$$

This index is defined as the number of common neighbors of nodes x and y [22]. $\Gamma(x)$ denotes the set of neighbors of node x . In a weighted directed network, when x is the start vertex, its neighbor set is represented as $\Gamma_{out}(x)$, and when x is the end vertex, it is denoted as $\Gamma_{in}(x)$. $w(x, y)$ denotes the weight of the edge from x to y . The index is redefined as:

$$S_{xy}^{CN} = \sum_{z \in \Gamma_{out}(x) \cap \Gamma_{in}(y)} w(x, z) + w(z, y) \quad (2)$$

- **Resource Allocation index (RA)**

$$S_{xy}^{RA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{k(z)} \quad (3)$$

where RA is a process based on resource allocation [23]. $k(x)$ is the degree of vertex x , which can be divided into the indegree $k_{in}(x)$ and outdegree $k_{out}(x)$. Each common neighbor is considered as a resource emitter, but the resources they contain are transformed from a unit to the “flow” through the path, denoted by $w(x, z) + w(z, y)$. The average number of shares allocated is also changed to the sum of the out-weights. Redefined as:

$$S_{xy}^{RA} = \sum_{z \in \Gamma_{out}(x) \cap \Gamma_{in}(y)} \frac{w(x, z) + w(z, y)}{s_{out}(z)} \quad (4)$$

- **Adamic-Adar index (AA)**

$$S_{xy}^{AA} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log k(z)} \quad (5)$$

This index assigns greater importance to a common neighbor node with fewer neighbors to measure how closely it is related to the node pair [24]. Similar to the idea of improving RA, improved it to:

$$S_{xy}^{AA} = \sum_{z \in \Gamma_{out}(x) \cap \Gamma_{in}(y)} \frac{w(x, z) + w(z, y)}{\log(1 + s_{out}(z))} \quad (6)$$

- **Preferential Attachment Index (PA)**

$$S_{xy}^{PA} = k(x) \times k(y) \quad (7)$$

The principle of PA is that the more links one already has, the greater the likelihood of new links being generated between that node pair [25]. Redefined as:

$$S_{xy}^{PA} = k_{out}(x) \times k_{in}(y) \quad (8)$$

The similarity calculation of node pairs in the weighted directed target layer will be performed later using the modified indexes.

3.3. Temporal Dependencies Weighted Exponential Model

Sections 3.1 and 3.2 construct a grain trade integrated relational network as the target layer. The information of the multiplexed network is compressed into the weighted directed target layer of each time snapshot separately, and the object of link prediction becomes the target layer. The Temporal dependencies of the network during co-evolution need to be considered in link prediction at the target layer. The similarity of a node pair in a temporal snapshot can be assessed using its similarity indices and those from earlier periods. This is due to the similarity of point pairs in the target layer at the current time being correlated with past periods, and this correlation may weaken as the time span increases. The weighted exponentially decaying model can be used to integrate the similarity of node pairs on each time snapshot for the comprehensiveness of the prediction [17] [26].

Take the improved similarity index CN as an example. Let

$\{S_t^{\text{CN}}(i, j), t = 1, 2, \dots, T\}$ represent the score of the similarity index of the directed edge from i to j at time t . The aggregate similarity score matrix, denoted as $Sim_{t_0}^{\text{CN}}$, is established by the weighted exponential decay model. It integrates the information from $(t_0 - \Delta t)$ to t_0 , and the time span is Δt . A larger time span indicates a wider information horizon for integration, but the information span is not proportional to the information accuracy. The matrix element $sim_{t_0}^{\text{CN}}(i, j)$ is expressed as:

$$sim_{t_0}^{\text{CN}}(i, j) = \sum_{t=t_0-\Delta t}^{t_0} \theta^{t_0+1-t} S_t^{\text{CN}}(i, j) \quad (9)$$

where $\theta \in [0, 1]$ is the smoothing weight of the previous time snapshot, the value of θ determines the contribution level of the score of the previous time to the similarity score of the current. Therefore, Δt and θ will be subsequently used as parameters to be estimated.

The aggregate similarity score matrix under the similarity index S^{CN} is as follows:

$$sim_{t_0}^{\text{CN}} = \begin{bmatrix} sim_{t_0}^{\text{CN}}(1, 1) & \dots & sim_{t_0}^{\text{CN}}(1, N) \\ \vdots & sim_{t_0}^{\text{CN}}(i, j) & \vdots \\ sim_{t_0}^{\text{CN}}(N, 1) & \dots & sim_{t_0}^{\text{CN}}(N, N) \end{bmatrix}_{N \times N} \quad (10)$$

The similarity matrices of the four similarity indices are calculated separately and denoted as $\{Sim_{t_0}^{\text{index}}, \text{index} = \text{CN}, \text{AA}, \text{RA}, \text{PA}\}$.

