Community structure in the World Trade Network based on communicability distances

Paolo Bartesaghi[1] · **Gian Paolo Clemente[2]** · **Rosanna Grassi[1]**

Abstract

In this paper, we investigate the mesoscale structure of the World Trade Network. In this framework, a specific role is assumed by short- and long-range interactions, and hence by any suitably defined network-based distance between countries. Therefore, we identify clusters through a new procedure that exploits Estrada communicability distance and the vibrational communicability distance, which turn out to be particularly suitable for catching the inner structure of the economic network. The proposed methodology aims at finding the distance threshold that maximizes a specific quality function defined for general metric spaces. Main advantages regard the computational efficiency of the procedure as well as the possibility to inspect intercluster and intracluster properties of the resulting communities. The numerical analysis highlights peculiar relationships between countries and provides a rich set of information that can hardly be achieved within alternative clustering approaches.

✉ Paolo Bartesaghi
  paolo.bartesaghi@unimib.it

  Gian Paolo Clemente
  gianpaolo.clemente@unicatt.it

  Rosanna Grassi
  rosanna.grassi@unimib.it

1 Department of Statistics and Quantitative Methods, University of Milano - Bicocca, Via Bicocca degli Arcimboldi, 8, 20126 Milan, Italy

2 Department of Mathematics for Economic, Financial and Actuarial Sciences, Università Cattolica del Sacro Cuore, Milan, Italy

⌖ Springer

1 Introduction

International trade is based on a set of complex relationships between different countries. Both connections between countries and bilateral trade flows can be modelled as a dense network of interrelated and interconnected agents. A long-standing problem in this field is the detection of communities, namely subset of nodes among which the interactions are stronger than average. Indeed, the community structure of a network reveals how it is internally organized, highlighting the presence of special relationships between nodes, that might not be revealed by direct empirical analyses.

In this framework, a specific role is assumed by the distance between nodes. Indeed, the neighbours of a given node are immediately connected to such a node and they can affect its status most directly. Nonetheless, more distant nodes can influence this node while passing through intermediary ones. In the economic field, a network perspective is actually based on the idea that indirect trade relationships may be important (see, for example, Fagiolo et al. 2015). For instance, Abeysinghe and Forbes (2005) explain the impact of shocks on a given country by indirect trade links. Based on a global VaR approach, Dées and Saint-Guilhem (2011) show that countries that do not trade (very much) with the USA are largely influenced by its dominance over other trade partners linked with the USA via indirect spillovers. In Ward et al. (2013), the bilateral trade is assumed not independent of the production, consumption and trading decisions made by firms and consumers in third countries. A measure of the distance between nodes that also considers indirect connections is therefore crucial to catch deep interconnections between nodes. In this work, we will focus on two measures of distance or metrics on the network: the Estrada communicability distance (Estrada and Hatano 2009) and the vibrational communicability distance (Estrada and Hatano 2010a). They both go beyond the limits of the immediate interaction between neighbours, and they look simultaneously, albeit differently, at all the possible channels of interactions between nodes. The nearest two nodes are in each metric, the stronger is their interaction or, in other words, the higher is the level of communicability between them.

With this paper, we contribute to the literature by proposing a specific methodology that exploits such metrics to inspect the mesoscale structure of the network, in search for strongly interacting clusters of nodes. Indeed, our purpose is twofold. We reveal hidden relationships between nodes due to non-immediate connections and long-range interactions, and we show how this approach turns out to be particularly suitable when applied to a dense network like the World Trade Network (WTN). More specifically, we exploit communicability and vibrational communicability metrics to group nodes whose mutual distances are below a given threshold, i.e. whose interactions are stronger than a given value. Then, we identify the optimal partition according to a maximum quality function criterion. It is well-known that classical modularity is a way to measure if a specific mesoscopic description of the network in terms of communities is more or less accurate. But, unlike the Girvan–Newman approach (Newman and Girvan 2004), we will refer to the partition quality index proposed in Chang et al. (2016) for general metric spaces. In this way, we can exploit the additional information contained in the metric structure of the network. Among all the different partitions we get at different thresholds, we select the one providing the maximum quality index,

according to the criterion described in Chang et al. (2016). Our proposal is very efficient from a computational viewpoint. Indeed, given the specific distance matrix, the optimal solution can be easily evaluated varying the threshold. We cluster nodes going beyond the interactions between neighbours and considering all possible channels of interaction between them. We allow for a degree of flexibility by introducing a threshold. Varying the threshold, it is possible to depart from the optimal solution so that only the strongest (or the weakest) channels of communications emerge.

The paper is organized as follows. After a short review of the literature in Sect. 2, main preliminaries and the definitions of the communicability functions are revised in Sect. 3. These functions lead to two important metrics on networks, which are described in Sect. 4. Section 5 contains the description of the proposed methodology, which is also tested on a suitable toy model. In Sect. 6, we apply our methodology to the World Trade Network. In particular, main characteristics of the network are described in Sect. 6.1. The steps of the methodology are summarized in Sect. 6.2. We report in Sect. 6.3 main results based on communicability and resistance distance, respectively. We show how the proposed methodology is able in capturing key economic clusters as well as in providing additional insights into intracluster and intercluster characteristics and of countries' relevance both in the community and in the whole network. Conclusions follow. Technical details are left in "Appendices A and B".

2 Literature review

Community detection is an important topic in the analysis of the topological structure of complex systems. Its importance has grown over time in light of the remarkable progress in the description of large networks, together with the development of new powerful data analysis tools (Lancichinetti and Fortunato 2009). These advances have made it possible to extend the field of applicability of the theory not only to networks of enormous dimensions but also to weighted networks and direct networks (Fagiolo 2007; Rattigan et al. 2007; Clemente et al. 2018; Cerqueti et al. 2018). Various methods and algorithms to detect communities on networks have been studied. Some methods are algorithm-based, such as methods based on hierarchical clustering or edge removal (Newman 2004). Other methods are based on the optimization of specific criteria over all possible network partitions. In this context, it is well-known optimization of a modularity function according to Newman's definition (Newman and Girvan 2004). An exhaustive review about methods and algorithms can be found in Fortunato (2010) and Fortunato and Hric (2016). Some authors proposed to detect communities by means of a quality measure called surprise (Traag et al. 2015; Nicolini et al. 2017). Inspired by this literature, recently Van Lidth de Jeude et al. (2019) deal with detection of general mesoscale structures, such as core-periphery structures.

More recently, the role of non-local interactions between nodes has been highlighted, that is, interactions that do not exclusively involve the immediate neighbours of a given node. In particular, results connected to the idea of communicability introduced by Estrada in 2004 have proved to be extremely effective (Estrada and Rodriguez-Velazquez 2005; Estrada and Hatano 2008, 2009; Estrada 2012a). All the more so by allowing a metric different from the shortest path metric to be introduced on the net-

work. The purpose of this new metric is precisely to take into consideration long-range interactions between institutions. Some important similarities can be found between this new metric and the resistance distance, a well-known metric in network theory derived from the study of electric circuits (Klein and Randic 1993; Estrada and Hatano 2010a; Lee et al. 2019), and its interpretation in terms of vibrational communicability (Estrada and Hatano 2010b; Bozzo 2013; Ferraz de Arruda et al. 2014; Van Mieghem et al. 2017).

