Mining World Indicators for Analyzing and Modeling the Development of Countries

HONG HUANG, MINGYUAN CHI, YU SONG, and HAI JIN, The National Engineering Research Center for Big Data Technology and System, Key Laboratory of Service Computing Technology and System, Ministry of Education, and School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China

The world indicators released by the World Bank or other organizations usually give the basic public knowledge about the world. However, separate and static index lacks the complex interplay among different indicators and thus cannot help us have an overall understanding of the world. To this end, we study the world indicators from a different angle. Firstly, we discover that there exist correlations between indicators either from a static view or from a dynamic view. Moreover, taking the trade and diplomatic relationships into consideration, we construct a multi-relational network to depict the interactions between different countries, and propose a Multiple Relations to Vector (MR2vec) model to study world indicators from a network perspective. The experimental results show the changes of world indicators are predictable with the proposed model, and our proposed MR2vec has wide adaptability in predicting multi-relation networks.

CCS Concepts: • **Information systems** → **Data analytics**; **Data mining;**

Additional Key Words and Phrases: World indicator, data mining, network embedding, dynamic network, multi-relation

ACM Reference format:

Hong Huang, Mingyuan Chi, Yu Song, and Hai Jin. 2022. Mining World Indicators for Analyzing and Modeling the Development of Countries. *ACM/IMS Trans. Data Sci.* 2, 4, Article 30 (March 2022), 27 pages.
https://doi.org/10.1145/3488059

30

1 INTRODUCTION

World development indicators are becoming indispensable tools for assessing and promoting global developments and reform strategies [25]. Indicators are widely used at the national level and are increasingly important in global governance. For different countries and regions in the world, several comprehensive indicators have been proposed that focus on social, economic, political, and environmental issues, such as the quality of life index, human happiness score, health index, and sustainable welfare index. Development agencies such as the World Bank have developed a wide range of indicators and regularly publish them. The widely used indicators, such as

The research was supported by National Natural Science Foundation of China (No. 61802140).
Authors' address: H. Huang, M. Chi, Y. Song, and H. Jin, The National Engineering Research Center for Big Data Technology and System, Key Laboratory of Service Computing Technology and System, Ministry of Education, and School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China; emails: {honghuang, mingyuan.chi, yusonghust, hjin}@hust.edu.cn.

2577-3224/2022/03-ART30 $15.00
https://doi.org/10.1145/3488059

the increasing GDP per capita, have been considered as one of the important indicators of the rapid development of a country. Therefore, understanding the development of each country in the world depends more and more on the indicator data.

World development indicators can accurately reflect the current state of development in various countries and regions in the world and reflect their commonalities and differences. As a result, studies of world development indicators have attracted scholars from various disciplines to participate. Sociologists often reveal the laws of social development by studying world indicators. For example, John R. Carter [6] analyzed the correlation between economic freedom and income inequality through statistics on per capita income, political structure, education, population, and industry composition of countries and regions around the world. Statisticians tend to study the statistical characteristics of the world's indicators to quantify the level of development of various countries and regions in the world. For example, the relationship between world education finance policy and higher education opportunities can be explored by studying the development indicators of 86 countries and regions [39]. These studies have achieved very influential success, revealing the development trend of the world and providing a basis for developing countries and regions in the world to formulate development strategies. However, the massive data on world indicators are not fully utilized. On one hand, researchers often focus on qualitative or quantitative studies on the data level or certain indicators, and cannot comprehensively consider the *interactions* among various indicators. On the other hand, the *dynamic* properties of these indicators have usually been ignored. To this end, we aim at studying the world indicators from different perspectives, not only to capture the dynamic properties of varying indicators but also considering the inter-relationship among them.

The world indicators are broadly correlated with each other. Thus it is possible to use some easily accessible world indicators to predict the other significant but hard-to-collect world indicators. In this article, we collect world indicators from multiple sources and study them from different perspectives. First of all, through the correlation found between the static world indicators, we obtain some interesting observations. For example, there is a positive correlation between sugar consumption and the GDP of a country. We also explore the dynamic development trends of different countries and regions by a time series clustering algorithm, namely, improved KSC. By such an algorithm, we discover the temporal patterns of world indicators' development from a dynamic perspective preliminarily. In the end, considering the trade and diplomatic relationships, we construct a network to describe the complex interactions between different countries and propose a **Multiple Relations to Vector (MR2vec)** model to study world indicators from the network perspective.

Our proposed MR2vec is an end-to-end multi-relational representation model composed of four parts: Variable Generator, Feature Extractor, Graph Extractor, and Time Extractor. The first one enables the missing data trainable. By reducing manual pre-processing and subsequent processing, the model is as far as possible from the original input to the final output, giving the model more space for automatic adjustment based on the data and increasing the model fit. For our model, we do not need to pre-process the input data. But for other algorithms, it's a must. Furthermore, the remaining three parts perform feature extraction from the correlation of indicators, bilateral relations, and time sequence to get a comprehensive representation. We did several experiments and case studies to prove the effectiveness of our model. The experiments validate our model's outstanding performance in predicting the world indicators and the relationships between countries and regions. Moreover, we also studied the adaptability of our model and found it performs well in various situations.

Our main contributions are as follows:

— We are the first to study the world indicators from a network view, and we study the world indicators' correlation from both the static and the dynamic perspectives.

Table 1. World Development Indicators

Indicator name	# Countries or regions	Start	End
Proportion of rural population	258	1960	2015
Population growth rate	258	1960	2015
Proportion of population aged 0–14	258	1960	2015
Life expectancy at birth	258	1960	2014
proportion of females survive to 65 years old	258	1960	2014
Fertility rate	258	1960	2015
CO2 emission	264	1960	2016
Government health expenditure	190	1995	2010
GDP	161	1980	2011
Sugar or sweeter consumption	148	1961	2004

— We propose a multi-relation representation learning model for the world indicators' network, which not only captures the topology structure, the attributes of entities, the time-sequential information, but also is robust to the missing values.
— Experiment results show that our proposed model outperforms all baseline methods. Moreover, our model can be used not only to predict world indicators, but also suitable for other tasks with time-series heterogeneous graphs or multi-relational network prediction.

The rest of this article is organized as follows. Section 2 introduces the dataset used in this article and defines the problem; Section 3 discusses observations from the world indicators from different views. Section 4 presents the proposed multi-relation representation learning model MR2vec; Section 5 presents our experiments and results, and Section 6 discusses related works. Section 7 concludes.

2 DATA AND PROBLEM DEFINITION

2.1 Dataset Description

The data we used in this study come from three parts: world development indicators, international trade, and diplomatic relationships among different countries and regions.

World development Indicators. The world development indicators include some indexes that demonstrate the world's health and development situation over more than 40 years for hundreds of countries and regions around the world. We collect these datasets from several sources, like the world bank and other organizations.[1,2,3] These datasets are able to describe world development using its population dynamics, birth rates, sugar consumption, and so on. The detailed statistic of the dataset is given in Table 1.

International Trade Relationship. The international trade dataset consists of import and export trade history among sovereign countries and regions in the international system over the years. This dataset is collected from the **Correlates of War (COW)** project homepage.[4] In order to be consistent with the world indicator data, we use international trade relations data from 1960 to 2014, including 754,386 trade relations records. Each international trade relation records includes the amount of import trade and export trade between two countries from each other [2, 3].

[1]https://www.kaggle.com/census/international-data.
[2]https://www.kaggle.com/worldbank/world-development-indicators.
[3]https://www.kaggle.com/angelmm/healthteethsugar.
[4]http://www.correlatesofwar.org/data-sets/bilateral-trade.

Diplomatic Relationship. The diplomatic relationship dataset mainly includes the diplomatic relations between hundreds of countries and regions in the world. There are in total of five types of diplomatic relations: non-diplomatic exchanges, agency level, minister-level, ambassador level, and others. In order to be consistent with the previous data on the year, we use data from the diplomatic relationships between 1960 and 2005, including 251,750 records of diplomatic relationships [4].

2.2 Problems

In this article, we aim at exploring the changes in world indicators of different countries or regions and attempt to bring up a robust model to modify the changes. To be more specific, we try to answer the following questions:

- **Q1:** Are the world indicators closely correlated with each other from a static perspective?
- **Q2:** Does the development of the world indicators follow some regular patterns from a dynamic perspective?
- **Q3:** Are the developments of world indicators and inter-country relationships predictable?
- **Q4:** How is the parameter sensitivity of our proposed model?
- **Q5:** How is the adaptability of our proposed model on different datasets?

In the rest of this article, we will answer the above five questions in order to make the context of the article more clear.

3 OBSERVATIONS FOR WORLD INDICATORS

In this part, we will present some observations for world indicators and then explore the correlation among all the world indicators from both a static and dynamic view.

3.1 World Indicators' Correlation from a Static Perspective (Q1)

Since there may be a positive or negative correlation between world indicators, it is of great significance to quantitatively explore the correlation coefficients. In this part, we will explore the correlation among all the world indicators. We use Spearman correlation coefficient [28] to calculate the correlation. Figure 1 gives the results. From the figure, we can find that agricultural land proportion and ratio of the female population have little correlation with other world indicators, that is to say, these two indicators are relatively independent with others. GDP is positively correlated with health-related indicators (e.g., adult literacy, sugar consumption, health expenditure), while negatively correlated with population index (e.g., fertility rate, population ages 0–14, population growth rate, and rural population rate).

3.2 World Development Patterns from a Dynamic Perspective (Q2)

In this part, we will examine the correlation of world development indicators from a dynamic view. As we know, the world is keeping developing, and its indicators are keeping changing over time as well. Thus, we exhibit its temporal dynamics and show the world indicators' dynamic development patterns.