3.4. Ranking Methods

The last step uses the rank aggregation method [27], where the input lists must contain each element's ranked value. Four matrices containing different similarity metric values were previously computed, which will be sorted based on the values from highest to lowest and transformed into four sorted lists. The lists are merged to obtain a final ranked list containing all node-pair elements. The final ranked list is the basis for link prediction to predict future time node potential and higher probability links. Of course, all this is done under the condition that variables t_0 , Δt and θ have specific values.

Borda's method is a type of rank aggregation method, originally proposed by Jean-Charles de Borda as a voting method. It is an absolute ranking method as it also requires the order of voters' preferences to be given in the form of ranking during the voting process. On each time snapshot, the set of lists is denoted as $L = \{L^{\text{CN}}, L^{\text{RA}}, L^{\text{AA}}, L^{\text{PA}}\}$, which consists of four sorting vectors due to four similarity indices, and the length of each vector is the number of node pairs. It is known that there are $N(N - 1) = 176 \times 175$ sets of node pairs in a network consisting of 176 countries. The ranking of node pairs (i, j) in the sorting vector L^{index} is expressed as $L^{\text{index}}(i, j)$. According to Borda's method, the score of an element in a ranking sequence is the number of sequence elements minus the

ranking value of the element, which is denoted as $b^{index}(i, j) = N - L^{index}(i, j)$. The total Borda score of a node pair is determined by the sum of its scores in each ranking vector, denoted as:

$$B(i, j) = \sum_{index \in \{CN, RA, AA, PA\}} b^{index}(i, j) \quad (11)$$

A node pair with a higher Borda score indicates that it is potential or has a higher probability of appearing in the future, based on which link prediction is achieved. $B(i, j)$ can also be written specifically as $B_{t_0}(i, j)$ when the prediction base period is taken as moment t_0 .

3.5. Definition of Evaluation Metrics for Algorithms

The available data allow us to use 2020 as the base period network to make a link prediction of the network in 2023. As a result, the link prediction's accuracy over a three-year time horizon must be evaluated. Before that, the prediction result data needs to be converted into a prediction network. The process is relatively straightforward. In each time t_0 , node pairs are selected to create links based on the total Borda score from highest to lowest, and the number of node pairs chosen is equal to the number of links in the original network at the moment $(t_0 + 3)$. The network consisting of these selected links is the predicted network. The prediction accuracy of the predictions spanning three years can be obtained by comparing the predicted network built at the base period at the moment t_0 with the original network at the moment $(t_0 + 3)$. The overall prediction accuracy of the algorithm is obtained by averaging the prediction accuracy with different moments as the prediction base period, and its calculation process is illustrated and explained in **Figure 3**.

The accuracy of the link prediction algorithm is usually calculated using the AUC criterion, which can be interpreted as the probability that a randomly selected edge in the test set has a higher score than a randomly selected non-existent edge. The prediction algorithm gives the score there, which in this paper is the total Borda score. However, links between pairs of nodes in trade networks may appear or disappear after a period of time as an indication of the occurrence and discontinuity of trade relationships. Considering this factor, a new index for algorithmic accuracy assessment is defined and named $AUC_{improved}$.

Using the sampling comparison method, one edge e_1 is randomly selected from the test set $E_{t_0}^P$ at time t_0 , the edges e_2 and e_3 are randomly selected from the set of "non-existent edges" $E_{t_0+3}^N$ and the complementary set of $E_{t_0+3}^N$ at time $(t_0 + 3)$. The total Borda scores of these three edges at time t_0 are compared, and the numerator is assigned according to the criteria in **Table 1**.

If e_1 is present in the network at the moment $(t_0 + 3)$, then the algorithm should assign it a higher Borda value compared to e_2 to reflect the accuracy. In this case, when $B_{t_0}(e_1)$ is greater than $B_{t_0}(e_2)$, the numerator plus one; when $B_{t_0}(e_1)$ is equal to $B_{t_0}(e_2)$, the numerator plus 0.5; when $B_{t_0}(e_1)$ is greater