An area in which these concepts allow us to gain a deep insight into the hidden structures of the network is properly the WTN. The topology of the world trade web has been extensively analysed over time (Serrano and Boguñá 2003; Li et al. 2003; Garlaschelli and Loffredo 2004, 2005; Garlaschelli et al. 2007; Fagiolo et al. 2008). The behaviour of international trade flows, the impact of globalization on the international exchanges, the presence of a core-periphery structure or the evolution of the community centres of trade are just some of the issues addressed by the recent developments (Serrano et al. 2007; Tzekina et al. 2008; Fagiolo et al. 2010; De Benedictis and Tajoli 2011; Blöchl et al. 2011). Many works have dealt with the network from a multi layers perspective (Snyder and Kick 1979; Barigozzi et al. 2011) or aim to emphasize financial implications of the world trade or contagion processes on the network (Wilhite 2001; Reyes et al. 2008; Schiavo et al. 2010; Fagiolo et al. 2013; Fan et al. 2014; Varela et al. 2015; Giudici and Spelta 2016; De Benedictis and Tajoli 2016; Cepeda-López et al. 2019; Cerqueti et al. 2019).

The impact of topology and metric properties on the stability and resilience of an economic or financial system has been widely studied in order to describe the large-scale pattern of dynamical processes inside the network (Smith and White 1992; Kali and Reyes 2007; Piccardi and Tajoli 2018). These processes determine the subsequent diversification of the export of a country, which can be compared with descriptive empirical indices of its potential growth, such as the one introduced in a very fruitful way in Hausmann et al. (2014).

3 Communicability in complex networks

The idea of communicability on a network is based on the ways in which a pair of nodes can communicate, namely through walks connecting them. In the literature, two different definitions of communicability have been introduced: the Estrada Communicability and the Vibrational Communicability (Estrada and Hatano 2008, 2010b). We recall them in this section.

3.1 Preliminary definitions

First of all, we briefly remind some preliminary definitions. A network is formally represented by a graph $\mathscr{G} = (V, E)$ where V and E are the sets of n nodes and m edges, respectively. Two nodes i and j are adjacent if there is an edge $(i, j) \in E$ connecting them. The network is undirected if (j, i) is an element of E whenever (i, j) is such. A $i - j$-path is a sequence of distinct vertices and edges between i and

j. The shortest path, or geodesic, between i and j is a path with the minimum number of edges. The length of a geodesic is called geodesic distance or shortest path distance $d(i, j) = d_{ij}$. A graph $\mathcal{G}$ is connected if, $\forall i, j \in V$, a $i - j$-path connecting them exists.

Adjacency relationships are represented by a binary symmetric matrix $\mathbf{A}$ (adjacency matrix). Graphs considered here will be always connected and without self-loops; in this case $a_{ii} = 0 \, \forall i = 1, \ldots, n$. We denote with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ the eigenvalues of $\mathbf{A}$, and $\varphi_i, i = 1, \ldots, n$ the corresponding eigenvectors.

The degree k_i of a node i is the number of edges incident on it. The diagonal matrix whose diagonal entries are k_i is $\mathbf{K}$. The Laplacian matrix is $\mathbf{L} = \mathbf{K} - \mathbf{A}$. $\mathbf{L}$ is a positive semidefinite symmetric matrix. We denote the eigenvalues of $\mathbf{L}$ by $\mu_1 \geq \mu_2 \geq \cdots > \mu_n = 0$ and $\psi_i, i = 1, \ldots, n$ the corresponding eigenvectors.

A graph $\mathcal{G}$ is weighted when a positive real number $w_{ij} > 0$ is associated with the edge (i, j). We define the strength s_i as the sum of the weights of the edges adjacent to i. The definition of geodesic path still holds, and it is a weighted path with the minimum sum of edge weights. In this case, the adjacency matrix is a nonnegative symmetric matrix $\mathbf{W}$. When $w_{ij} = 1$ if $(i, j) \in E$, then the graph is unweighted. Thus, the unweighted case can be viewed as a particular weighted one.

3.2 Estrada communicability

The *Estrada communicability* (Estrada and Hatano 2008) between two nodes i and j is defined as:

$$G_{ij} = \sum_{k=0}^{+\infty} \frac{1}{k!}[\mathbf{A}^k]_{ij} = \left[e^{\mathbf{A}}\right]_{ij}. \tag{1}$$

As the ij-entry of the k-power of the adjacency matrix $\mathbf{A}$ counts the number of walks of length k starting at i and ending at j, G_{ij} accounts for all channels of communication between two nodes, giving more weight to the shortest routes connecting them. It can also be interpreted as a measure of the probability that a particle starting at i ends up at j after wandering randomly on the complex network. The communicability matrix is denoted by $\mathbf{G}$.

By definition, it follows that $G_{ij} > 0$. Moreover, G_{ij} can be conveniently expressed using the spectral decomposition of $\mathbf{A}$ as follows (Estrada and Hatano 2008):

$$G_{ij} = \sum_{k=1}^{n} \varphi_k(i)\varphi_k(j)e^{\lambda_k},$$

where $\varphi_k(i)$ is the i-component of the kth eigenvector associated with λ_k.

It is worth noting that since G_{ii} characterizes the importance of a node according to its participation in all closed walks starting and ending at it, we recover the so-called *subgraph centrality* (see Estrada and Rodriguez-Velazquez 2005).

In the case of a weighted network, the communicability function is defined as

$$G_{ij} = \sum_{k=0}^{+\infty} \frac{1}{k!} \left[\left(\mathbf{S}^{-\frac{1}{2}} \mathbf{W} \mathbf{S}^{-\frac{1}{2}} \right)^k \right]_{ij} = \left[e^{\left(\mathbf{S}^{-\frac{1}{2}} \mathbf{W} \mathbf{S}^{-\frac{1}{2}} \right)} \right]_{ij} \tag{2}$$

where $\mathbf{S}$ is the diagonal matrix whose diagonal entries are the strengths of the nodes. We will call this quantity weighted communicability.

3.3 Vibrational communicability

Vibrational communicability represents an alternative definition of communicability, different from Estrada communicability, and which can be introduced through the following model. Let us suppose that nodes of the network are objects of negligible identical mass connected by springs in a plane grid. Nodes can oscillate in the direction perpendicular to the plane and the displacement of the node i from its rest position is z_i. The elastic force applied to node i is given by $F_i = \mathcal{K} \sum_j A_{ij}(z_i - z_j)$, where $\mathcal{K}$ is the common elastic constant of each spring. An elastic potential energy can be assigned to each perturbed spring, and the potential energy of all the springs connected with node i is given by $U_i = \frac{1}{2} \mathcal{K} \sum_j A_{ij}(z_i - z_j)^2$.

The overall potential energy of the network is therefore

$$U = \frac{1}{4} \mathcal{K} \sum_{i,j} A_{ij}(z_i - z_j)^2 = \frac{1}{2} \mathcal{K} \sum_{i,j} z_i L_{ij} z_j \tag{3}$$

where L_{ij} is the ij-entry of $\mathbf{L}$.

The reciprocal influence of two nodes i and j in their positions z_i and z_j is computed by means of the Green's function, according to the classical Boltzmann's distribution (Estrada and Hatano 2010a, b). This mutual influence can be interpreted as the correlation function between the displacements z of two nodes in the network:

$$G_{ij}^v(\beta) = \langle z_i z_j \rangle = \frac{1}{\mathcal{Z}} \int z_i z_j e^{-\beta U} \, d\mathbf{z}$$

where β is a constant and $\mathcal{Z} = \int e^{-\beta U} \, d\mathbf{z}$ is the partition function. Using the nonzero eigenvalues of $\mathbf{L}$, $\mathcal{Z}$ can be expressed as

$$\mathcal{Z} = \int e^{-\frac{1}{2} \beta \mathcal{K} \sum_{ij} z_i L_{ij} z_j} \prod_k dz_k = \prod_{k=1}^{n-1} \sqrt{\frac{2\pi}{\beta \mathcal{K} \mu_k}} \tag{4}$$

so that the correlation function can be rewritten in the final form

$$G_{ij}^v(\beta) = \sum_{k=1}^{n-1} \frac{\psi_k(i) \psi_k(j)}{\beta \mathcal{K} \mu_k} \tag{5}$$

where ψ_k is the eigenvector associated with μ_k. Introducing the Moore–Penrose pseudo-inverse of the Laplacian $\mathbf{L}^+$ (Bozzo 2013; Gutman and Xiao 2004), the vibrational communicability between nodes i and j is defined as

$$G_{ij}^v(\beta) = \frac{1}{\beta \mathcal{K}} L_{ij}^+ \tag{6}$$

The vibrational communicability matrix is denoted by $\mathbf{G}^v$. In the remainder of the paper, we will assume $\beta = 1$ and $\mathcal{K} = 1$, so that $G_{ij}^v = L_{ij}^+$.

