Eur. Phys. J. Special Topics **212**, 115–120 (2012)

DOI: 10.1140/epjst/e2012-01657-4

The European Physical Journal Special Topics

Regular Article

Competitors' communities and taxonomy of products according to export fluxes

M. Cristelli[1,3], A. Tacchella[1,2], A. Gabrielli[2,4], L. Pietronero[2]
A. Scala[3,4], and G. Caldarelli[5,3,4]

[1] "Sapienza", Università di Roma, Dip. Fisica, P.le A. Moro 2, 00185 Roma, Italy
[2] ISC-CNR, Via dei Taurini 19, 00185 Roma, Italy
[3] ISC-CNR, Dip. Fisica"Sapienza", Università di Roma, P.le A. Moro 2, 00185 Roma, Italy
[4] London Institute for Mathematical Sciences, South Street 35a, Mayfair London, UK
[5] IMT, Institute for Advanced Studies, Piazza S. Ponziano, 6, 55100 Lucca, Italy

Received 31 May 2012 / Received in final form 23 July 2012
Published online 17 September 2012

Abstract. In this paper we use Complex Network Theory to quantitatively characterize and synthetically describe the complexity of trade between nations. In particular, we focus our attention on export fluxes. Starting from the bipartite countries-products network defined by export fluxes, we define two complementary graphs projecting the original network on countries and products respectively. We define, in both cases, a distance matrix amongst countries and products. Specifically, two countries are similar if they export similar products. This relationship can be quantified by building the Minimum Spanning Tree and the Minimum Spanning Forest from the distance matrices for products and countries. Through this simple and scalable method we are also able to carry out a community analysis. It is not gone unnoticed that in this way we can produce an effective categorization for products providing several advantages with respect to traditional classifications of COMTRADE [1]. Finally, the forests of countries allows for the detection of competitors' community and for the analysis of the evolution of these communities.

1 Introduction

In the last two decades networks have become one of the main mathematical tools for the description of complex systems. In fact the graph theory permits for quantitative analysis, synthetic description, extraction of information and powerful visualization of a wide range of problems belonging to fields ranging from biological to social systems [2–4]. In this paper we show how the framework of networks can provide effective methods to address the problem of economic complexity.

This paper is in the spirit of a recently economical theory proposed in [5,6]. Authors of such papers propose to interpret the trade fluxes (in the form of a bipartite countries-products network) as the projection of a tripartite network where the intermediate level, linking countries to products, is the key level where the information about the complexity of the productive system of a country is stored. In other words the authors of [5] say that a country can produce a product if it owns all the capabilities (i.e. the element of this intermediate layer) required. On the other hand, it is very hard, if not impossible, to define a universal measure of this intermediate level. This is due to the fact that all capabilities required for the production of a product are the result of a complex interaction of several different aspects of a single country. For instance, education level, primary energetic resources, but also welfare level, efficency of burocracy and technological transfer can be considered as capabilities of a country. It is therefore impossible to determine all the capabilities owned by a country and it is even more difficult to try to quantitatively characterize them. Unfortunately, even if this tripartite graph is substantially a non-measurable object, it nevertheless plays a crucial conceptual role. In order to solve this dilemma, a proper characterization of the observed bipartite graph is needed. These measures become a proxy for the complexity of the productive system of a country in the sense that exported products inform about the capabilities owned by a specific country.

In this paper we discuss how to introduce a distance among countries on the basis of their mutual economical competition. This quantity can be reasonably defined starting from the measure of the similarity of the export of two countries as set by the overlap of the two export baskets which gives the level of mutual competition [14]. Furthermore a similar metrics can be introduced for products and this allows for the definition of a dynamical categorization of products. By *dynamical* we mean that we propose a method to group products which does not require periodic revisions as standard classification but is rather bottom-up defined from data. Actually, standard product classifications are usually top-down defined while the one here proposed is based on a metric criterion. As an example, our method is able to catch that nowadays colza seeds (mostly used for biofuels) should be grouped with machineries and cars rather than with the group of food as indicated by standard classification.

The paper is organized as it follows. In Sect. 2 we discuss the criterion to build the bipartite network countries-products starting from the raw export fluxes expressed in US Dollars for each country and product. In Sect. 3 we study the features of the two possible projections of the bipartite network and we introduce an new classification for products based on a Minimal Spanning Tree deriving from a metric criterion for the similarity of products. In Sect. 4 we draw conclusions and perspectives.