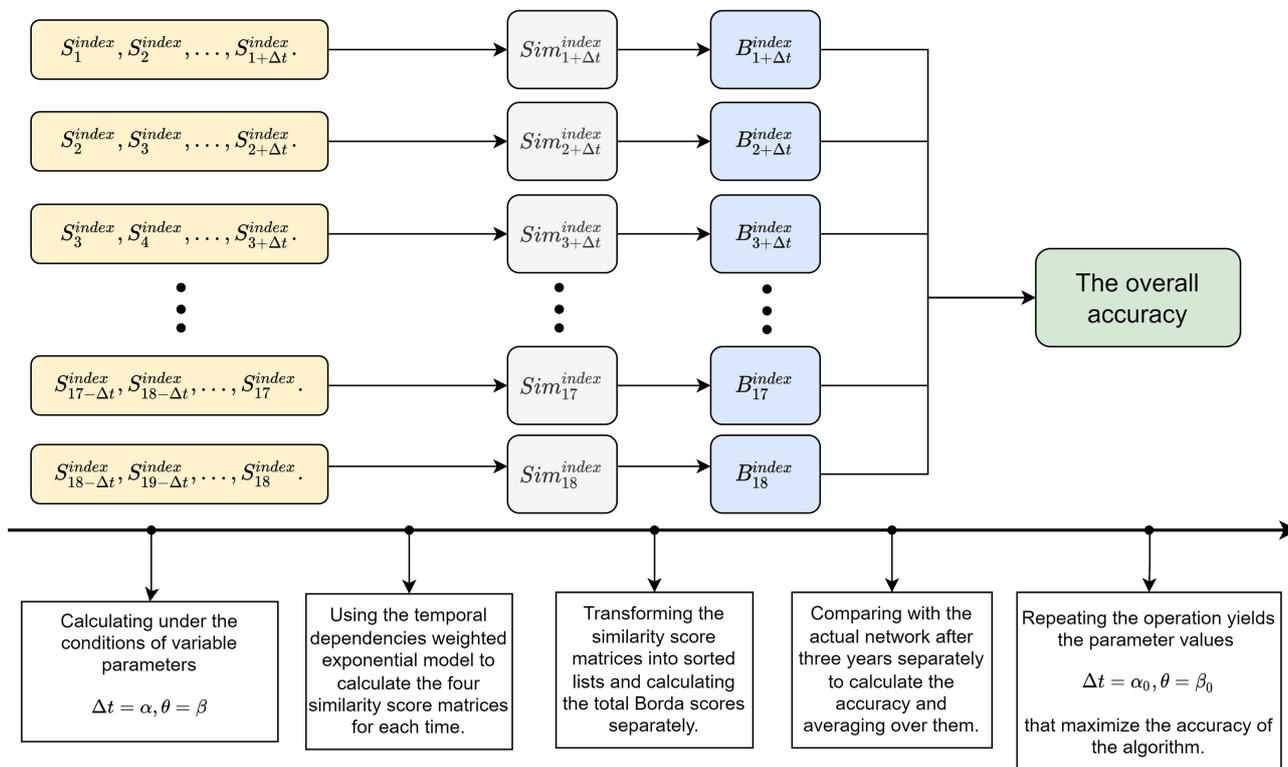


Figure 3. The overall accuracy calculation process of the algorithm.

Table 1. The assignment criteria of the improved link prediction accuracy rubric.

Condition 1	Condition 2	Molecular counting
$\{e_1 \notin E_{t0+3}^N \mid e_1 \in E_{t0}^P\}$	$B_{t0}(e_1) > B_{t0}(e_2)$	sum + 1
	$B_{t0}(e_1) < B_{t0}(e_2)$	sum + 0
	$B_{t0}(e_1) = B_{t0}(e_2)$	sum + 0.5
$\{e_1 \in E_{t0+3}^N \mid e_1 \in E_{t0}^P\}$	$B_{t0}(e_1) < B_{t0}(e_3)$	sum + 1
	$B_{t0}(e_1) > B_{t0}(e_3)$	sum + 0
	$B_{t0}(e_1) = B_{t0}(e_3)$	sum + 0.5

than $B_{t0}(e_2)$, the numerator is not added. Similarly, if e_1 does not exist in the network at the moment $(t0 + 3)$, then the algorithm should assign it a lower Borda value compared to e_3 in order to reflect the accuracy. The denominator of the indicator is the number of repetitions of the sample comparison, denoted as n . The final expression of the indicator is obtained: $AUC_{improved} = sum/n$.

The overall prediction accuracy is obtained by taking different moments as the base period and finding the average of the $AUC_{improved}$. Compared with the usual link prediction AUC metric, the improved AUC metric applies similar concepts of “true positive” and “false negative”. It follows the principle of the ROC curve more closely.

4. Results

Based on the methodology given in Section 3.1 of this article, the ratio of the overlapping edges between other networks and the network in layer (a) is used as the weight of each layer at each time snapshot. **Figure 4** displays heat maps representing the layer weights at certain times.

According to the research, there is a small increase with the year in the tightness of the linkage between all three indicators and cereal trade in general. The cereal import dependence network is the strongest linkage with the cereal trade flow network among the layers, and performance is the largest layer weight. And the trade agreement network has little overlap with the cereal trade flow network, even below 0.5. It follows that for an irreplaceable commodity like grains, the existence of generalized trade agreements between trading countries does not have a decisive impact on their grains' trade behavior unless it is a targeted one. For the sake of simplicity and efficiency of the model, the Trade Agreement Relationship Network is removed from the multiplexed network, and its influence is not added to the calculation of the connected edge weights of the target layer.