The detailed computations for previous formulas are reported in "Appendix A".

4 Metrics on networks

Metric properties play an important role in the study of the structure and dynamics of networks. The best known metric is the so-called shortest path distance. In the literature, other metrics have been defined, each one stressing different features of the network. We remind the definitions of communicability distance and resistance distance, in view of their following application to the WTN.

4.1 Communicability distance

The communicability distance ξ_{ij} is defined as (see Estrada 2012a):

$$\xi_{ij} = G_{ii} - 2G_{ij} + G_{jj}. \tag{7}$$

As already observed, G_{ii} is the subgraph centrality of i and it measures the amount of information that starts from and returns to node i after having wandered through the network. On the other hand, G_{ij} measures the amount of information transmitted from i to j. Notice that the word *information* is meant in its broadest sense. Therefore, information flow can be any kind of flow along edges: money, current, traffic and so on. Thus, the quantity ξ_{ij} accounts for the difference in the amount of information that returns to the nodes i and j and the amount of information exchanged between them.

The greater is G_{ij}, the larger the information exchanged and the nearer are the nodes; the greater are G_{ii} or G_{jj}, the larger the information that comes back to the nodes and the farther are the nodes. In a matrix form, ξ_{ij} can be expressed as follows:

$$\mathbf{\Xi} = \mathbf{g}\mathbf{u}^T - 2\mathbf{G} + \mathbf{u}\mathbf{g}^T$$

where $\mathbf{g} = [G_{11}, \ldots, G_{nn}]^T$ is the vector of subgraph centralities and $\mathbf{u}$ the all 1's $n-$vector. Since ξ_{ij} is a metric, $G_{ii} + G_{jj} \geq 2G_{ij}$, i.e. no matter what the structure of the network is, the amount of information absorbed by a pair of nodes is always larger than the amount of information transmitted between them.

4.2 Resistance distance

The vibrational communicability distance between i and j is defined as (see Estrada and Hatano 2010a; Van Mieghem et al. 2017):

$$\omega_{ij} = G_{ii}^v - 2G_{ij}^v + G_{jj}^v. \tag{8}$$

Formula 8 can be written in a more suitable way. Indeed, recalling that $G_{ij}^v = L_{ij}^+$, we have:

$$
\begin{aligned}
\omega_{ij} &= L_{ii}^+ - 2L_{ij}^+ + L_{jj}^+ \\
&= (\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{L}^+ (\mathbf{e}_i - \mathbf{e}_j) \\
&= (\mathbf{e}_i - \mathbf{e}_j)^T \left[\left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right)^{-1} - \frac{1}{n}\mathbf{J} \right] (\mathbf{e}_i - \mathbf{e}_j) \\
&= (\mathbf{e}_i - \mathbf{e}_j)^T \left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right)^{-1} (\mathbf{e}_i - \mathbf{e}_j)
\end{aligned}
\tag{9}
$$

where $\mathbf{e}_k$, $k = 1, \ldots, n$, is the standard basis in $\mathbb{R}^n$ and $\mathbf{J} = \mathbf{u}\mathbf{u}^T$ is the matrix whose entries are all 1. Note that in the previous chain of equalities, we made use of the following expression of the pseudo-inverse $\mathbf{L}^+ = \left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right)^{-1} - \frac{1}{n}\mathbf{J}$, proved in Gutman and Xiao (2004).

Equation 9 offers an interesting interpretation of the resistance distance. We synthesize here the main idea, referring to "Appendix B" for a more detailed discussion. Let $\mathbf{v} = [v_1, v_2, \ldots, v_n]^T$ be a vector representing attributes of the nodes—for instance, the gross domestic product (GDP) of a country or the assets of a financial institution—and suppose that there are currents or flows (of money, for instance) along the edges of the network. The operator $\left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right)^{-1}$ allows to obtain the state vector that gives rise to a given set of flows. In Formula 9, the vector $(\mathbf{e}_i - \mathbf{e}_j)$ refers to a global flow equal to $+1$ from node i, a flow equal to -1 into node j and a flow equal to 0 for the other ones. When we apply $\left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right)^{-1}$ to $(\mathbf{e}_i - \mathbf{e}_j)$, we get the state vector $\mathbf{v} = [v_1, v_2, \ldots, v_n]^T$ of attributes on nodes that gives rise to these flows. Finally, the left inner product with $(\mathbf{e}_i - \mathbf{e}_j)$ in Formula 9 gives $v_i - v_j$, namely the difference between attributes of nodes i and j. This gradient produces exactly the flow $+1$ from node i and -1 to node j. If $v_i - v_j$ is big, we need a big difference in order to produce such a unit flow and so we have a big resistance between nodes i and j. If $v_i - v_j$ is small, it is enough a low difference in order to produce such a unit flow and so we have a low resistance between nodes i and j. If ω_{ij} is big, we have a high resistance distance between i and j. Therefore, these two nodes do not communicate easily. Vice versa a low value of ω_{ij} means a high level of communication between the nodes. ω_{ij} is called *effective resistance* between nodes i and j and $\mathbf{\Omega} = [\omega_{ij}]$ is the resistance matrix.

In the literature, it is known an important close form for $\mathbf{L}^+$ in terms of $\mathbf{\Omega}$:

$$\mathbf{L}^+ = \frac{1}{2} \left[\frac{1}{n}(\mathbf{\Omega}\mathbf{J} + \mathbf{J}\mathbf{\Omega}) - \frac{1}{n^2}\mathbf{J}\mathbf{\Omega}\mathbf{J} - \mathbf{\Omega} \right]$$

which allows us to rewrite the diagonal elements of the matrix $\mathbf{L}^+$ in a useful form[1]

$$L_{ii}^+ = \frac{1}{n} \sum_j \omega_{ij} - \frac{R}{n^2}$$

where

$$R = \frac{1}{2} \sum_{i,j} \omega_{ij} = \sum_{i=1}^n \sum_{j=i+1}^n \omega_{ij} = \frac{1}{2} \mathbf{u}^T \mathbf{\Omega} \mathbf{u} = n \operatorname{tr} \mathbf{L}^+ = n \sum_{k=1}^{n-1} \frac{1}{\mu_k}$$

is the *effective graph resistance* (or *Kirchhoff index*) of the network, i.e. the sum of the resistances between all possible pairs of nodes in the graph (see, for example, Klein and Randic 1993). R reflects the overall transport capability of the network: The lower R, the better the network conducts flows. In particular, it has been shown that this index is able to catch the average vulnerability of a connection between a pair of nodes, and therefore, it is a suitable tool for assessing the ability of a network to well react when it is subject to failure and/or attack (see Ellens et al. 2011; Wang et al. 2014; Clemente and Cornaro 2019).