In order to mine temporal patterns of world indicators, we treat the data of each indicator in each year as a time series, and we aim at finding out the trend of each time series, namely, the trend of each world indicator. With the knowledge of all the trends, we may discover the country with similar ones and hereby uncover the development patterns of countries and regions. Here, we use K-SC clustering algorithm [38] to cluster the time series of different development indicators for all countries and regions by their trend. By using a similarity metric that is insensitive to scaling

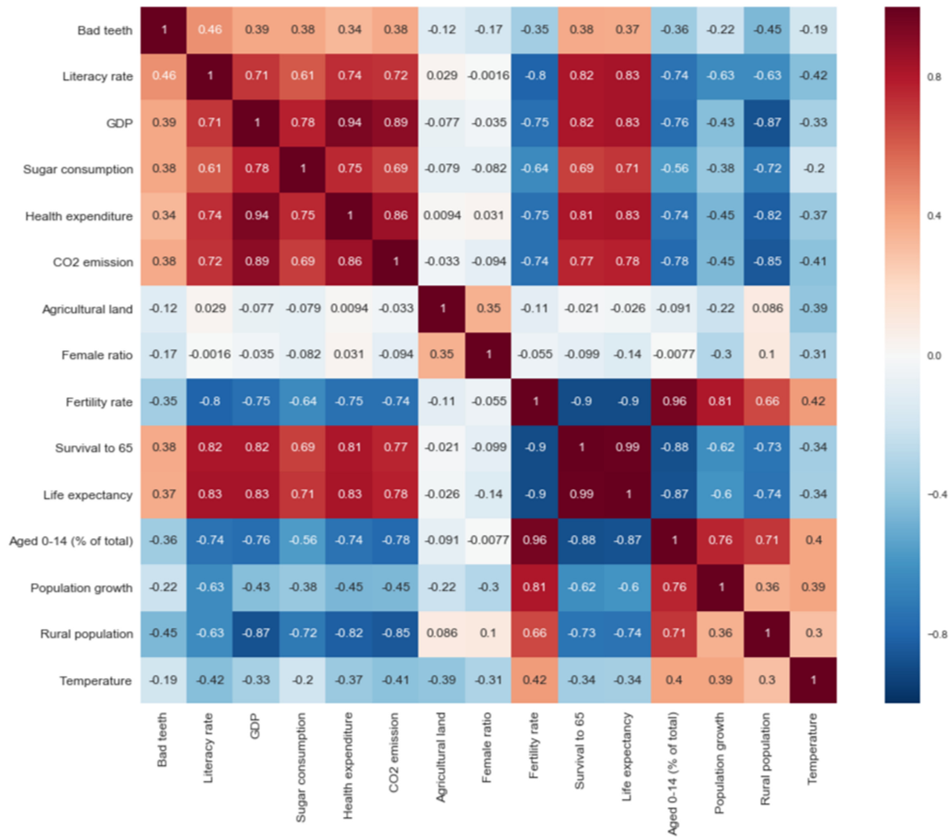


Fig. 1. Spearman correlation coefficient of world indicators.

and shifting, this algorithm can achieve a better clustering performance than traditional clustering algorithms like K-Means.

However, the K-SC algorithm cannot be directly applied here due to two reasons: Firstly, the length of our time series is not fixed; Secondly, the data points in each time series are usually not much enough to get a smooth trend in order to be examined by K-SC algorithm. In this end, we propose an improved K-SC algorithm to solve these two issues. The improved K-SC algorithm leverages polynomial fitting technique with cross-validation that prevents from overfitting. To notify that we can replace the fitting technique with other ones like linear fitting, exponential fitting, and logarithmic fitting. However, polynomial fitting is tested to perform the best in our scenario. To make it simple, we use polynomial fitting in this article.

Assume that there are N time series, then we do N times fitting to fix the order L. For the ith time series, we first calculate the coefficients of the polynomial, and then calculate the estimated value of the ith time series. After calculating the residual the sum of squares, we can find the best L that makes the objective the smallest, which is

$$arg \min_{L} \frac{1}{N} \sum_{i=1}^{N} \left[y_i - \sum_{l=1}^{L} x_i^l b_{i,l} \right]^2, \tag{1}$$

where $b_{i,l}$ is the polynomial coefficient of the l item for the ith time series. y_i is the actual value of the ith time series. The detailed improved K-SC algorithm is given in Algorithm 1.

ALGORITHM 1: Improved K-SC Clustering Algorithm

Input: Raw time series x_i, i = 1, 2, . . . , N, The number of cluster K, Initial assignments C = $(C_1, C_2, \ldots, C_k)$
Output: Cluster center,Cluster label
for *i = 1 to N* **do**
  Use equation (1) to determine the polynomial order L;
end
for *i = 1 to N* **do**
  $p_i \leftarrow$ polynomial fitting (data = x_i, order = L);
  (p_i means the polynomial coefficients);
  $y_i \leftarrow$ Use p_i to sample more time points to get a new time series;
end
C, Label = K-SC (y, C, K);
return C, Label

We then use our improved K-SC algorithm to cluster world indicators' time-series data. According to Equation (1), the best polynomial order is 4 for our selected world indicators. On the other hand, after several test experiments, the number of clusters can be set as 4 in order to get better temporal patterns.

K-SC Clustering Setup. We use all of the world development indicators to conduct our experiment. For each development indicator, we use the values over the years of the indicator to get a time series for each country. First, we apply the Equation (1) to determine the best polynomial order is 4. Next, we use polynomial fitting to sample more time points to get a longer time series as well as keep an all-time series having the same length. After obtaining these time series, we cluster all the time series for each development indicator. But the K-SC clustering algorithm, like other clustering algorithms, needs to first determine the number of clusters. Since this is an open question, we first tried different numbers of clusters. We are based on one principle: we want to summarize the development patterns of world development indicators with the least possible temporal patterns. We finally summarize four representative common temporal patterns of world indicators for both indicators of growing type and decline type. The illustration of cluster centroid is shown in Figure 2.

From Figure 2, we have four typical temporal patterns for world indicators' development: keeping stable increasing, keeping decreasing, keeping unstable increasing, and keeping stable.

Furthermore, we randomly choose 33 representative countries and regions. As an example, we show the classification for sugar consumption, rural population ratio, and GDP, respectively, in Table 2. We can see that in terms of sugar consumption, there is no significant change in Europe and North America, as temporal pattern C_4 shown, while changes in developing countries in Asia and northern Africa are intense, shown by the temporal patterns C_1 and C_3, which indicates that these country's sugar consumption is continuously improving, with a steady improvement of living standards in these countries and regions in the same period. At the same time, the proportion of the rural population in these countries and regions is also declining, showing strong correlation with the rapid development of these countries and regions.

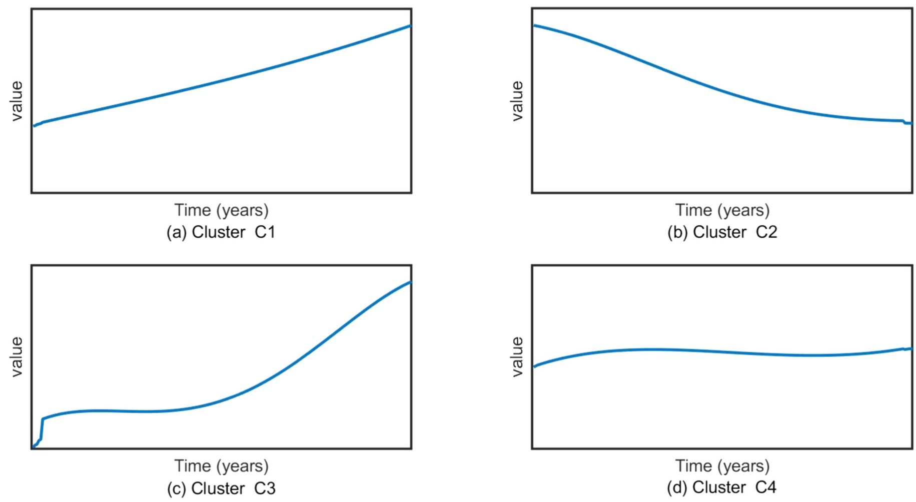


Fig. 2. Clusters identified by improved K-SC clustering algorithm. Each cluster represents a common temporal pattern of indicators.

In some countries and regions, the amount of sugar consumed is decreasing over the years. This is related to the development of GDP in countries in this region. In addition, we can see that in some countries in South America, the economy has been growing steadily as the proportion of the rural population in these areas has dropped significantly, but no significant change in sugar consumption. More in-depth research involves historical, social, and geographic factors, which are beyond the scope of our dataset, so we won't discuss it here.

3.3 Summary

Obviously, indicators are intricately interconnected with each other in both static and dynamic ways. Accordingly, we need a model combining both the static and dynamic patterns of the indicators to simulate interaction among entities and indicators better. For example, we want to know what impact of a change in one country's indicator on indicators of other countries and regions. (We will carry out the further discussion in Sections 5.9 and 5.10.)

4 PREDICTING THE DEVELOPMENT OF WORLD INDICATORS

Up to now, we have demonstrated that the world indicators are broadly correlated with each other. Thus it is possible to use some easily accessible world indicators to predict the other significant but hard-to-collect world indicators. To this end, we propose a graph-based model to integrate the world indicators and relationships between different countries and regions together, so as to make a prediction for world indicators.

In order to make a correct prediction, we first construct a network to describe the interactions between countries and regions. The node in the network represents each country, and its features are the world indicators associated with each country. The link in this network is the relationship between the two countries. There are two types of relationships: international trade and diplomatic relationships. Hence, the key problem in this task can be treated as two steps: first, we try to utilize **graph neural networks (GNNs)** to learn node representations for this multi-relational network; second, a predictor can be trained using the learned representations as features to predict the development of world indicators.