2 Definition of the binary matrix countries-products

The bipartite network countries-products is mathematically described considering a $N_c \times N_p$ matrix M, where N_c and N_p are the number of countries and products respectively. The generic element M_{cp} of this matrix is 1 if the country c is able to export the product p, and 0 otherwise. We point out that the two descriptions, i.e. the bipartite network and the export matrix, are completely equivalent, a link between the country c and the product p correspond to $M_{cp} = 1$ and *vice versa*. In order to determine if a country is an exporter of a product given the exported amount expressed in US Dollars we introduce the so-called Relative Comparative Advantage (RCA) [8]. The RCA compares the share of a product in the country's export basket with respect to the share of that product in the global economy, in formula RCA reads as

$$RCA_{cp} = \frac{\frac{W_{cp}}{\sum_{p'} W_{cp'}}}{\frac{\sum_{c'} W_{c'p}}{\sum_{c',p'} W_{c'p'}}} \tag{1}$$

where W_{cp} is the amount of the product p exported by the country c expressed in US Dollars. We say that a country has a Relative Comparative Advantage in the export of the product p when the RCA_{cp} is larger than a fixed threshold. Following previous works [6], we fix this threshold to be 1, in that way, if the ratio is larger than 1 it means that the share of the product p for the economy of the country c is higher than the average share of this product in the world economy, vice-versa if it is smaller than 1. Thus the elements of the binary matrix M, and therefore the link of the bipartite network countries-products too, result to be defined as

$$M_{cp} = \begin{cases} 1 \text{ if} RCA_{cp} \geq 1 \\ \\ 0 \text{ if} RCA_{cp} < 1. \end{cases} \tag{2}$$

3 Projection of the bipartite network

Given the M matrix (or equivalently the bipartite network countries-products) we define two complementary graphs corresponding to the projection of the original bipartite network on the country nodes and on the products nodes. In that way the projected graphs with respectively N_c and N_p nodes are homogeneous with respect to the two different types of nodes. The easiest way to perform this projection is to consider the following two matrix products

$$\begin{aligned} C &= MM^T \\ P &= M^T M \end{aligned} \tag{3}$$

where M^T is the transposed matrix and the square matrices C and P define the country-country network and the product-product network. The element $C_{cc'}$ defines the weight associated to the link between contries c and c' in the country-country network. Analogously $P_{pp'}$ gives the weight of the link between products p and p' in the product-product network. These weights have an interesting interpretation: if we write explicitly the expression of a generic element of the C matrix according to Eq. (3), we have that $C_{cc'} = \sum_p M_{cp} M_{c'p}$. Therefore the element $C_{cc'}$ (since M_{cp} is a binary unweighted matrix), counts the number of products exported by both countries c and c'. In a similar way the the element $P_{pp'}$ counts the number of countries which export both the products p and p'. The diagonal elements C_{cc} and P_{pp} are respectively the number of products exported by the country c and the number of exporters of the product p respectively.

Let us now analyze these two networks. Standard analysis of the matrix elements will allow to characterize the competition among countries and to define a self-consistent taxonomy for products starting from data.

3.1 A new geography: Similarity in countries' productive systems

The non-diagonal elements of the square matrix C inform about the level of competition between a pair of countries, because (as already remarked), the elements $C_{cc'}$ give the number of products that both countries export. Obviously we assume that countries which export the same products are likely to be competitors with respect to that product. When a close analysis is made on the strongest competition an unexpected geographical correlation appears. In practice this corresponds to perform a community analysis of the countries-countries network. Given the fact that

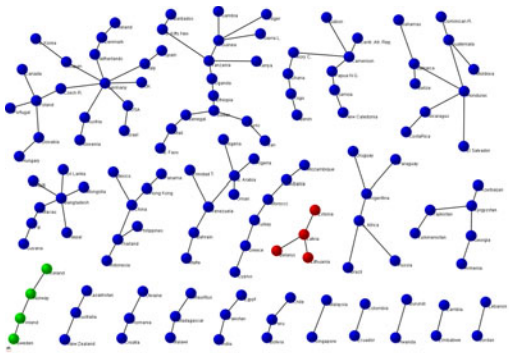

Fig. 1. A minimal spanning forest of the competition matrix C. A cluster of red (color online) countries indicates former Soviet Republics, while the green (color on line) one on bottom left, indicates Scandinavian countries.