Effective resistances allow to give a specific definition of the centrality of a node in the network. Indeed, the *best spreader* (or best connected) node in the network is the node $i^\star$ that minimizes the quantity $\sum_{j=1}^n \omega_{i^\star j} = (\mathbf{\Omega u})_{i^\star}$, i.e. the sum of all its resistance distances from any other node in the network. Since L_{ii}^+ equals the difference between the average resistance between node i and all the other nodes in the network and the overall network mean resistance, the best spreader node $i^\star$ is the one such that $L_{i^\star i^\star}^+ \leq L_{jj}^+$ for any $j \neq i^\star$. Node $i^\star$ can be regarded as the best diffuser of a flow to the rest of the network, and, to some extent, it is the most influential with respect to a diffusion process inside the network, since it guarantees the highest flow towards other nodes (see Van Mieghem et al. 2017). Best diffuser means that most of the information coming out from this node is absorbed by other nodes. If L_{ii}^+ is big, then most of this information comes back to node i and does not reach other nodes. The reciprocal of L_{ii}^+ can then be regarded as a centrality measure of a node, and it is called *vibrational centrality*.

5 Community detection based on communicability metrics

5.1 The model

As discussed in the previous section, $\xi_{ij} = G_{ii} - 2G_{ij} + G_{jj}$ and $\omega_{ij} = G_{ii}^v - 2G_{ij}^v + G_{jj}^v$ represent the two metrics induced on the network by the Estrada communicability and the vibrational communicability, respectively.

---

[1] $L_{ii}^+ = \frac{1}{2n}(\mathbf{\Omega J})_{ii} + \frac{1}{2n}(\mathbf{J\Omega})_{ii} - \frac{1}{2n^2}(\mathbf{J\Omega J})_{ii} - \frac{1}{2}(\mathbf{\Omega})_{ii} = \frac{1}{2n} \sum_j \omega_{ij} J_{ji} + \frac{1}{2n} \sum_j J_{ij} \omega_{ji} - \frac{1}{2n^2} \sum_{jk} J_{ij} \omega_{jk} J_{ki} - 0 = \frac{1}{2n} \sum_j \omega_{ij} + \frac{1}{2n} \sum_j \omega_{ij} - \frac{1}{2n^2} \sum_{jk} \omega_{jk} = \frac{1}{n} \sum_j \omega_{ij} - \frac{R}{n^2}.$

In an economical context, referring to the international trade network, they measure how well two countries, or companies, communicate in terms of commercial and trade exchanges. For instance, the attributes on nodes may be identified with the GDP and the currents along nodes with the total trade or money flow between two countries. Information on the network may be replaced by money flow. Therefore, the quantity ξ_{ij} of Eq. 7 accounts for the difference in the amount of money flow that returns to the nodes i and j and the amount of money flow exchanged between them. The bigger is G_{ij}, i.e. the money flow exchanged, the nearer are the nodes; the bigger are G_{ii} or G_{jj}, i.e. the amount of money flow that comes back to the each node, the farther they are. A similar interpretation holds for ω_{ij}. In a trade network, ω_{ij} accounts for the difference between the mean resistance to export a given money flow from each country and the correlation between them. The bigger is G_{ij}^v, the more interconnected they are and the nearer they are in the resistance metric; the bigger are G_{ii}^v and G_{jj}^v, the more isolated they are in the network and between them and the farther they are.

In light of these observations, we formulate our proposal,[2] considering as members of the same cluster nodes whose mutual distance is below a given threshold ξ_0. Specifically, we construct a new community graph where the elements of the adjacency matrix $\mathbf{M} = [m_{ij}]$ are given by:

$$m_{ij} = \begin{cases} 1 & \text{if } \xi_{ij} \leq \xi_0 \\ 0 & \text{otherwise} \end{cases}$$

with ξ_0 threshold distance such that $\xi_0 \in [\xi_{\min}, \xi_{\max}]$, being $\xi_{\min}$ and $\xi_{\max}$ the minimum and the maximum distances between couples of nodes, respectively. In this way, clustered groups of nodes that strongly communicate emerge, in dependence of the threshold. If ξ_0 is high enough, all nodes in the network are at a mutual distance lower than the threshold and the whole network behaves like a unique community. As ξ_0 decreases, there will be nodes too far, such that to be considered disconnected and then members of different clusters, entailing the emergence of islands of connected nodes. Hence, the number of communities depends on the threshold; precisely, it increases as ξ_0 decreases.

It is important to observe that, with the proposed methodology, we do not choose any a priori optimal number of communities. Our approach is more in line with the classic Girvan–Newman approach (Newman and Girvan 2004).

The optimal partition is determined according to an optimization problem whose objective function is based on the idea of cohesion between nodes. Specifically, since we deal with distances, following the approach for clustering in metric spaces proposed by Chang et al. (2016), we provide a cohesion measure γ_{ij} between two nodes i and j, as follows:

$$\gamma_{ij} = \left(\bar{\xi}_j - \bar{\xi} \right) - \left(\xi_{ij} - \bar{\xi}_i \right)$$

---

[2] In what follows, we will refer to the communicability distance ξ, but similar arguments may be repeated identically for the resistance distance ω.

where $\bar{\xi}_i = \frac{1}{n-1} \sum_{k \neq i} \xi_{ik}$ is the average distance between i and nodes other than i and $\bar{\bar{\xi}}$ is the average distance over the whole network. Thus, $\xi_{ij} - \bar{\xi}_i$ represents the *relative distance* between nodes i and j and $\bar{\xi}_j - \bar{\bar{\xi}}$ represents the *relative distance* from a random node to the node j.

Two nodes i and j are said to be cohesive (or incohesive) if $\gamma_{ij} \geq 0$ ($\gamma_{ij} \leq 0$). Notice that $\gamma_{ij} \geq 0$ yields $\xi_{ij} + \bar{\bar{\xi}} \leq \bar{\xi}_i + \bar{\xi}_j$, i.e. intuitively, two nodes are cohesive if they are close to each other and, on average, they are both far away from the other nodes. In other words, γ_{ij} can be interpreted as the gain (when positive) or the cost (when negative) related to the grouping of nodes i and j in the same cluster of a given partition.

We assume to maximize an objective function that represents the global cohesion function based on the mutual relative distances between every pairs of nodes. Therefore, we refer to a specific partition quality index defined as

$$Q = \sum_{i,j} \gamma_{ij} x_{ij} \tag{10}$$

where x_{ij} is a binary variable equal to 1 if two nodes are in the same cluster and 0 otherwise and γ_{ij} is the cohesion measure between nodes i and j. It is worth to notice that when the partition is made up of a unique community, equal to the entire network, $x_{ij} = 1 \ \forall i, j$. In this case[3]

$$\begin{aligned} Q = \sum_{i,j} \gamma_{ij} &= \sum_{i,j} \bar{\xi}_j + \sum_{i,j} \bar{\xi}_i - \sum_{i,j} \bar{\bar{\xi}} - \sum_{i,j} \xi_{ij} \\ &= n \sum_j \bar{\xi}_j + n \sum_i \bar{\xi}_i - n^2 \bar{\bar{\xi}} - \sum_{i,j} \xi_{ij} \\ &= 2n^2 \bar{\bar{\xi}} - n^2 \bar{\bar{\xi}} - n(n-1)\bar{\bar{\xi}} \\ &= n\bar{\bar{\xi}}. \end{aligned}$$

On the other hand, when the partition consists of n isolated nodes, $x_{ij} = 0 \ \forall i \neq j$ then

$$Q = \sum_i \gamma_{ii} = \sum_i (\bar{\xi}_i - \bar{\bar{\xi}}) - (\xi_{ii} - \bar{\xi}_i) = 2 \sum_i \bar{\xi}_i - n\bar{\bar{\xi}} = n\bar{\bar{\xi}}.$$

Thus, in these two extreme cases, Q provides the same value $n\bar{\bar{\xi}}$.