After calculating the edge weights and measuring the similarity indexes based on the algorithm described in Section 3, the selection of parameters is performed in order to ensure the accuracy of the link prediction algorithm. As mentioned earlier, there are two parameters to be estimated in the model, namely θ and Δt , which represent the smoothing weight and the time span in the temporal dependencies weighted exponential model, respectively. Now change their values and calculate the overall accuracy for different scenarios.

After fitting the surface, it is found that the algorithm has the highest accuracy for the time span of 10 years and the smoothing weight between 0 and 0.4. Based on this range, to further refine the selection of parameters. See **Figure 5(c)**, the highest accuracy of the algorithm was obtained for the case of $\Delta t = 10$, $\theta = 0.26$, which was 0.885, and the values of this parameter were substituted into the model for link prediction.

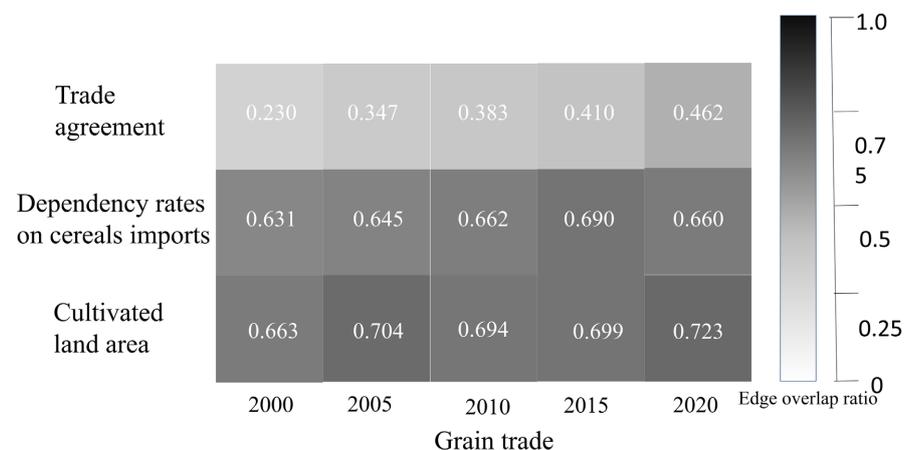


Figure 4. Heat maps were taken for each layer weight at 5-year intervals from 2000 to 2020.