Table 2. Clustering Result of Selected Countries and Regions

Cluster	Sugar or sweet consumption	Proportion of rural population	GDP
c1	Brazil, Chile, China, Germany, Iceland, India, Indonesia, Iran, Israel, Italy, Mexico, New Zealand, Russia, South Korea, Turkey, Uganda, United Kingdom, United States, Vietnam		Brazil, Chile, Iceland, India, Italy, Kenya, Romania, South Korea, Spain, Switzerland, United Kingdom, Vietnam, Zambia
c2	Kenya	Canada, Chile, China, Iceland, Mexico, Romania, Thailand, United Kingdom	
c3	Romania, Thailand		Austria, China, Colombia, France, Germany, Indonesia, Japan, South Africa, Thailand, Turkey, Uganda, Ukraine, United States
c4	Austria, Canada, Colombia, France, Japan, Pakistan, South Africa, Spain, Switzerland, Ukraine, Zambia	Austria, Brazil, Colombia, France, Germany, India, Indonesia, Iran, Israel, Italy, Japan, Kenya, New Zealand, Pakistan, Russia, South Africa, South Korea, Spain, Switzerland, Turkey, Uganda, Ukraine, United States, Vietnam, Zambia	Canada, Iran, Israel, Mexico, New Zealand, Pakistan, Russia

4.1 The Proposed Model

To construct a heterogeneous graph to illustrate the relationships between different countries and regions, we treat each country as a node, and the bilateral relationship as edges. Considering the temporal dynamics property of international trade and diplomacy, given a dynamic heterogeneous graph $\boldsymbol{G}_t$ at time t, our purpose is to learn the graph representation $\boldsymbol{X}_t$ on the basis of current incomplete observation of $\boldsymbol{G}_t$. Our purposed model MR2vec mainly consists of four components: Variable Generator, Feature Extractor, Graph Extractor, and Time Extractor.

- **Variable Generator:** For the incomplete feature matrix $\boldsymbol{M}_t^f$ of all nodes in each time step, it extracts the position of the missing data as variable mask $\boldsymbol{M}_t^v$ and adopts linear interpolation

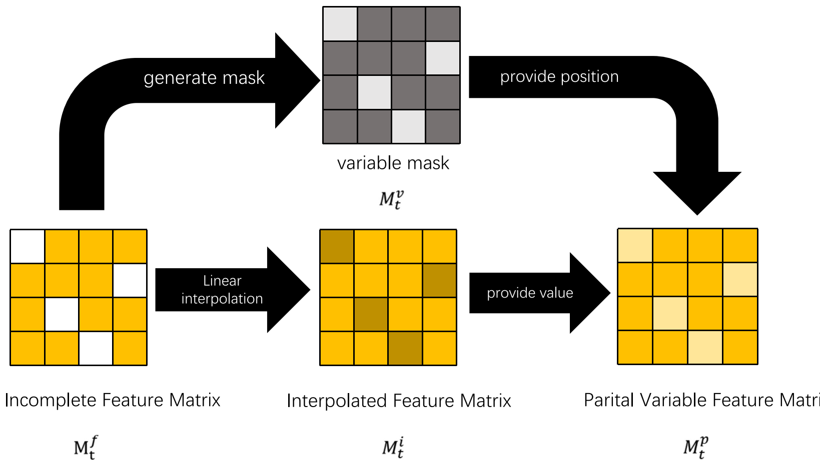


Fig. 3. Variable Generator.

to fill in all missing values as interpolated feature matrix M_t^i. The value of partial variable feature matrix M_t^p is the same as the interpolated feature matrix M_t^i and the values in the position of variable mask M_t^v are trainable. The processing flow is shown in Figure 3.

— **Feature Extractor:** It is fixed as a three layer **Multilayer Perceptron** (**MLP**) in our model. After we obtain the M_t^p for Variable Generator, the feature matrix are embedded in a higher dimension by feature extractor to an abstract representation.
— **Graph Extractor:** We further introduce a **Relational Graph Convolution Network** (**RGCN**) [33] as our Graph Extractor. When the output of feature extractor is available, it will be bound to the heterogeneous graph G_t. Then, a four layer RGCN will applied to the graph to obtain a representation for each node.
— **Time Extractor:** In order to capture time sequential information, we consider a **Gate Recurrent Unit** (**GRU**) preceding a linear layer as the Time Extractor of our purposed model. The illustration of the above three stages are shown in Figure 4.

As we have generated the low-dimensional representation for each node, it is possible and convenient to make predictions for the development of world development. In this work, we focus on two meaningful and significant prediction problems: one is predicting the temporal pattern, and the other is predicting the relationships between countries and regions.

For predicting of temporal patterns of world indicators, it can be treated as a regression problem. We directly use the representation of each node as the feature to predict the future indicators. The predicting of the relationships between countries and regions can be converted to a link prediction problem. The similarity of representing a pair of nodes could be considered the probability of the potential link's existence. In the next two sections, we also illustrate the priority of our proposed method to traditional approaches.

4.2 Relational Graph Convolutional Network with Gated Recurrent Unit

Relational Graph Convolutional Network. Our proposed model is motivated as an extension of the existing GCNs that aggregate the neighborhoods' features to update the feature of the central node for relational graph data. The relational GNN is designed for considering the diplomatic and trade relationships simultaneously. Typically, the GNNs can be understood as a message-passing

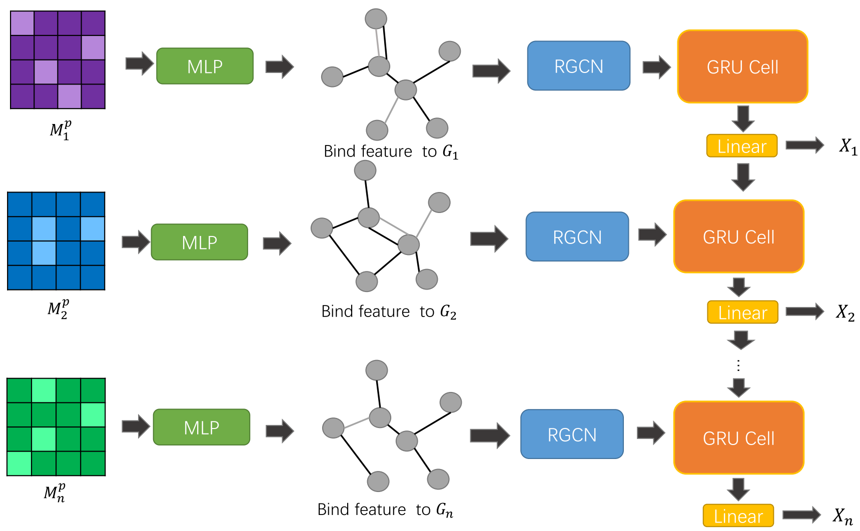


Fig. 4. Three stages extractor: Feature Extractor, Graph Extractor, and Time Extractor.

process [13]:

$$h_i^{(i+1)} = \sigma\left(\sum_{j \in N_i} f_m\left(h_i^{(l)}, h_j^{(l)}\right)\right), \tag{2}$$

where $h_i^{(l)} \in \mathbb{R}^{d^{(l)}}$ is the hidden representation of node v_i in the lth layer, and $d^{(l)}$ is the dimension of this layer. To propagate the feature information of the neighborhoods, a designed function $f(\cdot)$ is chosen and the activation function $\sigma(\cdot)$ is used for non-linear transformation. N_i denotes the neighborhood set of the central node v_i. The message passing framework has been demonstrated to be pretty effective at learning feature representations for complex data structure, especially for graph data, and has led to significant improvements in several fields such as node classification, graph classification, and link prediction [37].

To expand this framework to multi-relational graph data, we design the following simple yet effective model to learn the feature representation of each node:

$$sigmoid(x) = \frac{1}{1+e^{-x}}, \tag{3}$$

$$h_i^{(i+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in \hat{N}_{r,i}} \frac{1}{c_{i,r}} W_r^{(l)} h_j^{(l)}\right), \tag{4}$$

where W is the layer-specific and relation-specific weight matrix to perform a linear transformation as in [20]. Besides, to include the information of central node itself, we redefine the neighborhood set $\hat{N}_{r,i} = N_{r,i} \cup \{v_i\}$, which means we add a self-connection of a specific relation type to each node. More importantly, the normalization coefficient for v_i with respect to relation r is defined as $c_{i,j} = |\hat{N}_{r,i}|$. Normally, we adopt sigmoid function as $\sigma(\cdot)$ here.

Gated Recurrent Unit. Since our data are temporal and sequential, it is of great significance to consider the dynamic property for learning the final feature representation. LSTM (Long Short-term Memory RNN) [12] has been proven to be successfully applied in various sequential applications, such as machine translation, speech recognition, and natural language processing. However, due to the large computational overloads and hard parallelization, the LSTM module is usually inefficient and resource consuming. Recently, the GRU [10] has attracted more and more attention from academia and industry since it reduces the gating signals to two from the LSTM module but also achieves equivalent performance as LSTM. In view of this, we choose GRU as our network component to learn sequential information from all timestamps to yield the final network representation.

With the reset gate and update gate embedded in the GRU module, its formulations are presented as follows:

$$\begin{aligned} r_t &= \sigma(\boldsymbol{W}_r \cdot [h_{t-1}, x_t]) \\ z_t &= \sigma(\boldsymbol{W}_z \cdot [h_{t-1}, x_t]) \\ \tilde{h}_t &= tanh(\boldsymbol{W}_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \\ h_t &= \boldsymbol{W}_l \cdot ((1 - z_t) * h_{t-1} + z_t * \tilde{h}_t) + \boldsymbol{B}_l \end{aligned}, \tag{5}$$

where the $*$ represents element-wise multiplication, $[\cdot]$ represents concatenation operation, and $\sigma(\cdot)$ is the activation function. To prompt the representation ability of GRU. We add a linear layer following the GRU to make it more robust. Generally, we adopt the output of linear layer h_t from the hidden representation of the last timestamp as the final representation.