5.2 An illustrative example

We start by testing our methodology on a simple example. Let us consider the weighted undirected network displayed in Fig. 1. The network has 10 nodes and 32 edges. The thickness of links is proportional to weights. The network allows to easily identify two

---

[3] Notice that $\bar{\bar{\xi}} = \frac{1}{n} \sum_i \bar{\xi}_i = \frac{1}{n(n-1)} \sum_{i,j} \xi_{ij}$.
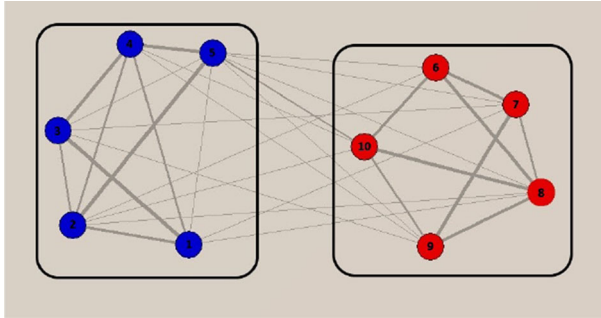
Fig. 1 A weighted undirected network with 10 nodes and 32 edges. Edges weights have been randomly sampled with replacement from integers between 1 and 6. The thickness of edges is proportional to the weights. Nodes of two relevant communities are highlighted in blue and red (colour figure online)

natural communities, which are highlighted by the two closed lines containing nodes 1 to 5 (on the left) and nodes 6 to 10 (on the right).

We compute the Estrada communicability matrix $\mathbf{G}$; then, we get the communicability distance matrix $\mathbf{\Xi}$. The nearest nodes are 1 and 3 with a communicability distance equal to $\xi_{\min} = \xi_{13} = 1.18$ and farthest nodes are 3 and 6 with a communicability distance equal to $\xi_{\max} = \xi_{36} = 1.49$. Figure 2 summarizes the number of communities identified at different thresholds. The blue line represents the number of communities, while the red line represents the quality index Q of the corresponding partition. When the threshold is greater than or equal to $\xi_0 = 1.38$, all nodes are connected and the network is partitioned in a single community, with quality index $Q = n\bar{\xi}$. As the threshold decreases below 1.38, the network begins to split into disconnected components. When the threshold becomes lower than the minimum distance, the network is partitioned into ten communities and each node belongs to a different community. The best partition according to the maximum quality index criterion splits the network into two clusters, which are easily identified with the two expected natural communities. The composition of the communities for alternative thresholds is reported in Fig. 3. It is noticeable that, lowering the threshold, the procedure allows to disentangle tightest relationships. For instance, when $\xi_0 = 1.23$, only nodes connected by edges with highest weights are kept in the same community.

Similar results are derived by applying the procedure based on the vibrational communicability. The nearest nodes are 1 and 3 with a resistance distance equal to $\omega_{\min} = \omega_{13} = 1.22$, and farthest nodes are 3 and 8 with a resistance distance equal to $\omega_{\max} = \omega_{38} = 1.69$. Again if we move the threshold from the maximum distance to the minimum distance, we get an increasing number of communities from 1, the whole network, to 10, isolated nodes. The best partition according to the maximum quality index criterion splits the network into the two expected communities, as shown in Fig. 4.
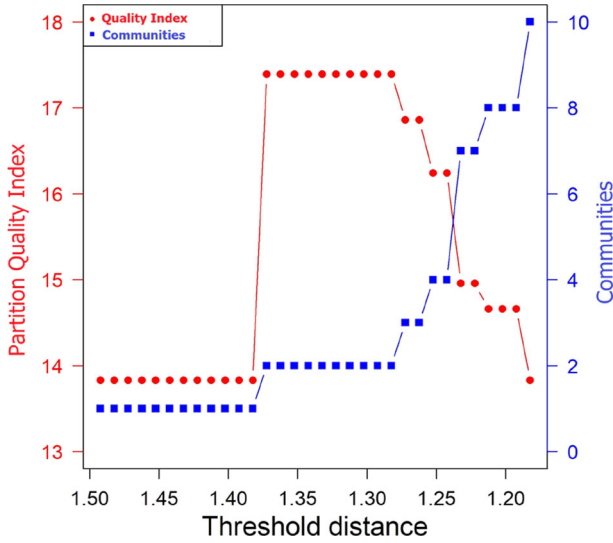
Fig. 2 Quality index Q of the partition computed according to Formula 10 and number of components (on the secondary scale) for different threshold values. The communicability distance has been used for the identification of the communities
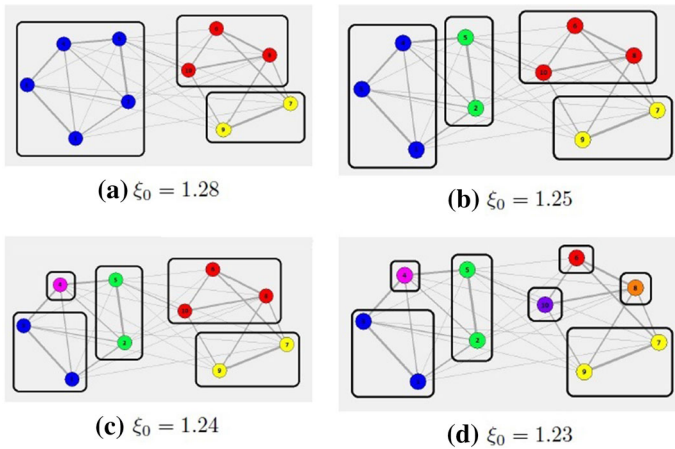

(a) $\xi_0 = 1.28$

(b) $\xi_0 = 1.25$

(c) $\xi_0 = 1.24$

(d) $\xi_0 = 1.23$

Fig. 3 Community structure at different thresholds

6 Application to the World Trade Network

In this section, we apply the proposed model in order to detect relevant communities of countries in the WTN. As described before, the method aims at grouping strongly interacting countries by means of their mutual distances. Two alternative distance functions will be tested. On the one hand, we find clusters exploiting communicability distance. Therefore, we detect how much two countries are close in the network considering all possible weighted walks connecting them. On the other hand, we select
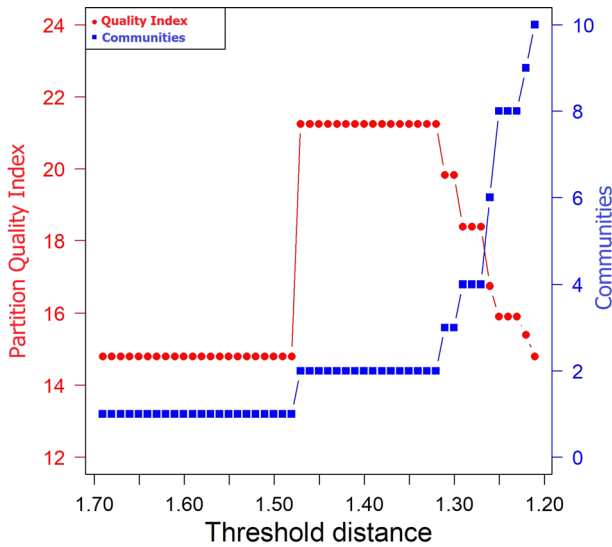
Fig. 4 Quality index Q of the partition and number of components (on the secondary scale) for different thresholds. The resistance distance has been used for the identification of the communities

clusters by means of resistance distance. In this case, countries are grouped together if they have a similar relevance in the network in terms of vibrational centralities as well as if they are correlated in terms of their expositions towards common countries.

We start with a general description of the dataset and the main characteristics of the WTN. Then, we briefly summarize the primary steps of the methodology, providing a pseudo-code of the algorithm. Finally, we report the results in terms of community structure with the related discussion.

6.1 Dataset and main characteristics of the WTN

We refer to the World Trade Data, available on the Observatory of Economic Complexity database.[4] The database has been developed by the Research and Expertise Center on the World Economy at a high level of product disaggregation, and it is based on original data provided by the United Nations Statistical Division (UN Comtrade). In particular, a harmonization procedure, that reconciles the declarations of exporters and importers, enables to extend considerably the number of countries for which trade data are available, as compared to the original dataset. In this analysis, we refer to the last version published in 2017, based on the Harmonized Commodity Description and Coding System, and that provides aggregated bilateral values of exports for each couple of origin and destination countries, expressed in billion dollars. We focus on the aggregated data of last available year, namely 2016.

