Physica A 436 (2015) 188–199

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Prediction in complex systems: The case of the international trade network

Alexandre Vidmer [a,*], An Zeng [a,b], Matúš Medo [a], Yi-Cheng Zhang [a]

[a] Department of Physics, University of Fribourg, Fribourg CH-1700, Switzerland

[b] School of Systems Science, Beijing Normal University, Beijing 100875, PR China

HIGHLIGHTS

- The international trade data are represented by a country-product network.
- Diffusion on a network is used to predict the future links.
- Different ways to alter the original diffusion scheme are proposed.
- Diffusion performs best when coupled with a time-aware metric for product similarity.
- The investigated methods can be applied to any network with time information.

ARTICLE INFO

Article history:
Received 16 January 2015
Received in revised form 2 April 2015
Available online 12 May 2015

Keywords:
Link prediction
Complex networks
Node similarity
Economic system
Recommender system

ABSTRACT

Predicting the future evolution of complex systems is one of the main challenges in complexity science. Based on a current snapshot of a network, link prediction algorithms aim to predict its future evolution. We apply here link prediction algorithms to data on the international trade between countries. This data can be represented as a complex network where links connect countries with the products that they export. Link prediction techniques based on heat and mass diffusion processes are employed to obtain predictions for products exported in the future. These baseline predictions are improved using a recent metric of country fitness and product similarity. The overall best results are achieved with a newly developed metric of product similarity which takes advantage of causality in the network evolution.

1. Introduction

In the past few years, the international trade network has attracted the attention of researchers from various fields, and especially from the complex networks scientists. The international trade was studied for the first time under the network framework in Ref. [1], using a blockmodel consisting in partitioning countries together according to their trade flows, military interventions, diplomatic exchanges, and conjoint treaty memberships. The complex network approach was used in Ref. [2] and showed that the international trade exhibits common features with the World Wide Web network. It has been shown that many features of the international trade can be explained using models based on the gravity equation [3–5]. Recently, not only the countries, but also their exports have been analyzed under the complex network framework. The Product Space attempts to explain how the nations develop by projecting their exports on a 2D map, and observing how

* Corresponding author. Tel.: +41 026 300 9156.
E-mail address: alexandre.vidmer@unifr.ch (A. Vidmer).

http://dx.doi.org/10.1016/j.physa.2015.05.057
0378-4371/

they diffuse in the Product Space [6]. The Economic Complexity aims to rank products by the technological requirements needed for a country to be able to manufacture a product, and to rank the countries by their competitiveness [7,8].

The prediction of quantity and price of exports in the international trade has been studied using various models [9] and additional information such as geographical distance between countries and common language [10]. By contrast, we employ here a recommendation approach that is usually applied to e-commerce systems [11,12] in order to predict what an individual country will export in the future. The prediction of which products a country will add to its export basket can help to understand how countries grow. The countries' future exports are particularly complex to predict, as their evolution depends on many external factors, such as geographical position, diplomatic relations between countries, available natural resources, and technologies. Nevertheless, previous studies [6,13] showed that it is possible to measure the competitiveness of a country, estimate its future growth and even predict its future exports using solely the international trade data. The last aspect, personalized prediction of future exports for each country, further suffers from the lack of a traditional approach with conventional metrics and renowned prediction algorithms.

Recommender systems were created to filter the relevant information in information systems [14,15]. For instance, an algorithm can help a person to choose which movie to watch by creating a list of most relevant movies that this particular person might enjoy, whereas it is a tremendous task to find a movie of interest among the thousands of existing ones. Recommendation using temporal and causal effects has recently attracted attention [16,17]. In this paper, we use a network representation for countries and their exports. A network is made of nodes connected by links. Nodes represent countries and products, and a link connects a country to a product if the country exports the product. We call this type of network *bipartite*, as the nodes are formed by two disjoint sets: countries and products. This allows us to treat the problem within a *recommender system* framework [12]. The approach that we use slightly differs from a link prediction approach because we aim to predict the future exported products for each country instead of predicting the most likely future exports for the network as a whole.

We use a recent recommender algorithm based on diffusion [18] as a tool for prediction. However, as demonstrated in Fig. 1, the international trade data fundamentally differs from online systems data on which recommendation is typically done by the absence of preferential attachment [19]. This makes it impossible to predict the future popularity of a product from its current popularity and the predictions thus have correspondingly lower accuracy. To improve the recommendation accuracy, we adopt a temporal approach to devise a new definition of mutual distance between products, and couple it with the diffusion method in order to enhance the prediction of the links. This approach is closely related to the proximity of products [6], defined as a distance between each couple of products depending on the number of co-occurrences in countries' export baskets. Based on this distance, it was shown that the closer a product is to one of the products the country is currently exporting, the more likely it is to be added to the country's export basket in the future. The recommendation of movies to users was improved by the use of co-occurrences in Ref. [16]. In Ref. [17], the authors studied the temporal dependence of movies ratings and used it to improve the prediction of future ratings.

In order to improve the prediction performance of the diffusion method, we also use Economic Complexity as defined in Ref. [13]. The first definition of Economic Complexity was made by the use of two self-consistent linear equations [20,21], which were then successfully applied to predict the long-term growth of countries' export basket. A more recent definition was given as a set of two self-consistent non-linear equations [13] that reflect the non-linear relations between complexity of products and competitiveness of countries. This non-linear definition of complexity was shown to capture more information than the linear one on the countries' hidden growth potential in a toy model of countries' exports [22]. The use of the complexity of products in the prediction process allows us to improve the diversity of the prediction, while maintaining the same level of accuracy. In contrast with the mutual distances introduced previously, the Economic Complexity theories aim to assign each country and product an individual score on an absolute scale.

2. Methods

2.1. Data

We use the NBER-UN dataset which was described and cleaned in Ref. [23]. We cleaned it further by removing aggregate categories and keeping only the countries for which complete mutual export data are available. Products having zero total export volume for a given year while having nonzero total export volume for the previous and the following years were removed from the dataset. Products and countries with no entries after year 1993 were removed as well. After the cleaning procedure, the network consists of 65 countries and 770 products. To decide if we consider country i to be an exporter of product α or not, we use the Revealed Comparative Advantage (RCA) [24] which is defined as

$$\mathrm{RCA}_{i\alpha} = \frac{e_{i\alpha}}{\sum_{\beta} e_{j\beta}} \bigg/ \frac{\sum_{j} e_{j\alpha}}{\sum_{j\beta} e_{j\beta}}, \tag{1}$$

where $e_{i\alpha}$ is the volume of product α that country i exports in thousands of US dollars. RCA characterizes the relative importance of a given export volume of a product by a country in comparison with this product's exports by all other countries. We use a bipartite network representation with two different kinds of nodes, one for countries and one for