The input x_t at timestamp t is the output of the RGCN at this timestamp, as described in Equation (4). In other words, the RGCN will be adopted at each timestamp to learn the time-specific feature representation.

Regularization. A key issue is simply applying Equation (4) tends to make the model overfitting for training set due to a huge amount of parameters dealing with multi-relational data. It is demonstrated that adding regularization term is quite effective and necessary to help GNNs learn generalized representations. As a result, we introduce a regularization term to constrain the parameters using L2 norm, leading to the following regularization loss:

$$\mathcal{L}_{l2} = \sum_{r,l} \left\| \boldsymbol{W}_r^{(l)} \right\|_2. \tag{6}$$

Moreover, since different relationships are usually correlated with each other implicitly, their corresponding weights (i.e., $\boldsymbol{W}_r$ is for relationship r) can be shared with each other "softly". To this end, we further add another constraint, that is modeling the correlations explicitly using the following formula:

$$\mathcal{L}_{cor} = \sum_{r1,r2,l} \left\| \boldsymbol{W}_{r1}^{(l)} - \boldsymbol{W}_{r2}^{(l)} \right\|_2, \quad r1! = r2, \quad r2 > r1, \tag{7}$$

where $r2 > r1$ means we only consider each relation pair once regardless of the relation order. Through Equation (7) we can make the learned weights close to each other so they can be treated as shared "softly". The overall regularization loss is the sum of the Equations (6) and (7), resulting the following regularization term:

$$\mathcal{L}_{reg} = \mathcal{L}_{l2} + \mathcal{L}_{cor}. \tag{8}$$

5 EXPERIMENT AND DISCUSSION

Based on the model proposed in last section, we carry out a serial of experiments to reveal the feasibility of predicting the world indicators and inter-entities relations in this section.

5.1 Data Processing

After calculating the missing rate of each kind of world indicator, we can find out that the missing rate of Sugar and Sweet Consumption is surprisingly high compared to other indicators. According

Table 3. Overview of the Relationships Dataset

Dataset	Countries & Regions	Average Edges per Year	Start Year	End Year
National trade	193	901	1960	2016
Diplomatic	198	1078	1960	2016

to the three kinds of datasets referred to in Section 2, we preprocess them to ensure their authenticity and suitability for the input of models except our proposed one. Data items for each country of severe loss are discarded. Others are filled to the same start year and end year by fitting them individually with a polynomial to make the input consistent. The details of processed relations are displayed in Table 3. Moreover, the raw data which are incomplete are directly used as the input of our model. For each year t, we then build a heterogeneous graph G_t according to Section 4. The inter-country relationships are considered the two types of edges between nodes. The regularized indicators of each country are fixed as the feature vector for the node. Considering the specialty of GCN and RGCN, we add a self-loop for each node to the original graphs if the model contains them.

5.2 Experiment Setup

To guarantee the integrity of the heterogeneous graphs, we focus on the years between 1995 and 2004 with 100 randomly chosen countries and regions as our experiment dataset. The first eight years are regarded as training data, the 9th year as the validating data, and the last year as testing one. For the relational prediction task, we randomly sample 100 existent and 100 nonexistent edges for each edge type as the input for the models to assure the invariability of the input size and the labels for them are directly fixed as 1 and 0. The input for the models in the feature prediction task is the feature vectors of all countries and regions.

5.3 Comparison Methods

The following models are adopted as baselines including attribute-considering baselines, homogeneous GNN baselines, heterogeneous GNN baselines, and **knowledge graph embedding** (**KGE**) baselines.

The **attribute-considering baselines** include:

- **CN (Common Neighbor) [30]:** A conventional method used for link prediction. The number of the same neighbors of two nodes shares determines the connectivity of them.
- **KI (Katz Index) [17]:** It measures the similarity of the two nodes by summing up one-hop to k-hops connectivity of them with an attenuation.
- **Logistic Regression:** It is a generalized linear model used for two class classification by gradient descending.
- **Linear Regression:** It is a linear approach to model the relationship between a scalar response and vector or a scalar by minmizing the error rate.
- **Bayesian Ridge Regression:** It estimates a probabilistic model of the non-linear regression problem. Its parameters are set as follows: $\alpha_1 = 10^{-6}$, $\alpha_2 = 10^{-6}$, $\lambda_1 = 10^{-6}$, $\lambda_2 = 10^{-6}$, $iteration = 300$. α is the shape parameter for Γ prior distribution, while λ is the reciprocal of the scale parameter for it. The Γ prior distribution is the intial distribution of Bayesian Ridge Regression.
- **GRU [10]:** It is an RNN based model with gates deciding how much information to forget or remember. For each iteration, it will both learn from the a portion of the current data and the old state of itself to update its current state. We set its hidden size as 32.

Table 4. Parameter Settings for Relational Prediction Task

Model	Train epoch	Embedding dimension	Learning rate	Optimizer
Logistic Regression	100	64	0.0001	L-BFGS [40]
KGE	400	32	0.001	Adam [19]
DeepWalk	5	32	0.025	SGD
Others	100	32	0.001	Adam

The **homogeneous GNN baselines** include:

- **DeepWalk [31]:** It is a well-known baseline for the embedding model. It adopt random walk and skip-gram model to generate representation for each node. We set number and length for each walk as 5 and 20, respectively, and the window size as 5.
- **GCN [20]:** It is composed of a list of GCN layers, which aims at aggregating the attributes from neighbor nodes get a lower dimension of representations. We set the number of layers 2, and the hidden layers of size as [32].

The **heterogeneous GNN baselines** include:

- **RGCN [33]:** It could fuse the structure of different kinds of edge types to get a lower dimension embedding by stacking GCN layers for each edge type. For each layer, it will sum up the vector of different edge types. We set the number of layers and hidden size the same as GCN.

The **KGE baselines** include:

- **TransE [5]:** It is a representative translational distance model that represents the entity and relations as a lower dimension vector in the same semantic space. In terms of computation, it assumes the sum of vector of the head entities and the relations is equal to the vector of tail entities.
- **TransR [23]:** Based on TransE, it represents the entities and relations into different semantic spaces. A projection matrix M is employed to project the vector of relations into the space of entities.
- **Analogy [24]:** It builds the embeddings in complex field for both entities and relations. And a score function is adopted to evaluate whether the relation exists or not by computing the embeddings of entities and relations.

Different models require different parameters to yield a relatively better result. For the national prediction task, empirically, we fix parameters according to Table 4 to obtain a better performance. It is worth noting that embedding dimension for Logistic Regression is the concatenation of the endpoints' feature vectors of the edge. On the basis of Section 4, we adopt Binary cross-entropy loss with l_2 punishment as our loss function. Considering the particularity that KGE models could only build embedding for the edges present in the training data and it has a loss function, we train them on the same edge type as the one to predict and use its loss function. Homogeneous GNN baselines are trained on one kind of edge type at a time for their homogeneity. For example, DeepWalk (trade) means the input of a DeepWalk model is the trade relationships. In this experiment, we also present other combination models such as GCN+GRU to validate the combination method's effectiveness.

The various task also demands diverse parameters to achieve a relative optimal result. Reasonable parameters setting for feature prediction task are presented in Table 5. According to Section 4, we adopt Soft l_1 loss with l_2 punishment as our loss function. Squaring up the fact that the KGE+GRU model cannot be trained synchronously for the difference between loss functions of

Table 5. Parameter Settings for Feature Prediction Task

Model	Train epoch	Embedding dimension	Learning rate	Optimizer
Bayesian Ridge Regression	300	10	0.001	SGD
Linear Regression	1	10	\	\
DeepWalk	5	32	0.025	SGD
Others	200	32	0.001	Adam

Table 6. Test Result of Predicting Trade and Diplomatic

Model	Trade AUC	Diplomatic AUC
Common Neighbor	0.463 (±0.008)	0.498 (±0.011)
Kalz Index	0.500 (±0.003)	0.498 (±0.005)
Logistic Regression	0.539 (±0.008)	0.562 (±0.006)
TransE	0.901 (±0.036)	0.917 (±0.034)
TransR	0.860 (±0.033)	0.924 (±0.037)
Analogy	0.979 (±0.013)	0.977 (±0.014)
DeepWalk (trade)	0.883 (±0.039)	0.966 (±0.016)
DeepWalk (diplomatic)	0.871 (±0.045)	0.870 (±0.038)
GCN (trade)	0.854 (±0.040)	0.925 (±0.039)
GCN (diplomatic)	0.757 (±0.060)	0803 (±0.050)
RGCN	0.911 (±0.026)	0.862 (±0.062)
GRU	0.897 (±0.047)	0.976 (±0.013)
GCN (trade)+GRU	0.915 (±0.025)	0.978 (±0.009)
GCN (diplomatic)+GRU	0.914 (±0.032)	0.974 (±0.015)
MR2vec	**0.995** (±0.005)	**0.992** (±0.007)

two models, we first train the KGE model for the epochs mentioned in the relational prediction task to gain a well-trained embedding and then train the GRU on this embedding for 200 epochs to solve the incompatible of the loss function.

5.4 Experiment Results

Relational Prediction. The ROC_AUC are employed as the evaluation metrics. The result is listed in Table 6.

$$ROC\_AUC = \frac{1}{m^+m^-}\sum_{x^+\in D^+}\sum_{x^-\in D^-}(W(f(x^+) > f(x^-))), \tag{9}$$

where D^+ and D^- are sets containing all the positive and negative samples. In this experiment, D^+ contain all the existent trade or diplomatic relationships. Analogically, D^- covers all the nonexistent relationships. $f(x)$ is the prediction from sample x. $W(x)$ is a logic function, which turns to 1 when x is true and turns 0 on the contrary. m^+ and m^- are the numbers of positive and negative samples. Specifically, the number of existent and nonexistent trade or diplomatic relationships.