Hence, we construct a weighted network where each node is a country and weighted links represent the amount of product traded between couple of countries (see Fig. 5).

---

[4] The Observatory of Economic Complexity (OEC) is the world's leading data visualization tool for international trade data. Data can be found at: https://atlas.media.mit.edu/en/.

Fig. 5 WTN based on 2016 data. Nodes are countries, and links are product trades between pair of countries. The size of the node is proportional to its strength
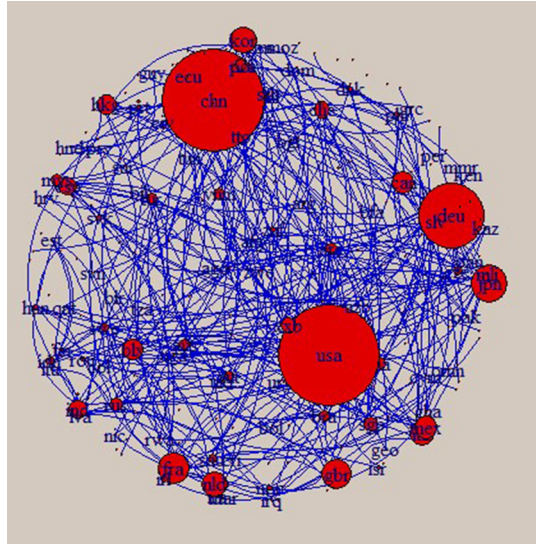
The mutually exchanged products between two countries are different in terms of entity, so that they can be better represented by oriented links from a country to another one. However, we observed a strict relation between in and out strength distribution with a Spearman correlation coefficient equal to 0.956. Hence, countries are ranked in a very similar way in terms of in and out strength. Thus, we perform all the analysis assuming the network as undirected.

The undirected network is characterized by 221 nodes and 14933 links. The network is connected, and its density is approximately 0.614: on average, each country has trades with more than a half of the entire network. However, the network is not regular and is far from being complete. In other words, most countries do not trade with all the others, but they rather select their partners. Furthermore, main trade flows tend to be concentrated in a specific subgroup of countries and a small percentage of the total number of flows accounts for a disproportionately large share of world trade. For instance, the top 10 countries export more than 50% of the total flow. The maximum weight corresponds to the channel between China and the USA and its value amounts to 277 billion dollars. Minimum, non- null, weights are involved in the trade between a number of very small countries, far from each others, and they are approximately around 1 thousand dollars.

Finally, we expect that several countries trade with their geographical neighbours so that we investigate the correlation between flows and geographical distance of countries. We computed the Spearman rank correlation between link weights (i.e. monetary flows between countries in the network) and the great circle distance between capital cities in kilometres. We obtained a rank correlation of -0.27 that confirms a little preference for trading with physical neighbours. However, as stressed before, our aim is to go beyond immediate neighbours by means of both communicability and resistance distances.

6.2 Summary of the methodology

In this section, we summarize by means of a pseudo-code the main steps of the methodology we are proposing. The code has been written taking into account the communicability distance matrix $\mathbf{\Xi}$, but the same procedure can be easily applied by considering the resistance matrix $\mathbf{\Omega}$.

1. let $\mathscr{G}$ be the original directed weighted network with n nodes and weighted adjacency matrix $\mathbf{W}$;
2. build the undirected weighted network $\mathscr{G}_1$ with a symmetric adjacency matrix defined as $\mathbf{W}_1 = \frac{1}{2}(\mathbf{W} + \mathbf{W}^T)$;
3. build the undirected weighted network $\mathscr{G}_2$ with normalized weighted adjacency matrix $\mathbf{W}_2 = \mathbf{S}^{-1/2}\mathbf{W}_1\mathbf{S}^{-1/2}$, where $\mathbf{S}$ is the diagonal matrix of the strengths of the network $\mathscr{G}_1$;
4. construct the distance matrix $\mathbf{\Xi} = \mathbf{g}\mathbf{u}^T - 2\mathbf{G} + \mathbf{u}\mathbf{g}^T$ based on the communicability matrix $\mathbf{G}$;
5. define the threshold interval $[\xi_{\min}, \xi_{\max}]$, where $\xi_{\min}$ and $\xi_{\max}$ represent the minimum and the maximum communicability distances between couples of nodes, respectively, and set $\xi_h = \xi_{\min}$, with $h = 0$;
6. define a $n \times n$ matrix $\mathbf{M}_h = [m_{ij}]$ such that

$$
m_{ij} = \begin{cases} 1 & \text{if } \xi_{ij} \leq \xi_h \text{ and } i \neq j \\ 0 & \text{otherwise} \end{cases};
$$

7. build the undirected unweighted network $\mathscr{G}_{3,h}$ from the binary adjacency matrix $\mathbf{M}_h$;
8. select the partition P_h given by the components of the network $\mathscr{G}_{3,h}$;
9. compute the quality index $Q = \sum_{i,j} \gamma_{ij} x_{ij}$ of the network $\mathscr{G}_2$ with respect to the partition P_h;
10. set the number of iterations r, compute $k = \frac{\xi_{\max} - \xi_{\min}}{r}$, set $\xi_h = \xi_{h-1} + k$ and $h = h + 1$ and repeat steps 6-9 while $\xi_h \leq \xi_{\max}$;
11. select the optimal partition $P_h^\star$ as the partition P_h that provides the maximum quality index Q.

We stress some key points of the presented methodology. We aim at clustering countries on the basis of a specific distance. The two distances we have chosen highlight relationships of a different nature between countries, and the different community structure emerging will support this fact. Varying the threshold, we can disentangle the role of very tight relationships between couples of countries. Of course, reducing the threshold distance a great number of isolated nodes may appear. They are typically very small countries whose trade volume is very low and whose commercial partners are few. They play a marginal role in the WTN, and they do not affect in a significant way the structure of the network in terms of relevant communities. This is the reason why we will focus our attention on the main communities that are produced by our methodology.

6.3 Results

6.3.1 Results in terms of communicability metric

We initially applied the methodology described in Sect. 6.2 by using the communicability distance. The rationale for using the communicability metric on the WTN is the following. Two countries share a total volume of trade because they exchange a given set of products, of any kind. But they can be linked even if they do not exchange each other a given product; that is, there is no direct flow of such product between them. A higher order exchange may occur between them. For instance, a country A exports some raw materials—let's say, iron—to a country B; country B produces mechanical parts from iron and exports them to country C. A and C communicate via a higher-order walk, and they depend on each other even if the two countries are not neighbours in the network. Indeed, communicability takes into account precisely all possible weighted walks between two nodes.

Therefore, we calculate the communicability matrix $\mathbf{G}$ on the normalized network $\mathcal{G}_2$ and the corresponding communicability distance matrix $\mathbf{\Xi}$. Using this metric, we find that the nearest countries are the USA and Canada with a distance $\xi_{\min} = 1.242$ and the farthest countries are the USA and Seychelles Islands with a distance $\xi_{\max} = 1.470$. Lowering the threshold distance value from maximum to minimum with a 0.001 step, we look at the corresponding partition in communities. In Fig. 6, we plot the value of partition quality index Q (in red) and the number of communities (in blue), counting each isolated node as an independent one. Both values are expressed as functions of the threshold ξ_h. The maximum of Q is reached at a threshold distance $\xi_h = 1.392$. It corresponds to 106 communities, among which we have 87 isolated nodes. Hence, we observe 19 significant communities other than isolated nodes.

We display in Fig. 7 communities in the optimal partition, and we list in Table 1 the countries belonging to the ten biggest communities in terms of numerousness.