the countries-countries network is a weighted network we can easily define a distance matrix $\bar{C}$ suitably normalizing the elements of C as it follows

$$\bar{C}_{cc'} = 1 - 2\frac{C_{cc'}}{C_{cc} + C_{c'c'}} \tag{4}$$

so that $\bar{C}_{cc'}$ ranges from 0 to 1 and small values of $\bar{C}_{cc'}$, that is *neighboring* countries according to this metrics, correspond to a large overlap of the export baskets as wanted, while large values of $\bar{C}_{cc'}$ correspond to couples of countries with a very different export basket. See [10] for similar approaches to define the distance between nodes of a network and [6] for a similar approach to define distance among countries even if the distance proposed in [5] appears to be rather artificial.

The easiest approach to visualize the most similar nodes is to build a Minimal Spanning Tree (MST hereafter) as is often done for stock correlation prices [9–11]. In such a way we are able to hierarchically connect countries according to their level of competition. The MST is constructed by considering the ordered list of all the $N(N-1)/2$ possible distances $\bar{C}_{cc'}$ between two countries according to their increasing distance $\bar{C}_{cc'}$. The first edge of the MST is the edge linking the pair of countries with the lowest distance $\bar{C}_{cc'}$ (i.e. the largest overlap of export basket), and then we sequentially scroll the above-defined list by keeping all the edges with the only condition of forbidding loops. Each branch of the MST can be seen as community of very close competitors where the distance is defined from an economical point of view (not geographical). In order to better detect and track the evolution of these communities we would like to define a disjointed forest of sub-trees. This can be done generalizing the method to build the MST by adding a further condition to the MST algorithm: an edge is also forbidden if it links vertices that are both already drawn (i.e. already connected to another vertex). In that way we build a set of disconnected trees (i.e. a forest) embedded in the MST. We call this forest as the Minimal Spanning Forest (MSF hereafter) and it allows for the detection of communities and neighboring countries according to the competition criterion by the visual inspection of the sub-trees.

In Fig. 1 we show the MSF of the competing countries for year 2000.

The most striking evidence from visual inspection is that many competing communities are formed by countries geographically nearby. This evidence contrasts the canonical vision of Economics according to which neighboring countries should specialize their productive systems on different products or sectors in order to avoid competition. In fact standard theory predicts that geographically far countries tend to be similar from the export point of view since transportation costs would prevent

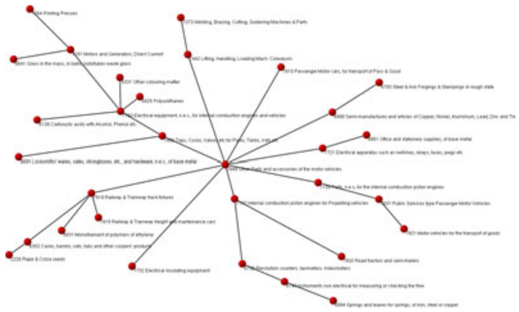

Fig. 2. A minimal spanning forest of the product similarity matrix P.

competition between these countries in the countries close to them. Actually, the data evidence points to an opposite situation, where neighboring countries tend to have similar productive systems. Even though this is not the most rational choice, this is nevertheless a general feature of economics systems [12], where banks [13] or countries [14] trade preferentially with similar partners, thereby affecting the whole robustness of the system [15–17]. Under this respect, the MSF is a tool allowing for the systematic studies of the evolution of the competition among countries.

3.2 A similarity criterion for a new categorization of products

Similarly to the countries-countries network, we define a distance matrix $\bar{P}$ whose expression in term of the square matrix P reads as

$$\bar{P}_{pp'} = 1 - 2\frac{P_{pp'}}{P_{pp} + P_{p'p'}}. \quad (5)$$

Following the same prescription discussed for the countries-countries network we can define the minimum spanning tree and the minimum spanning forest in order to obtain a hierarchical categorization of products. In particular we can interpret the disjointed subtrees as the highest level at which products can be grouped, that is the lowest level of resolution at which products are classified according to this method. In Fig. 2 we report the largest tree obtained in the MSF for the products community matrix. We point out that this community analysis for products can be a powerful categorization for products providing several advantages with respect to the top-down grouping procedures which are usually adopted as in COMTRADE, when import/export fluxes are recorded. Indeed, differently from standard method the categorization here proposed takes naturally into account changes in world economy and to self-update the groups of products according to the changes of their relative similarity. In this respect, we propose a self-adaptable categorization which is able to grasp the dynamics of product communities and does not require periodic revisions to incorporate those changes as in the case of official categorization.