Table 6 shows a comparison between predictions of trade and diplomatic relationships. Overall, most score obtain high AUC scores over 0.9, indicating that the bilateral relationships among countries and regions are predictable to a large degree. More observations are represented below:

- Traditional topology-based methods, such as CN and KI, perform poorly on the link prediction task. Compared to other methods, they can only learn from a topology view ignoring the features of each node. It may result in an unprecedented ill score in this task.
- On the contrary, the Logistic Regression method cannot utilize the topology structure of the graph. It is normally used to classify vectors spread in Euclidean space other than with a network structure, so its result is intuitively poor.
- Analogy is a highly robust model in these two tasks. It has perfect performance better than most models, but still inferior to our proposed one. The performance of our model is far better than others, which may mainly be caused by two aspects: First, Our model considers the missing data as a variable rather than a constant value compared to others.
- One thing worth noting that DeepWalk trained on national trade relationships is much more robust than that trained on diplomatic relationships. The performance of the former one is better than that of the latter one in the relationship prediction task. The same phenomenon also happens in GCN. It can be inferred that entities' relationships are closely linked together, inferring that it is possible to predict the hard-to-obtain relationships with readily accessible ones. However, it did not show up during the training process of our model, showing the ability of our model to extract relationships strongly.
- As can be seen, simple GRU performs quite well in both prediction tasks. Moreover, if we use the embedding of GCN as the input for GRU, it becomes a more robust model than simple GRU and could correctly reflect the entities' relationships. Moreover, the result of diplomatic relation prediction is better than that of national trade relations, indicating that these models may not be good at predicting national trade relationships compared to the diplomatic ones. But overall, intern-entities relations are predictable for our model.

Feature Prediction. In addition to the relationships between countries and regions, it is equally important to predict each country's indicators. We adopt the error rate to measure how bad the performance is in this experiment. The higher the error rate, the worse the model performs. The error rate is the absolute value of the difference between the prediction data $f(x)$ and the real data y. To be more specified, it is the absolute value of the difference between the actual entities' attributes and the prediction from the model.

$$error\_rate = |y - f(x)|. \tag{10}$$

As can be seen from Tabel 7, the error rates of most models are around 2.5%, suggesting that the world indicators can be predicted to a certain extent. Further observations are listed as follows:

- As shown in the table above, Bayesian Ridge Regression is obviously a light and robust model to reflect the future status of countries and regions. It could achieve a very low error rate close to that of our model at around 2.4 percent. Because GCN and RGCN are static models, their performances are relatively worse than those of other models.
- In most cases, the combination of models will substantially improve performance. Closer inspection of the table shows that, when TransE is combined with GRU, its performance drops a bit. Furthermore, its error rate raises 2%, implying that the embedding of TransE may not suit the GRU input.

Table 7. Test Result of Prediction Error Rate for Country Features

Model	Test error rate (%)
Linear Regression	5.30 (±0.00)
Bayes Ridge Regression	2.42 (±0.00)
GCN (trade)	15.03 (±0.03)
GCN (diplomatic)	15.20 (±0.01)
RGCN	15.00 (±0.02)
GRU	3.13 (±0.02)
TransE (trade)+GRU	5.04 (±2.94)
TransE (diplomatic)+GRU	2.36 (±0.04)
TransR (trade)+GRU	2.24 (±0.01)
TransR (diplomatic)+GRU	2.25 (±0.02)
DeepWalk (trade)+GRU	2.50 (±0.04)
DeepWalk (diplomatic)+GRU	2.52 (±0.12)
GCN (trade)+GRU	2.45 (±0.03)
GCN (diplomatic)+GRU	2.38 (±0.04)
MR2vec	**1.53 (**±0.03**)**

— Our MR2vec model outperforms all baselines and almost 30% better than the second one (TransR+GRU). Moreover, it has a comparably small variance, demonstrating the robustness of our model.

5.5 Predictability of World Indicators (Q3)

In our collected dataset, different kind of world indicator has different missing rate. Some indicators are hard to collect such as the sugar or sweet consumption while other indicators like proportion of rural population are more accessible. To explore the relationship between various world indicators and whether the hard to collect one could be predicted by easily accessible ones. We set up a new experiment to explore which world indicators are easier to be predicted by MR2vec.

In this experiment, we set the raw indicators except the one to study as the feature matrix and the one being studied as the label of the regression. For example, if we study the case of sugar consumption, we delete it from the origin feature matrix and fix it as the target output of MR2vec. The learning rate and training epoch are empirically set to 0.0001 and 50. We did each experiment three times to get the mean and standard deviation values.

The results of the experiment are listed in Table 8. There are several conclusions that can be inducted from the table:

— The overall average error rates are apparently higher than the one when predicting all the 10 indicators with all of them as input, inferring that the past of an indicator plays a significant role in predicting the future of itself.

Table 8. Test Result of Data Analysis

Indicator	Test error rate (%)
Proportion of rural population	6.51 (±0.66)
Population growth rate	9.37 (±0.01)
Proportion of population aged 0-14	12.11 (±0.02)
Life expectancy at birth	10.26 (±5.10)
Proportion of females survive to 65 years old	4.50 (±0.29)
Fertility rate	5.08 (±0.01)
CO2 emission	6.38 (±0.00)
Government health expenditure	7.92 (±0.30)
GDP	8.35 (±0.61)
Sugar or sweet consumption	7.54 (±0.03)
Average	7.80 (±0.73)

— The proportion of the population aged 0–14 has the maximum error rate, indicating that it is hard to predict with other indicators.
— For the hard to collect indicator: Sugar or sweet consumption, its error rate is around the average level. It can be said that this indicator can be predicted by other indicators to some degree.

5.6 Parameter Sensitivity (Q4)

We investigate the parameter sensitivity in this section. Individually, we evaluate how the different numbers of the RGCN layer β in the graph extractor of MR2vec influence the results. In order to make the result distinguishable, the model is trained for 50 epochs on a feature prediction task with various β chosen from [1,2,3,4,5]. Figure 5 shows the curves of the error rate of training and validating in each epoch for different β. The curves in the figure of different β are generally monotonic.

It is worth noting that the curve rebounds at about 15 epochs after a period of decline. But eventually, they converge at the error rate of around 0.2. Moreover, the error rate of validation is slightly below that of the train, indicating each update has a more obvious effect on the validation set than on the training set. We can infer that the world indicators do have some hidden laws and follow some imperceptible rules.

As can be seen, no matter how the β changes, the performance of the model changes very little. Overall, our model is not very sensitive to this parameter.

5.7 Ablation Study

From an intuitive point of view, the world indicators is a time series topology changing data, which needs to be considered from the following four parts:

— Features of each node;
— The relationships between different nodes;

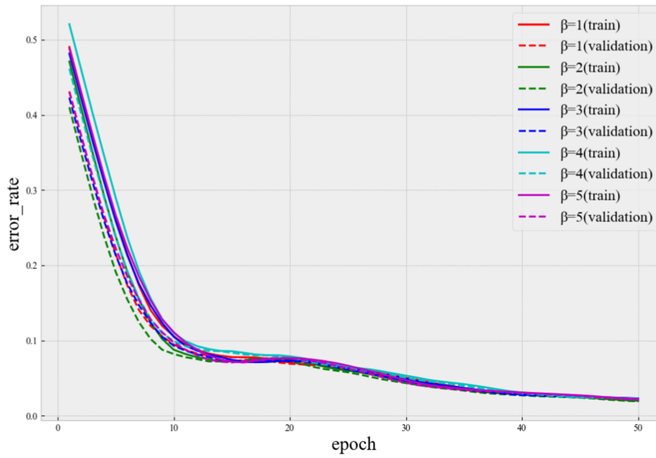


Fig. 5. Parameter sensitivity.

Table 9. Ablation Study Results of MR2vec

Variable Generator	✓	×	✓	✓	✓
Feature Extractor	✓	✓	×	✓	✓
Graph Extractor	✓	✓	✓	×	✓
Time Extractor	✓	✓	✓	✓	×
Test Error Rate (%)	2.63 (±0.14)	2.84 (±0.04)	2.70 (±0.12)	2.75 (±0.11)	3.04 (±0.43)

— The changes of the features of each node;
— The changes in relationships between nodes.

There are four components in our model, and each component has a different function.

— Variable Generator is responsible for the processing of missing data to remain its illegibility;
— Feature Extractor is used for extracting the feature of nodes;
— Graph Extractor is responsible for learning the topology of the nodes;
— Time Extractor remembers the changes of the feature and connection of the nodes.

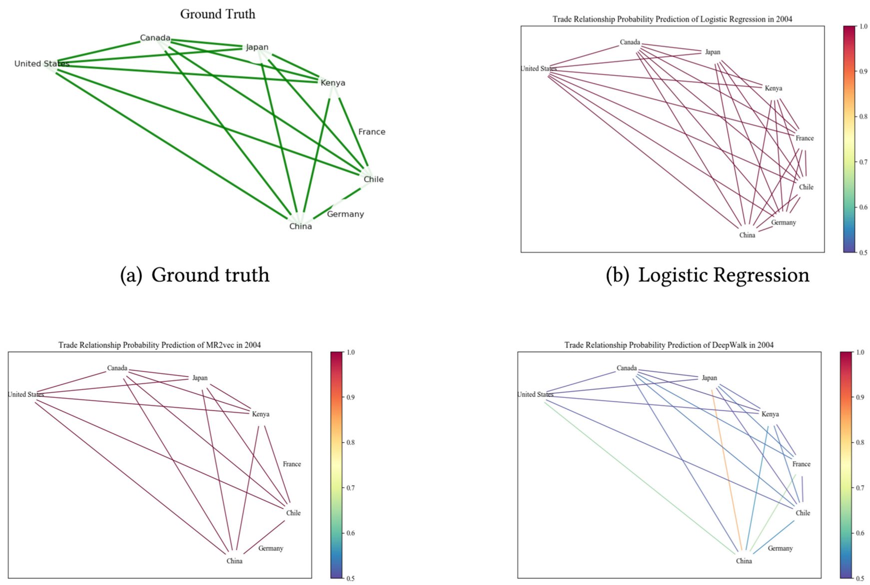
Although these four parts are very intuitive, it cannot prove that they all contributed to the performance of the model.