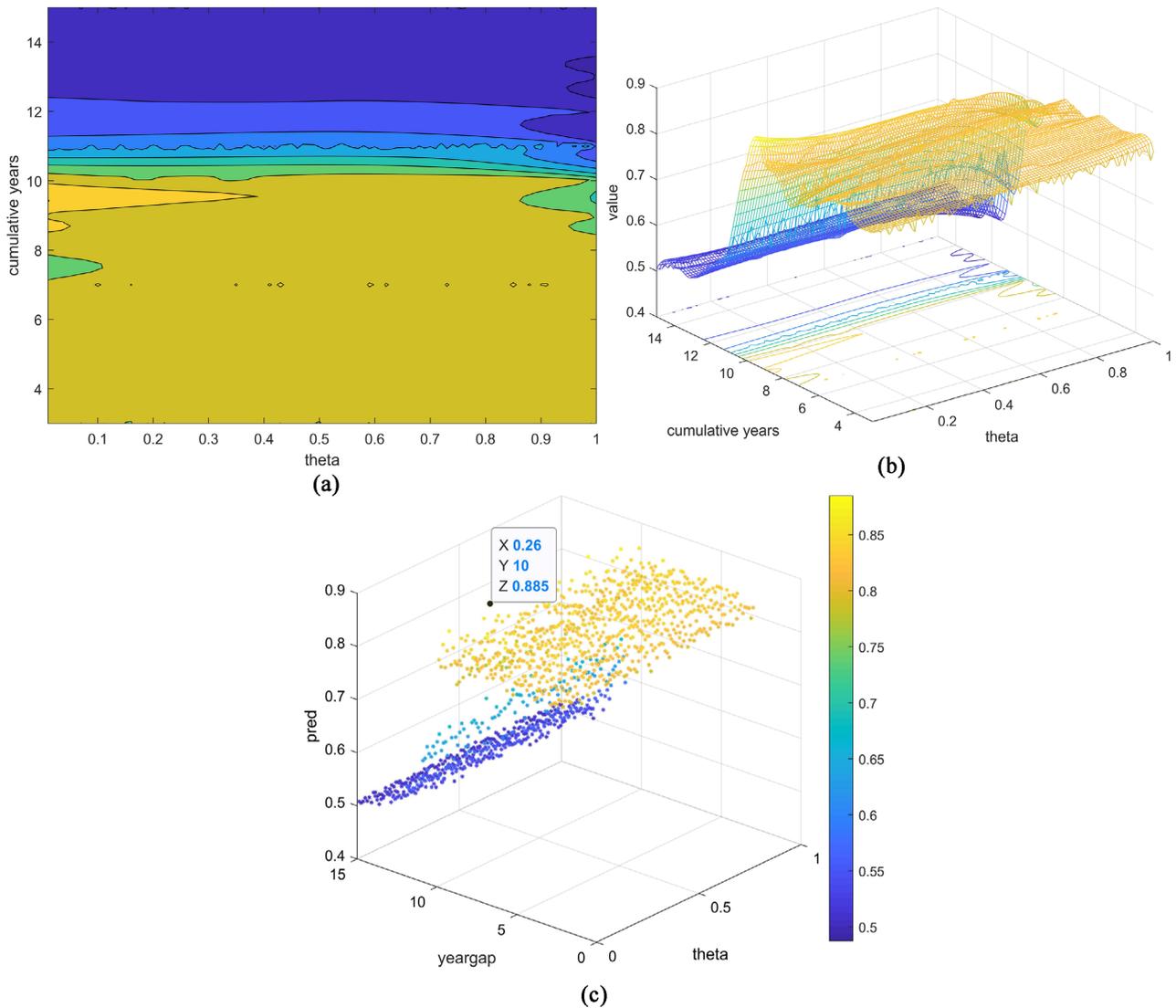


Figure 5. The improved AUC values for different parameters, (a) and (b) are the fitted thermograms, (c) is a discrete plot representing the actual optional parameter points.

The predicted results are shown in **Figure 6**. The nodes of different colors indicate the trade participating countries from different continents. The number indicates the country number, and the size of the points indicates the degree of increase and decrease. To distinguish more clearly, the red and blue edges represent the possible broken and generated links obtained from the link prediction, respectively. The dark red and dark blue indicates the parts of them with a higher probability of occurrence.

Most of the vanishing edges occur within continents, and most of the increasing edges occur between continents. In 2023, intercontinental trade will increase relative to the previous grain trade network, and intra-continental grain trade transactions will decrease. However, projections suggest that trade in grain between European countries will grow closer. In particular, India’s cereal exports to Africa are likely to decline. In countries such as the United States and Australia,

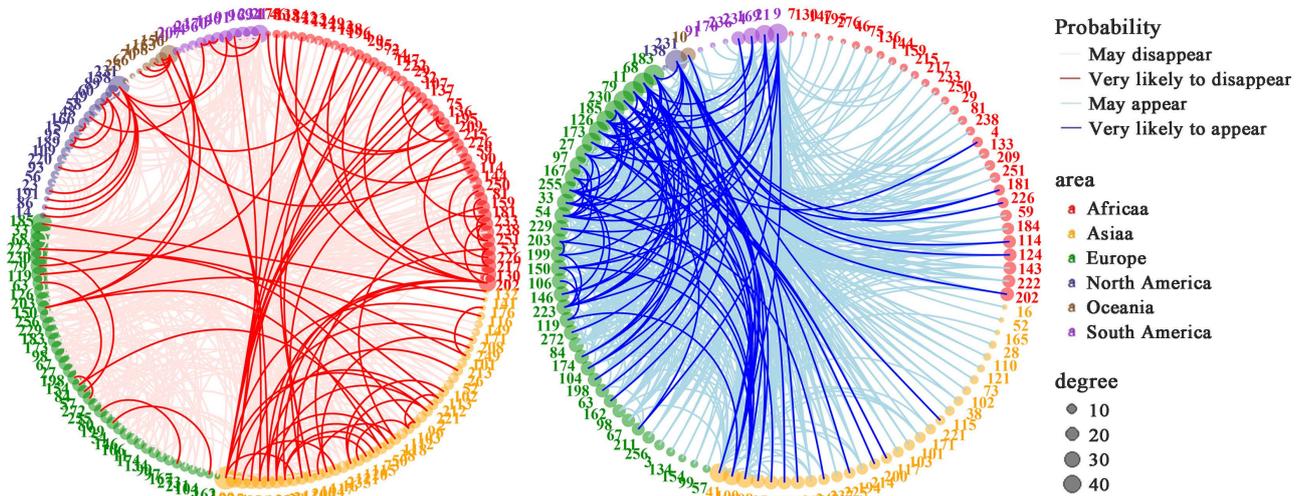


Figure 6. The results of the link projection with the base period of 2020, based on $\Delta t = 10$, $\theta = 0.26$. The left panel shows the edges that have the probability of disappearing from the world grain trade network in 2023, and the right panel shows the new linked edges that may be created.