Going deeper into the composition of the communities, the biggest one (see community 1 in blue) includes almost all continental European countries, with Great Britain and Ireland. This community acts on the screen of the global network as single player. It is worth pointing out also the presence of Morocco, confirming positive effects of bilateral trade agreements (see, for example, Van Berkum 2013). We also notice the presence of South Asian countries that are economically linked together by the South Asian Association for Regional Cooperation. The presence of these countries in the community is also an effect of the bilateral foreign relations between the European Union (EU) and the Association of Southeast Asian Nations (ASEAN). The partnership between the EU and ASEAN dates back to 1972 when the EU countries became ASEAN's first formal dialogue partner. Finally to the same community belong African countries that are characterized by close economic and cultural ties to European countries, in particular to France (see, for instance, Ivory Coast, Burkina Faso, Angola, Senegal).

Opposed to this community, we see the second largest community (see community 2 in red) which sees the USA and China as main actors. This means that in Europe, there are preferential channels of internal exchanges, whereas, outside Europe, most communication channels seem to be polarized around the exchange channel between
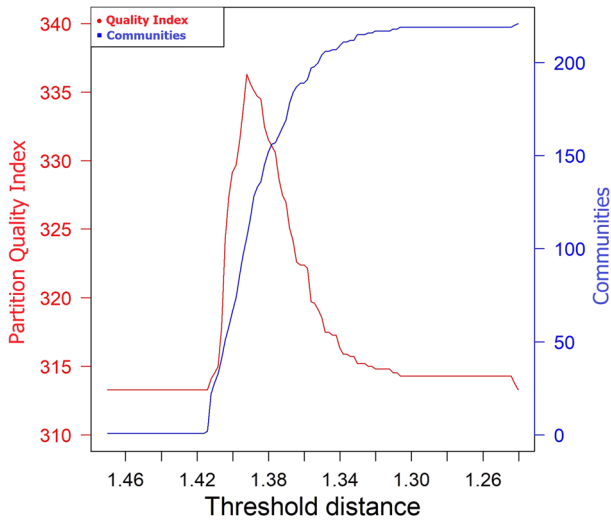
Fig. 6 Partition quality index Q (red line) and number of communities (blue line) as functions of the threshold communicability distance ξ_h. Maximum Q is observed for $\xi_h = 1.392$ (colour figure online)

Fig. 7 Optimal community structure based on communicability distance. Results have been derived by means of a maximum quality index criterion, with threshold $\xi_h = 1.39$ (Max Q Partition). Isolated nodes appear in white

China and the USA and all their satellites countries. Moreover, we can recognize other well-identified and coherent communities.

Furthermore, it is interesting the decomposition of post-Soviet States. While Baltic and Eastern Europe States (except for Ukraine) have main partners in European countries, and Central Asian countries have Russia as their leading trade and economic partner (see community 3). Although a positive trade balance and a priority of Russian government of an increasing participation in the economic relations of Asia-pacific

Table 1 Members of the top ten communities in terms of number of countries

	Size	Members
Community 1	54	AFG AGO ARE AUT BFA BGR BHR BIH BLX
		CHE CIV CYP CZE DEU DNK ESP EST FIN
		FRA GBR GRC GRL HRV HUN IND IRL IRN
		IRQ ITA JOR LTU LVA MAR MDA MKD MLI
		MNE NGA NLD NOR NPL OMN PAK POL PRT
		ROU SAU SEN SRB SVK SVN SWE TUR YEM
Community 2	21	AUS CAN CHN HKG IDN JPN KHM KOR LAO
		MEX MHL MMR MYS NZL PHL PNG SGP THA
		USA VNM XXB
Community 3	7	BLR KAZ KGZ RUS TJK UKR UZB
Community 4	6	ARG BOL BRA CHL PRY URY
Community 5	6	BDI COD KEN RWA SSD UGA
Community 6	5	BES COL ECU PAN PER
Community 7	5	CRI GTM HND NIC SLV
Community 8	4	GUY JAM SUR TTO
Community 9	4	MOZ ZAF ZMB ZWE
Community 10	3	ARM AZE GEO

region (see Kuznetsova et al. 2016), at moment, results show preferential channels with border countries. Transcaucasia is instead detected as a separate community (see community 10).

Except for Mexico, characterized by strong ties with the USA, the Latin American and the Caribbean Economic System is decomposed into four relevant communities (see communities 4, 6, 7 and 8). In particular, it is noticeable community 4 developed on the basis of the South Common Market, namely the so-called Mercosur. Mercosur's purpose is to promote free trade and the fluid movement of goods, people and currency in south America. Since its foundation, Mercosur's functions have been updated and amended many times; it currently confines itself to a customs union, in which there is free intra-zone trade and a common trade policy between member countries. In 2019, the Mercosur had generated a nominal gross domestic product (GDP) of around 4.6 trillion US dollars, reaching the fifth economy of the world.

Finally, significant blocks are also observed in central and south Africa (communities 5 and 9, respectively), polarized around Democratic Republic of the Congo and Republic of South Africa.

If we reduce the threshold, we let very strong channels of communication between countries emerge. For instance, Figs. 8 and 9 show the community structure lowering the threshold distance (equal to $\xi_h = 1.37$ and $\xi_h = 1.35$, respectively). Moving from 1.39 to 1.37, some loose connections are lost (see Fig. 8). Scandinavia and the Nordic Region split up from community 1 creating a separate cluster together. The South East Asian and former Yugoslavia appear as separate communities characterized only by most relevant partnerships, Australia goes out from community 2, and the strong
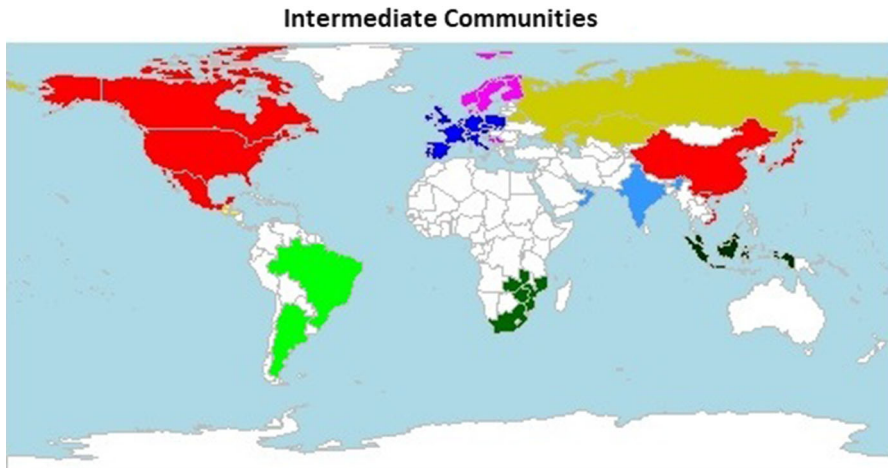
Intermediate Communities

Fig. 8 Intermediate connected community structure—$\xi_h = 1.37$

community in the South of Africa loses some country. Furthermore, in South America, only the relation between Brazil and Argentina survives. This result is in line with the fact that the strategic relationship between Argentina and Brazil is considered to be at the highest point in history: Brazil accounts indeed for Argentina's largest export and import market.

Reducing further the threshold to 1.35, only the most closely interrelated communities survive. The strongest community counts now, among its members, all North America, Mexico, China and Japan (in red in Fig. 9). In Europe, two communities are saved. On the one hand, the relation between Spain and Portugal is preserved. On the other hand, a community emerged in central Europe around the channel between France and Germany. Finally, community 3 in Table 1, including Russia and Central Asian countries, resists also when the threshold is lowered.