Accordingly, we perform an ablation study on our model. The configuration is fixed as followed. We select 158 entities and regions whose features were relatively complete and employ the data from 1995 to 2002 as the training set, the 2003 one as the validation set, and the 2004 one as the test set. The result is represented in Table 9. Each component is essential for our model MR2vec.

5.8 Relation Prediction by Different Models

Now, we present a experiment to illustrate the effectiveness of our model specifically. Figure 6 provides an example generated from three chosen models on diplomatic prediction tasks, including Logistic Regression, DeepWalk, and MR2vec. To make the figure more intuitive, we randomly select eight countries and regions to visualize.

To cover more countries and regions, we choose 155 countries and regions that outnumber the 100 countries and regions mentioned in the relational prediction experiment. The ground truth is



(a) Ground truth (b) Logistic Regression

(c) MR2vec (d) DeepWalk

Fig. 6. Ground truth and model's prediction for diplomatic relationships among countries and regions in 2002.

represented in green lines in the figure. In other subfigures, the links between countries and regions are prediction probabilities for diplomatic relationships among countries and regions. In order to make the result more reliable, we test each model for 10 times and get the average probabilities as the expected probabilities of each model. We consider a probability threshold of 0.5 to make the figure more succinct. Only the edges whose average probabilities above the threshold are visible in the graph. Moreover, the color bar is fixed to the same range to make it easier to observe. The redder the edge is, the higher the probability is. Some observations are listed as follows:

— It is apparent from the figure that Logistic Regression predicts a lot of nonexistent edges. Consequently, it is a relatively weak model in this task.
— Better than Logistic Regression, DeepWalk does not have confidence in non-existent edges. However, the cost is that for almost all edges, the confidence is not so firm.
— The result of MR2vec is closest to the ground truth compared to Logistic Regression and DeepWalk. All edges in ground truth have been predicted, and the reliability is quite high. At the same time, the edges that do not exist in ground truth are also distinguished.

Overall, our model MR2vec outperforms the Logistic Regression and DeepWalk in this study. Its prediction is almost the same as the ground truth, indicating its robustness.

5.9 Interaction among Countries and Regions

We have found some interesting conclusions about the potential interaction between different countries and regions' various features. We increase a particular indicator of a country by a certain ratio ζ and observe this change's effect on other countries and regions' indicators.

For example, we increased the United States' "Government health expenditure", "GDP", and "Fertility rate" by 10 percent separately and compared the new prediction for other countries and regions before this change. This operation takes the form as Equation (12), where X_i is the

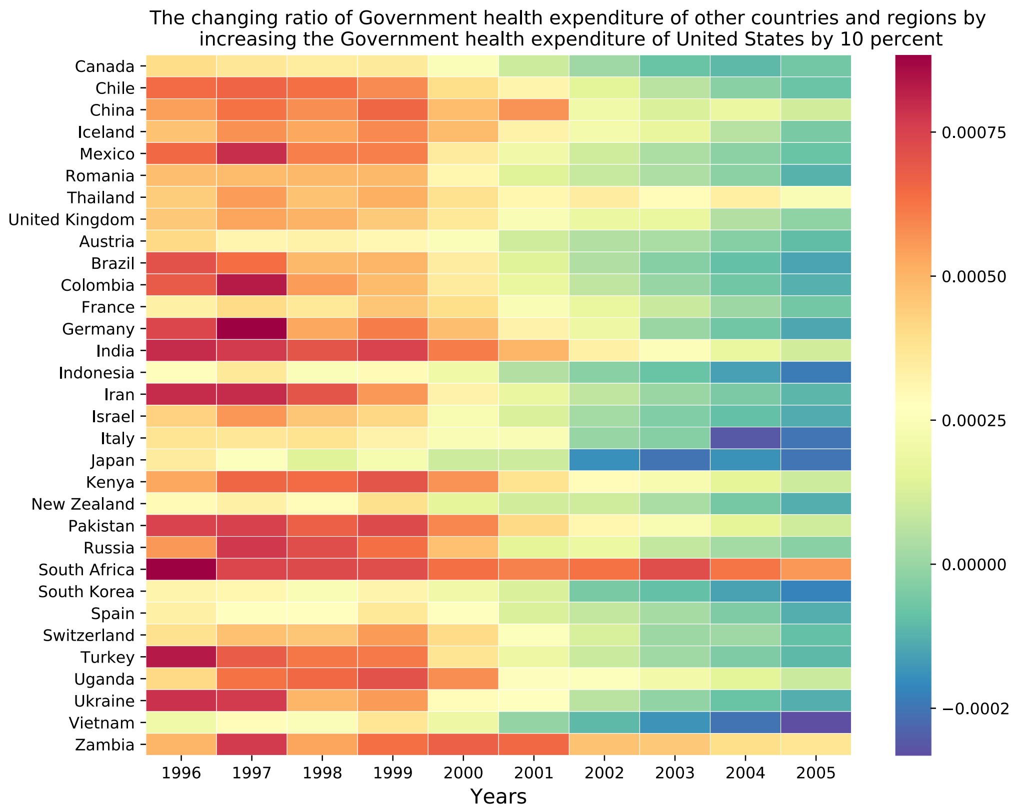


Fig. 7. The changing ratio of Government expenditure of other countries and regions by increasing the Government expenditure of United States by 10 percent.

augmented attribute of the data, specifically, “Government health expenditure”, “GDP”, and “Fertility rate” of United State here and ζ is hyper parameter that control this transformation. To measure the fluctuate of the output. We propose a metric called “changing ratio” or C, which could be defined according to Equation (12). The Y_{pred} and Y^{ζ}_{pred} are the prediction of MR2vec from the original data X and multiplied data X^{ζ}. In this study, we adopt the ζ as 10.

$$X_i^{\zeta} = \frac{100 + \zeta}{100} X_i, \tag{11}$$

$$C = \frac{Y^{\zeta}_{pred} - Y_{pred}}{Y_{pred}}. \tag{12}$$

We choose the countries and regions mentioned in Table 2. The result of “Government health expenditure” and “Fertility rate” are shown in Figures 7, 8, and 9. A significant conclusions is apparent. The change in health expenditure of the government is subtle compared to that of fertility rate and GDP. One cause may be that the health expenditure is up to the entities’ income, government fiscal capacity, demographic structure, disease pattern, and so on [18]. Other countries and regions can hardly influence all these factors as for the GDP and fertility rate, the more stable the international environment, the higher those indicators.

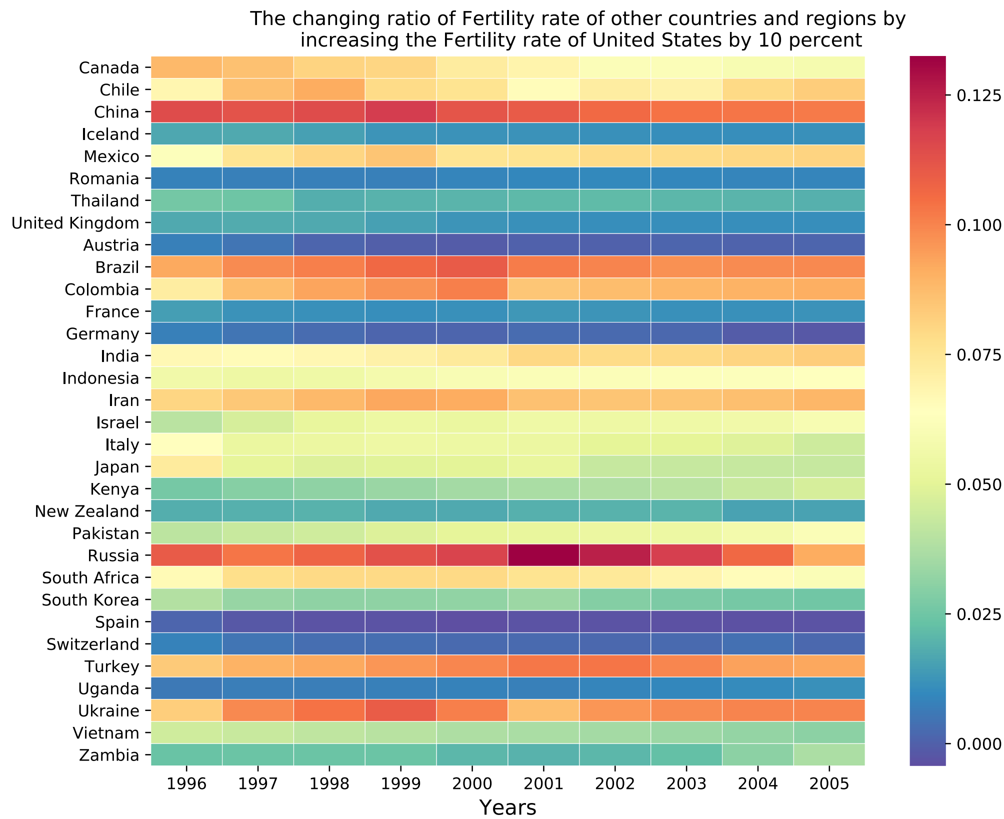


Fig. 8. The changing ratio of Fertility rate of other countries and regions by increasing the Fertility rate of United States by 10 percent.

5.10 Interaction among Indicators

We have also found some consistent conclusions about the potential interaction between different indicators to Section 3. We increase the "Sugar or sweet consumption" indicator of every country by 10 percent and observe the average effect (changing ratio according to Equation (12) of this adjustment on other indicators each year.