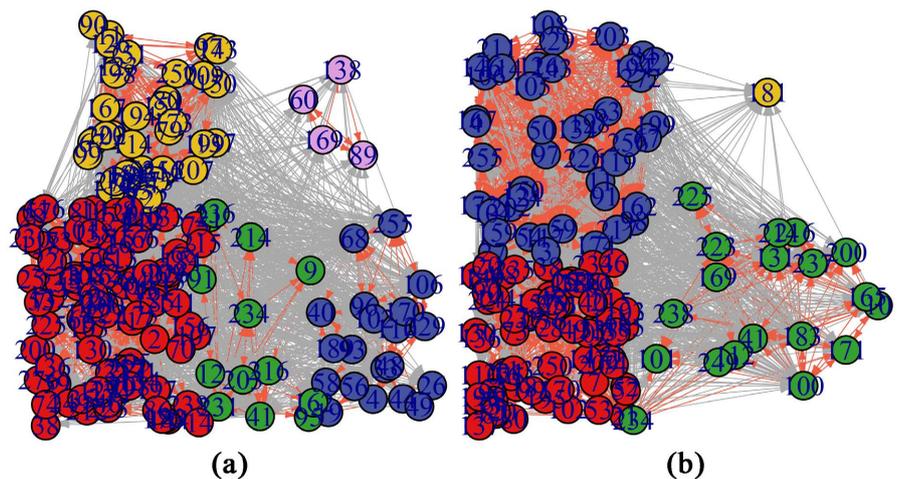


Figure 7. Schematic diagram of the community partition results before (a) and after (b) the prediction. Points with different numbers in the figure denote different countries participating in trade, and countries are numbered according to the FAO uniform standard. Points with different colors indicate that they are divided into different communities. Grey lines indicate cross-community trade, while red lines represent intra-community trade operations.

the frequency of trade with countries within their respective continents may decrease. The export of grain from Austria and Argentina is expected to increase substantially. On the other hand, Brazil, Canada, Russia, Ukraine, France, and other major grain-trading countries show a decline in intra-continental exports and an increase in intercontinental exports in their forecasting results.

The results of the link prediction were used to build a new network and perform community partition, which was compared with the division results of the previous data (see **Figure 7**). It was found that the number of communities was reduced from five to four, and the modularity was reduced accordingly. The densi-

ty of connected edges within communities' increases, and the membership composition of the societies will also move from a biased division based on geography to a more integrated direction.

5. Discussion

Research and prediction work on world trade networks has been widely carried out in various technical fields, whether it is the description of the static characteristics of the network, the analysis of clustering effects, or the prediction of possible future links. Link prediction based on dynamic multiplexed networks has far-reaching implications, using a wider range of data and making the experimental results more interpretable, especially compared to "black box" neural network-like methods.

The algorithm in this paper is based on the MLP framework with several changes: 1) Extending the MLP framework to weighted directed networks and changing the weight calculation method based on it. 2) The superposition term of the weighted exponential decay model is fixed to reduce the time complexity of the computation. 3) Constructed the algorithm accuracy evaluation index $AUC_{improved}$ that better fits the principle of the ROC curve. 4) Improved the four similarity indexes CN, RA, AA, and PA to apply to the weighted directed network, and these improvements are based on the respective construction principles of the indexes. In order to illustrate the rationality and applicability of the metrics, several weighted directed open network datasets are applied to be tested in **Table 2** [28] [29] [30] [31] [32].

The tests found that the improved similarity indices perform well overall on weighted directed open data networks, especially in denser networks where the number of edges is much larger than the number of points.

Of course, there are some limitations to this paper. Subjective judgment plays a part in the choice of calculation methods for layer weights and edge weights. There may be more accurate and scientific computational formulae that have not been tried. Also, due to the data being in years and the time span being large, the data cannot be divided very finely, and the accuracy of the calculation is reduced. Similarly, link predictions might be more accurate if complete layers of network data were available for the most recent year.

Table 2. Applicability test results of weighted directed network similarity index.

Networks	No. of nodes	No. of edges	Average weight	AUC			
				CN	RA	AA	PA
Subreddit hyperlink network	6555	3879	1.246	0.5950	0.5705	0.5860	0.5830
Trust network	3683	22650	1.996	0.8025	0.6501	0.7085	0.8045
E-mail network	309	2728	20.912	0.9525	0.8035	0.8830	0.9495
Messages network	1899	18267	2.951	0.739	0.5855	0.6845	0.7670

6. Conclusions and Future Work

Ultimately, our model has a high prediction accuracy for grain trade networks and achieves some significant results. Overall, the results show that the number of generalized trade agreements has a small impact on grain trade across countries. Summarizing the reason may be that food-importing countries must import food from some large exporting countries to meet their domestic demand even though there are no trade preference agreements between them. Unless the data specific to grain trade agreements are used, such data are difficult to obtain comprehensively. The impact of the per capita arable land indicator is relatively large and likely to increase for grain imports and exports. The once-major cereal-exporting countries still play an essential role in this trade network. The world cereal trade will develop in the direction of more frequent intercontinental trade and gradually weaken the intracontinental cereal trade. The results of the community partition also show this trend, which is reflected in a smaller number of communities compared to the previous ones.