4 Conclusions and perspectives

In this paper we have shown how the framework of networks can be effectively used to define a proper metrics both in the space of countries and in the space of products. In particular, while countries can be aggregated on the basis of their competition, products can be associated by their co-occurence in countries economies.

We believe that the product taxonomy as discussed in this paper can be a starting point for a new categorization for products. The method proposed here clusters products according to a similarity criterion. In particular, a pair of products are more and more similar, if several countries produce both. A typical case is represented by the example shown in Fig. 2. When many countries which export cars and machineries, export colza seeds too, it is likely that colza seeds may play a role for this industrial sector. A traditional top-down classification would instead put colza in the category of foods. That is not wrong, since colza seeds, before the introduction of bio-fuel, was mainly a food product. Nevertheless, in the present time, colza seeds have become an industrial product rather than a food. Indeed, our clustering method does not require any periodic revision of the classification since it considers automatically changes and evolution in space of the product similarity. In addition, the changes of the minimal spanning tree and forest over time can be tracked back. In such a way we can obtain a dynamical categorization of products which is self-adaptable and self-consistently defined with respect to the evolution of the world economy. In this sense we propose a dynamic and self-adaptable categorization for products.

On the other hand the community analysis deriving from the minimal spanning tree and forest of countries permits for grouping the countries according to their mutual level of competition, in fact two countries are close to each other if they have export basket with a large overlap. As in the case of products, year by year we are able to detect the emergence of new economical communities and forecast which are and will be the main competitors of a given country tracking the changes of the tree and the forest of nations.

We thank EU FET Open project FOC nr.255987 for support.

References

1. World Trade Flows: 1962-2000, `www.nber.org/data`, and `http://cid.econ.ucdavis.edu/data/undata/undata.html`
2. G. Caldarelli, *Scale-Free Networks* (Oxford University Press, Oxford 2007)
3. F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, Science **325**, 422 (2009)
4. N. Johnson, T. Lux, Nature **469**, 302 (2011)
5. C.A. Hidalgo, B. Klinger, A.-L. Barabási, R. Hausmann, Science **317**, 482 (2007)
6. C.A. Hidalgo, R. Hausmann, Proc. Nation. Acad. Sci. (USA) **106**, 10570 (2009)
7. G. Caldarelli, M. Cristelli, A. Gabrielli, L. Pietronero, A. Scala, A. Tacchella, `http://arxiv.org/abs/1108.2590`
8. B. Balassa, *Trade Liberalization and Revealed Comparative Advantage*, Manchester School **33**, 99123 (1965)
9. R.N. Mantegna, Eur. Phys. J. B **11**, 193 (1999)
10. G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, Phys. Rev. E **68**, 046130 (2003)
11. G. Bonanno, G. Caldarelli, F. Lillo, S. Micciché, N. Vandewalle, R.N. Mantegna, Eur. Phys. J. B **38**, 363371 (2004)
12. S. Battiston, D. Delli Gatti, M. Gallegati, B. Greenwald, J.E. Stiglitz, J. Econo. Dyn. Cont. **31** 20612084 (2007)
13. G. De Masi, G. Iori, G. Caldarelli, Phys. Rev. E **74**, 66112 (2006)
14. G. Caldarelli, M. Cristelli, A. Gabrielli, L. Pietronero, A. Tacchella, `http://arxiv.org/abs/1108.2590`
15. K.-M. Lee, J.-S. Yang, G. Kim, J. Lee, K.-I. Goh, I.-M. Kim, PLoS ONE **6**, e18443 (2011)
16. W. Chen, R.M. D'Souza, Phys. Rev. Lett. **106**, 115701 (2011)
17. G. D'Agostino, A. Scala, V. Zlatić, G. Caldarelli, `http://arxiv.org/abs/1105.3574`