As can be seen from Figure 10. Different indicators are potentially connected. We could find several interesting conclusions:

— The "Sugar or sweet consumption" has a strong correlation with "CO2 emission". One possible explanation could be the positive correlation between sugar consumption and energy consumption. And a higher energy consumption often means a higher quantity of CO2 emission [29].
— It associates with the rising of GDP for most developing countries such as China and Thailand, similar to the conclusion of Ismail, Amid I. and Tanzer, Jason M. and Dingle, Jennifer L. [16]. It may result from the change of diet due to the increase in the economy. Take China in the 1990s, for example. The rapid economic growth [7] has boosted China's massive demand for sugar.
— The increasing demands for sugar associate with a decrease in the "Fertility rate". It may result from the potential negative effect of sugar and artificial sweeteners on the fertility [34].

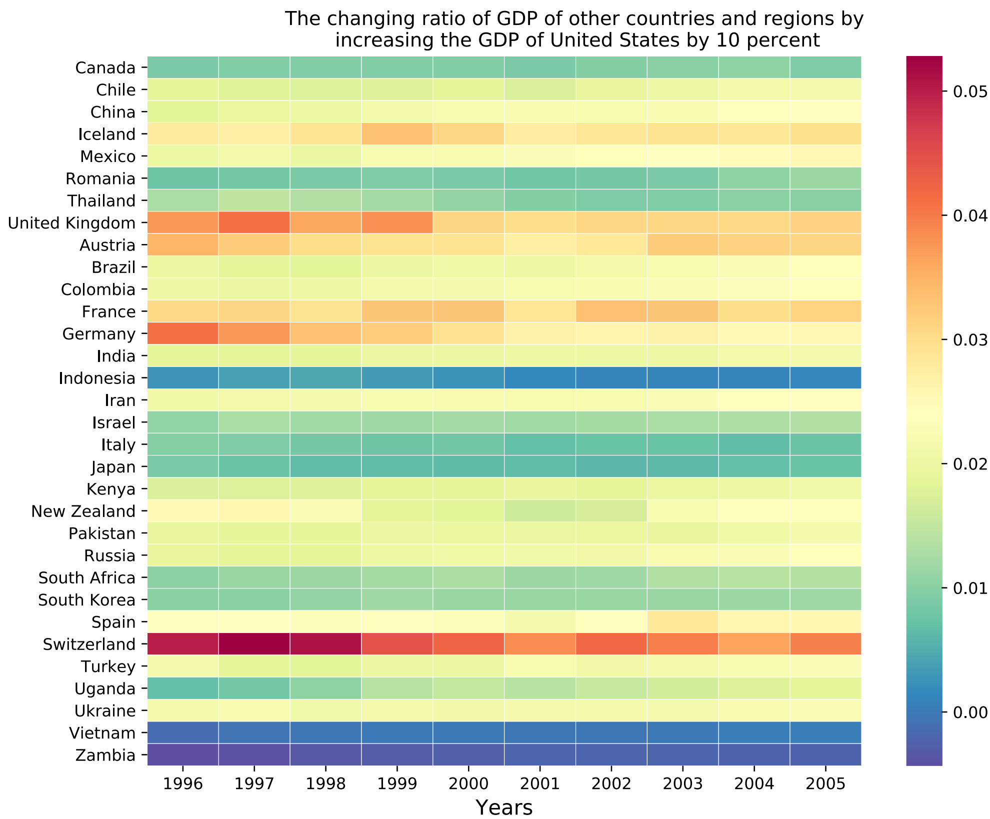


Fig. 9. The changing ratio of GDP of other countries and regions by increasing the GDP of United States by 10 percent.

5.11 Wide Adaptability of MR2vec (Q5)

Although our method achieves good results on the world indicators and intern-entities relationships, our model's scalability has not been validated. Therefore, we further verify it on a larger-scale transportation dataset, accurately, part of TLC Trip Record Data [8]. We collect the records of Yellow taxi (Yellow) trips, and **For-Hire Vehicle** (**FHV**) trips for NYC city from January 2019 to June 2019. Considering repeated trips, the average number of trips per month for them is 7 million and 5 million, as represented in Table 10. Similar to the graphs built in the last experiment, we build a heterogeneous graph for each month in this experiment. For each graph, the nodes and edges are taxi zones and two kinds of trips among them. Considering the particularity that the attributes of pull up and drop off locations will not change during time, the attributes of each node are static, including the Borough, Zone, and service zone. We normalized these attributes as feature vectors for nodes.

We compare baselines mentioned in the relational prediction experiment with our proposed models on the task of trip existence prediction. Specifically, the task is to predict whether there exists a trip record between the zones. To gain a better result, we train each model for 200 epochs with a learning rate of 0.001 except for DeepWalk. And DeepWalk is trained for five epochs with a learning rate of 0.025. Similar to the relational prediction task, for each epoch, we randomly sample 100 existent links and 100 nonexistent links as input. and employ the AUC score to measure

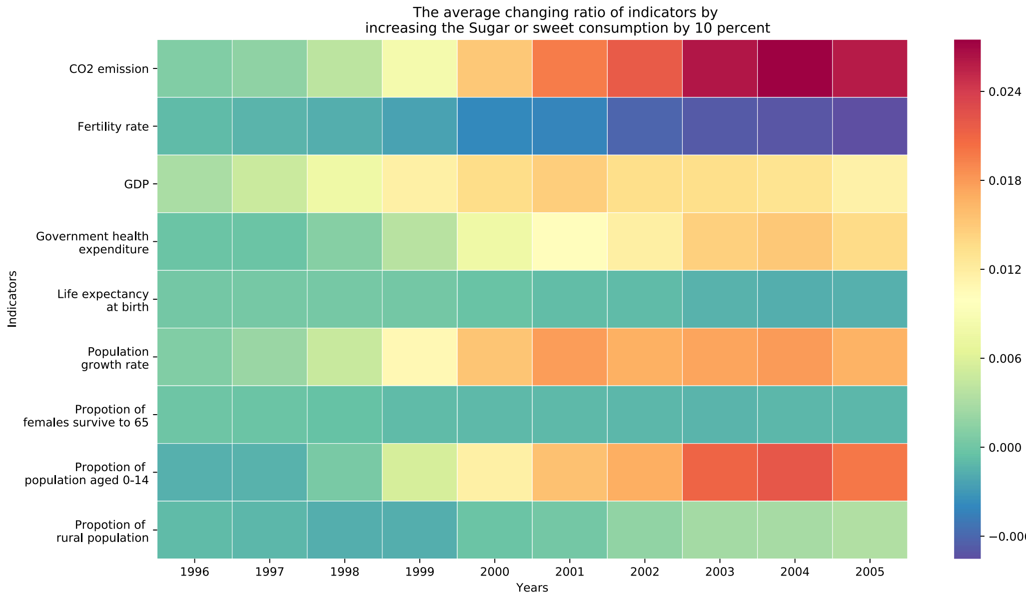


Fig. 10. The average changing ratio of indicators by increasing the sugar or sweet consumption by 10 percent.

Table 10. Overview of the Part of TLC Trip Record Data

Dataset	Number of Zones	Average Trips per Month	Start Time	End Time
Yellow Taxi	265	7M	2019 Jan	2019 June
For-Hire Vehicle	265	5M	2019 Jan	2019 June

the performance. We adopt the first four months for training, the fifth month for validating, and the last month for testing. The results are listed in Table 11, It is evident that our model has an outstanding performance, and could obtain almost the highest AUC score in both tasks. From the results above, it can be inferred that our proposed model has robust scalability in predicting a wide range of relationships.

6 RELATED WORK

In this article, we mainly study the development patterns of world indicators and the relationships between countries and regions in the world with various models including GNN models and KGE models.

6.1 Works on World Indicators

The analysis of world indicators data has always been the focus of the research society, and the World Bank and other organizations regularly publish world indicators to the public. They will show the development of each indicator and make predictions for the future [1]. For example, the results of the Sally Engle Merry study show that world indicators are rapidly becoming tools for assessing and promoting various social justice and reform strategies around the world [26]. However, they mainly focused on separate indicators and seldom considered them from a network view.

Table 11. Test Result of Predicting Yellow Taxi Trips and For-Hire Vehicle Trips

Model	Yellow AUC	FHV AUC
Logistic Regression	0.506 (±0.008)	0.396 (±0.013)
TransE	0.983 (±0.011)	0.916 (±0.119)
TransR	0.977 (±0.014)	0.611 (±0.286)
Analogy	**0.993** (±0.004)	0.822 (±0.029)
DeepWalk (Yellow)	**0.991** (±0.005)	0.870 (±0.138)
DeepWalk (FHV)	0.560 (±0.060)	0.815 (±0.232)
GCN (Yellow)	0.838 (±0.058)	0.678 (±0.253)
GCN (FHV)	0.503 (±0.046)	0.612 (±0.101)
RGCN	0.585 (±0.309)	0.654 (±0.293)
GCN (Yellow)+GRU	**0.992** (±0.006)	0.943 (±0.026)
GCN (FHV)+GRU	**0.991** (±0.039)	0.958 (±0.026)
MR2vec	**0.995** (±0.005)	**0.984** (±0.035)

Through the excavation of the temporal patterns, plenty of research work involves various aspects. After the temporal patterns have been discovered, researchers have done a lot of work, such as predicting the popularity of news [35], discovering topic intensity flow [21], and so on. Xiaodi Du et al. proposed a new trajectory mining algorithm to find the migration patterns in the financial market and solve the immigration patterns in the stock market problem [11]. Manish Gupta et al. found communities through time-pattern mining [15]. Jaewon Yang and Jure Leskovec analyzed the temporal pattern of the evolution of online content in online social media over time by defining a clustering algorithm [38]. Poon, Leonard KM adopt three clustering algorithms to analyze world indicators in WDR dataset [32]. Different from our model, it aims at analyzing and cluster the indicators rather than predict them. Inspired by these efforts, we hope to find some temporal patterns to represent the development patterns of world indicators.