During the study, some exciting ideas arose that could serve as a prospect for future research. It may be useful to not only improve existing similarity indices based on them for weighted directed networks, but also to try to create new indices. When indexes are too difficult to create, combinations of metrics with higher precision can be found by combining them.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Duan, J., Nie, C.L., Wang, Y.Y., Yan, D. and Xiong, W.W. (2021) Research on Global Grain Trade Network Pattern and Its Driving Factors. *Sustainability-Basel*, **14**, Article 245. <https://doi.org/10.3390/su14010245>
- [2] Rosegran, M.W. and Cline, S.A. (2003) Global Food Security: Challenges and Policies. *Science*, **302**, 1917-1919. <https://doi.org/10.1126/science.1092958>
- [3] Brooks, D.H., Ferrarini, B. and Go, E.C. (2013) Bilateral Trade and Food Security. *Journal of International Commerce Economics and Policy*, **4**, Article ID: 1350015. <https://doi.org/10.1142/S1793993313500154>
- [4] Coyle, W., Gehlhar, M., Hertel, T.W., Wang, Z. and Yu, W. (1998) Understanding the Determinants of Structural Change in World Food Markets. *American Journal of Agricultural Economics*, **80**, 1051-1061. <https://doi.org/10.2307/1244203>
- [5] Puma, M.J., Bose, S., Chon, S.Y. and Cook, B.I. (2015) Assessing the Evolving Fragility of the Global Food System. *Environmental Research Letters*, **10**, Article ID: 024007. <https://doi.org/10.1088/1748-9326/10/2/024007>
- [6] Battiston, F., Nicosia, V. and Latora, V. (2014) Structural Measures for Multiplex Networks. *Physical Review E*, **89**, Article ID: 032804. <https://doi.org/10.1103/PhysRevE.89.032804>
- [7] Kivela, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y. and Porter, M.A. (2014) Multilayer Networks. *Journal of Complex Networks*, **2**, 203-271.

- <https://doi.org/10.1093/comnet/cnu016>
- [8] Wang, L., Hu, K. and Tang, Y. (2014) Robustness of Link-Prediction Algorithm Based on Similarity and Application to Biological Networks. *Current Bioinformatics*, **9**, 246-252. <https://doi.org/10.2174/1574893609666140516005740>
- [9] Liben-Nowell, D. and Kleinberg, J. (2007) The Link-Prediction Problem for Social Networks. *The Journal of the Association for Information Science and Technology*, **58**, 1019-1031. <https://doi.org/10.1002/asi.20591>
- [10] Kaya, B. (2020) Hotel Recommendation System by Bipartite Networks and Link Prediction. *Journal of Information Science*, **46**, 53-63. <https://doi.org/10.1177/0165551518824577>
- [11] Wu, J.H., Shen, J., Zhou, B., Zhang, X.Y. and Huang, B.H.. (2019) General Link Prediction with Influential Node Identification. *Physica A: Statistical Mechanics and its Applications*, **523**, 996-1007. <https://doi.org/10.1016/j.physa.2019.04.205>
- [12] Abbas, K., Abbas, A., Dong, S., Niu, L., Yu, L., Chen, B., Cai, S.M. and Hasan, Q. (2021) Application of Network Link Prediction in Drug Discovery. *BMC Bioinformatics*, **22**, Article No. 187. <https://doi.org/10.1186/s12859-021-04082-y>
- [13] Ahmed, N.M. and Chen, L. (2016) An Efficient Algorithm for Link Prediction in Temporal Uncertain Social Networks. *Information Sciences*, **331**, 120-136. <https://doi.org/10.1016/j.ins.2015.10.036>
- [14] Ahmed, N.M., Chen, L., Wang, Y.L., Li, B., Li, Y. and Liu, W. (2016) Sampling-Based Algorithm for Link Prediction in Temporal Networks. *Information Sciences*, **374**, 1-14. <https://doi.org/10.1016/j.ins.2016.09.029>
- [15] Güneş, İ., Gündüz-Öğüdücü, Ş. and Zehra, Ç. (2016) Link Prediction Using Time Series of Neighborhood-Based Node Similarity Scores. *Data Mining and Knowledge Discovery*, **30**, 147-180. <https://doi.org/10.1007/s10618-015-0407-0>
- [16] Niladri, S., Saptarshi, B., Sukumar, N. and Ranbir, S.S. (2018) Temporal Link Prediction in Multi-Relational Network. *World Wide Web*, **21**, 395-419. <https://doi.org/10.1007/s11280-017-0463-z>
- [17] Hajibagheri, A., Sukthankar, G. and Lakkaraju, K. (2016) A Holistic Approach for Link Prediction in Multiplex Networks. 2016 *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, San Francisco, 18-21 August 2016, 1079-1086. <https://doi.org/10.1109/ASONAM.2016.7752375>
- [18] UN FAO (2022) FAOSTAT Statistics Database. <https://www.fao.org/faostat/en/#data/TM>
- [19] Takesue, H. (2021) Symmetry Breaking in the Prisoner's Dilemma on Two-Layer Dynamic Multiplex Networks. *Applied Mathematics and Computation*, **388**, Article ID: 125543. <https://doi.org/10.1016/j.amc.2020.125543>
- [20] Barabási, A.-L. (2009) Scale-Free Networks a Decade and Beyond. *Science*, **325**, 412-413. <https://doi.org/10.1126/science.1173299>
- [21] Lü, L. and Zhou, T. (2011) Link Prediction in Complex Networks: A Survey. *Physica A: Statistical Mechanics and its Applications*, **390**, 1150-1170. <https://doi.org/10.1016/j.physa.2010.11.027>
- [22] Newman, M.E.J. (2001) Clustering and Preferential Attachment in Growing Networks. *Physical Review E*, **64**, Article ID: 025102. <https://doi.org/10.1103/PhysRevE.64.025102>
- [23] Zhou, T., Lu, L. and Zhang, Y.C. (2009) Predicting Missing Links via Local Information. *The European Physical Journal B*, **71**, 623-630. <https://doi.org/10.1140/epjb/e2009-00335-8>