However, the few of the articles mentioned above consider the world indicators and bilateral relations from the perspective of time series heterogeneous graphs. The time series heterogeneous graphs can better express the correlation between world indicators and bilateral relations. So we innovatively consider them from the perspective of time series heterogeneous graphs.

6.2 Graph Neural Network

Word embedding is a collective term for mapping of words from one hot vector space to a dense vector space. Word2vec is also a kind of word embedding and its goal is to learn the word vector according to the co-occurrence information between words. The random walk-based network embedding method draws on the Word2vec idea, which is represented by Deepwalk [31] and Node2vec [14]. While GCN [20] originated from CNN [22]. Network embedding can represent nodes in a complex network in a low-latitude vector space while preserving the node's structural information [9]. To apply GCN to a heterogeneous graph, RGCN [33] learns structures in different edge type and represent each node at a lower dimension.

The essence of network embedding is to find a nonlinear function to transform the raw network into a low-dimensional latent space and use the network structure and property to constraint the model [36]. For example, methods such as node2vec [14], and so on, consider the network as a graph, use a random walk from a node to generate sequence data similar to text, and then use the skip gram to train the node as a "word" to obtain a "word vector" [27]. GCN [20] aggregates the features of neighbor nodes and using linear transformation to reduce dimension each layer. Based on this, RGCN [33] apply GCN to each kind of edge type and average the aggregations as the embedding. These methods model the network on the basis of a static network, and then used for subsequent machine learning or data mining tasks, and have achieved very efficient performance.

Based on the work of RGCN [33], we proposed the model MR2vec. However, different from the traditional GNN, our model is a time-sequential network. It could extract a correlation in the time axis. Moreover, it is also an end2end model, which reduces the manual pre-processing of the input data.

6.3 Knowledge Graph Embedding

However, there is limited work to solve the problem of multi-relational network representation. TransE [5] proposes a simple and effective algorithm to solve the problem of multi-relational data processing. Inspired by word2vec [27], it uses the translation-invariant phenomenon of word vectors. Thinking of relation in each triple instance (head, relation, tail) as a translation from the head entity to tail entity. However, it is more reasonable to map the entities and relations into different spaces. So TransR [23] was brought out. Besides the translation model, the bilinear model is also an essential method for KGE. Analogy proposes a complex space embedding for both entities and relations [24]. By maximize a score function, it could learn a robust embedding from different datasets. However, the TransE, TransR as well as Analogy models are mainly applied to the static knowledge graph. And it requires a series of preprocessing to enable itself to process dynamic data. On the contrary, our model MR2vec is a time-sequential model that can process the time-sequential heterogeneous graphs easily.

7 CONCLUSIONS

In this article, we focus on the dynamic world indicators and multiple relationships between countries and regions. Firstly, we find out the temporal patterns of world development from a dynamic view and examine the correlation among those indicators from a static view. In addition, taking into account the trade and diplomatic relationship between countries and regions, we propose a model called MR2vec to represent the world indicators. This model considers the fusion of multi-relations and experimental results show that our proposed method outperforms most of the baseline methods. Furthermore, we study the parameter sensitivity, ablation study, special cases, and scalability of it. They validate that the world indicators are predictable and verify the road adaptability and superiority of our model.

REFERENCES

[1] The World Bank. 2000. *World Development Indicators 2000.* Oxford University Press.
[2] Katherine Barbieri, Omar Keshk, and Brian Pollins. 2008. Correlates of War Project Trade Data Set Codebook. Retrieved 17 November, 2020 from https://correlatesofwar.org/data-sets/bilateral-trade.
[3] Katherine Barbieri, Omar M. G. Keshk, and Brian M. Pollins. 2009. Trading data: Evaluating our assumptions and coding rules. *Conflict Management and Peace Science* 26, 5 (2009), 471–491.

[4] Reşat Bayer. 2006. Diplomatic Exchange Data Set, v2006. 1. Retrieved 17 November, 2021 from http://correlatesofwar.org.

[5] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems*. 2787–2795.

[6] John R. Carter. 2007. An empirical note on economic freedom and income inequality. *Public Choice* 130, 1–2 (2007), 163–177.

[7] Tanya Clark. 1995. Emerging-market indicators the Economist. *Oct. l4* 28 (1995).

[8] NYC Taxi&Limousine Commission. 2019. TLC Trip Record Data. Retrieved 17 November, 2020

[9] Peng Cui, Xiao Wang, Jian Pei, and Wenwu Zhu. 2018. A survey on network embedding. *IEEE Transactions on Knowledge and Data Engineering* 31, 5 (2018), 833–852.

[10] Rahul Dey and Fathi M. Salemt. 2017. Gate-variants of gated recurrent unit (GRU) neural networks. In *Proceedings of the 2017 IEEE 60th International Midwest Symposium on Circuits and Systems*. IEEE, 1597–1600.

[11] Xiaoxi Du, Ruoming Jin, Liang Ding, Victor E. Lee, and John H. Thornton Jr. 2009. Migration motif: A spatial-temporal pattern mining approach for financial markets. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1135–1144.

[12] Felix A. Gers, Jürgen Schmidhuber, and Fred Cummins. 1999. Learning to forget: Continual prediction with LSTM. In *Proceedings of the 9th International Conference on Artificial Neural Networks.*

[13] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70, JMLR. org, 1263–1272.

[14] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 855–864.

[15] Manish Gupta, Jing Gao, Yizhou Sun, and Jiawei Han. 2012. Community trend outlier detection using soft temporal pattern mining. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 692–708.

[16] Amid I. Ismail, Jason M. Tanzer, and Jennifer L. Dingle. 1997. Current trends of sugar consumption in developing societies. *Community Dentistry and Oral Epidemiology* 25, 6 (1997), 438–443.

[17] Leo Katz. 1953. A new status index derived from sociometric analysis. *Psychometrika* 18, 1 (1953), 39–43.

[18] Xu Ke, Priyanka Saksena, and Alberto Holly. 2011. The determinants of health expenditure: A country-level panel data analysis. Retrieved on 17 November, 2021 from https://www.who.int/health_financing/documents/report_en_11_deter-he.pdf.

[19] Diederik Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. CoRR

[20] Thomas N. Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations.*

[21] Andreas Krause, Jure Leskovec, and Carlos Guestrin. 2006. Data association for topic intensity tracking. In *Proceedings of the 23rd International Conference on Machine Learning*. ACM, 497–504.

[22] Alex Krizhevsky, I. Sutskever, and G. Hinton. 2012. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 25, 2 (2012), 1097–1105.

[23] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence.*

[24] Hanxiao Liu, Yuexin Wu, and Yiming Yang. 2017. Analogical inference for multi-relational embeddings. In *Proceedings of the 34th International Conference on Machine Learning*. Vol. 70, JMLR. org, 2168–2178.

[25] Sally Engle Merry. 2018. Measuring the world: Indicators, human rights, and global governance. In *The Palgrave Handbook of Indicators in Global Governance*. D. Malito, G. Umbach, and N. Bhuta (Eds.), Palgrave Macmillan, 477–501.

[26] Sally Engle Merry and John M. Conley. 2011. Measuring the world: Indicators, human rights, and global governance. *Current Anthropology* 52, S3 (2011), 000–000.

[27] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proceedings of the International Conference on Neural Information Processing Systems*. Vol. 26.

[28] Leann Myers and Maria J. Sirois. 2004. Spearman correlation coefficients, differences between. In *Encyclopedia of Statistical Sciences*. 12. Wiley Online Library.

[29] Shuwen Niu, Yongxia Ding, Yunzhu Niu, Yixin Li, and Guanghua Luo. 2011. Economic growth, energy conservation and emissions reduction: A comparative analysis based on panel data for 8 Asian-Pacific countries. *Energy Policy* 39, 4 (2011), 2121–2131. Retrieved from https://doi.org/10.1016/j.enpol.2011.02.003.

[30] Fragkiskos Papadopoulos, Rodrigo Aldecoa, and Dmitri Krioukov. 2015. Network geometry inference using common neighbors. *Physical Review E* 92, 2 (2015), 22807–22807.

[31] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 701–710.

[32] Leonard K. M. Poon. 2017. Clustering with multidimensional mixture models: Analysis on world development indicators. In *Proceedings of the International Symposium on Neural Networks*. F. Cong, A. Leung, and Q. Wei (Eds.), Springer, 153–160.

[33] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne vanden Berg, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *Proceedings of the European Semantic Web Conference*.

[34] Amanda Souza Setti, Daniela Paes de Almeida Ferreira Braga, Gabriela Halpern, Rita de Cássia S. Figueira, Assumpto Iaconelli, and Edson Borges. 2017. Is there an association between artificial sweetener consumption and assisted reproduction outcomes. *Reproductive Biomedicine Online* 36, 2 (2017), 145–153.

[35] Gabor Szabo and Bernardo A. Huberman. 2010. Predicting the popularity of online content. *Communications of the ACM* 53, 8 (2010), 80–88.

[36] Daixin Wang, Peng Cui, and Wenwu Zhu. 2016. Structural deep network embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1225–1234.

[37] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S. Yu Philip. 2020. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems* 32, 1 (2020), 4–24.

[38] Jaewon Yang and Jure Leskovec. 2011. Patterns of temporal variation in online media. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. ACM, 177–186.

[39] Lijing Yang and Brian McCall. 2014. World education finance policies and higher education access: A statistical analysis of world development indicators for 86 countries. *International Journal of Educational Development* 35, 1 (2014), 25–36.

[40] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. 1997. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software* 23, 4 (1997), 550–560.

Received July 2020; revised March 2021; accepted September 2021