- [24] Adamic, L.A. and Adar, E. (2003) Friends and Neighbors on the Web. *Social Networks*, **25**, 211-230. [https://doi.org/10.1016/S0378-8733\(03\)00009-1](https://doi.org/10.1016/S0378-8733(03)00009-1)
- [25] Wang, X. and Sukthankar, G. (2014) Link Prediction in Heterogeneous Collaboration Networks. In: Missaoui, R. and Sarr, I., Eds., *Social Network Analysis—Community Detection and Evolution*, Springer, Berlin, 165-192. https://doi.org/10.1007/978-3-319-12188-8_8
- [26] Acar, E., Dunlavy, D.M. and Kolda, T.G. (2009) Link Prediction on Evolving Data Using Matrix and Tensor Factorizations. *Proceedings of the IEEE International Conference on Data Mining Workshops*, Miami, 28 December 2009, 262-269. <https://doi.org/10.1109/ICDMW.2009.54>
- [27] Sculley, D. (2007) Rank Aggregation for Similar Items. *Proceedings of the 2007 SIAM International Conference on Data Mining*, Minneapolis, 26 April 2007, 587-592. <https://doi.org/10.1137/1.9781611972771.66>
- [28] Kumar, S., Hamilton, W.L., Leskovec, J. and Jurafsky, D. (2018) Community Interaction and Conflict on the Web. *Proceedings of the 2018 World Wide Web Conference on World Wide Web*, Lyon, April 2018, 933-943. <https://doi.org/10.1145/3178876.3186141>
- [29] Kumar, S., Hooi, B., Makhija, D., Kumar, M., Faloutsos, C. and Subrahmanian, V.S. (2018) REV2: Fraudulent User Prediction in Rating Platforms. *Proceedings of the 11th ACM International Conference on Web Search and Data Mining*, Los Angeles, February 2018, 333-341. <https://doi.org/10.1145/3159652.3159729>
- [30] Kumar, S., Spezzano, F., Subrahmanian, V.S. and Faloutsos, C. (2016) Edge Weight Prediction in Weighted Signed Networks. *Proceedings of the IEEE 16th International Conference on Data Mining (ICDM)*, Barcelona, December 2016, 221-230. <https://doi.org/10.1109/ICDM.2016.0033>
- [31] Paranjape, A., Benson, A.R. and Leskovec, J. (2017) Motifs in Temporal Networks. *Proceedings of the 10th ACM International Conference on Web Search and Data Mining*, Cambridge, 2 February 2017, 601-610. <https://doi.org/10.1145/3018661.3018731>
- [32] Panzarasa, P., Opsahl, T. and Carley, K.M. (2009) Patterns and Dynamics of Users' Behavior and Interaction: Network Analysis of an Online Community. *The Journal of the Association for Information Science and Technology*, **60**, 911-932. <https://doi.org/10.1002/asi.21015>