A significant feature of our approach is the fact that it allows to get deeper insight into the internal structure of each community and to give a measure of the mutual relationships between communities. Let us refer now to the clusters depicted in Fig. 7 and detected with the maximum quality index criterion. In this regard, we display in Fig. 10 the distributions of the communicability distances between pair of countries that belong to the same community. In particular, we compare the distributions for the first two relevant communities listed in Table 1

In fact, if we focus, for instance, on communities 1 and 2, we can inspect and compare their internal structure by providing some synthetic indicators in Table 2. From the analysis of Fig. 10 and of the values shown in Table 2, we can say that the community 2 (let's say, the USA–China) shows slightly more intense interactions than community 1 (let's say, Europe) since in the former, the average intracluster distance is slightly lower than in the latter. However, although the largest number of countries that belong to community 1, a more compact distribution is observed with a lower volatility. Trading interactions between countries in community 1 appear indeed somehow more homogeneous than between countries in community 2. This is partially
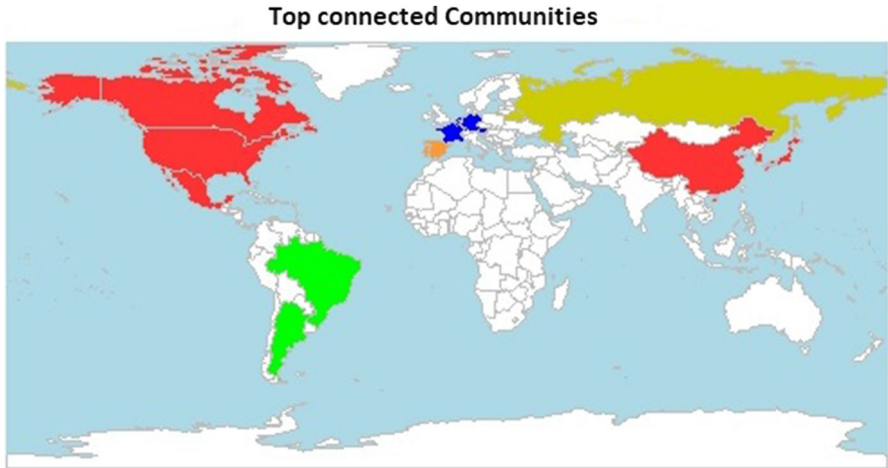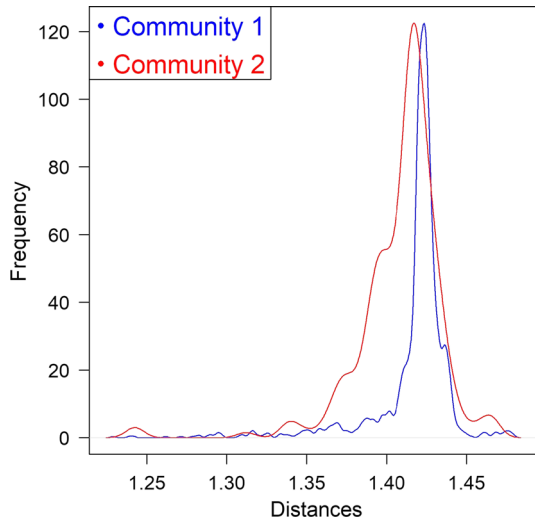
Top connected Communities

Fig. 9 Top connected community structure—$\xi_h = 1.35$

Fig. 10 Distributions of communicability distances between countries of the same community. We display only the distributions related to the two main communities summarized in Table 1

related to the geographical distribution of the countries inside the two communities. We have indeed that community 2 can be interpreted as the aggregation of different blocks mainly developed around the USA, China and Japan.

Last column of Table 2 provides the same indicators computed on intercluster basis. This analysis allows to provide additional information in terms of heterogeneity in the group and between groups. It is worth pointing out the lower intercluster standard deviation. It means that couple of countries that belong to a different community has a similar distance between them.

It is noteworthy that additional insights can be provided by assessing the relevance of each country in the community. Indeed, communicability distance matrix provides a metric on the network and on each subnetwork, like a community. Therefore, we

Table 2 Intercluster and intracluster characteristics of the distributions of communicability distances

	Intracluster		Intercluster
	Community 1	Community 2	Community 1 versus 2
Number of nodes	54	21	–
Mean distance	1.414	1.409	1.423
Min distance	1.325	1.242	1.393
Closest countries	NLD-BLX	USA-CAN	SAU-KOR
Max distance	1.444	1.467	1.469
Furthest countries	DEU-AFG	USA-LAO	USA-MNE
Standard deviation	0.012	0.028	0.011

Columns community 1 and community 2 refer to the intracluster properties of the two main detected communities, in terms of number of nodes. Last column reports the corresponding intercluster properties computed between the same two communities

adapt the idea of closeness to our context, by providing the following communicability closeness to assess how effectively a node is supposed to spread trade flows through the network. Similarly to the definition of closeness, we define the *communicability closeness* as:

$$C_i = \frac{1}{\sum_{j \in \mathscr{C}} \xi_{ij}} \tag{11}$$

where the sum is over all the internal nodes of the cluster $\mathscr{C}$ to which the node i belongs.

To exemplify, we rank in Fig. 11 (left-hand side) the top 20 countries of community 2 on the basis of values of C_i. It is worth to stress that the centre of this community is located in China, Japan and South Korea and not in the North American subcommunity. The three Asian nations are nowadays major traders, and their high-level economic cooperation has been strengthened also because of the speed-up of the negotiations on the trilateral Free Trade Agreement. The three parties unanimously agreed to further increase the level of trade and investment liberalization based on the consensus reached in the Regional Comprehensive Economic Partnership Agreement.[5]

Moreover, it is interesting to see that most central country in a community has not necessarily the same relevance on the whole network. We have indeed that, in terms of subgraph centrality, when we deal with the whole network (see Fig. 11, right-hand side), USA appears as the key player followed by China and Germany. This ranking is in line with the top three countries provided by the World Trade Organizations, in terms of world's leading traders of goods and services (WTO 2017).

Additionally, it is interesting to highlight that the relevance of countries reported in Fig. 11 (right-hand side) is consistent with the Economic Complexity Index (ECI), introduced by Hausmann et al. (2014). The ECI allows to rank countries in the WTN according to the diversification of their export flows, which reflects the amount of

---

[5] See "Fifteenth Round of Negotiations on a Free Trade Agreement among Japan, China and the Republic of Korea", 12 April 2019, Ministry of Foreign Affairs of Japan and Free Trade Agreement (FTA) and Economic Partnership Agreement (EPA), 4 November 2019, Ministry of Foreign Affairs of Japan.
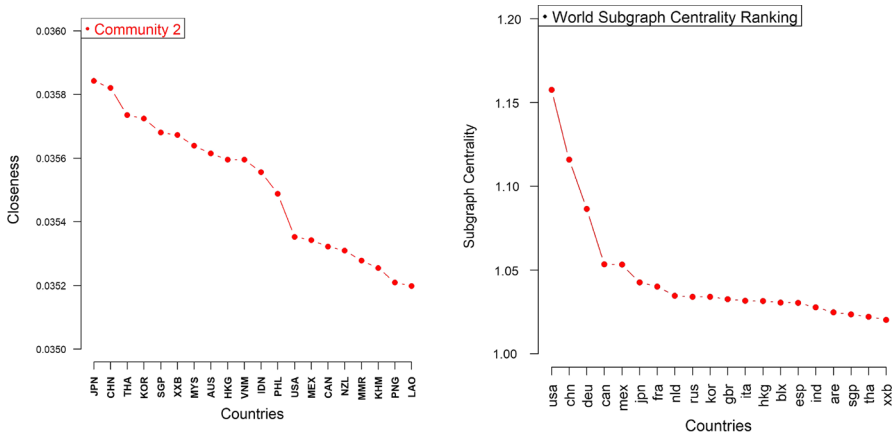
Fig. 11 On the left-hand side, values of communicability closeness C_i for the top 20 countries inside community 2; on the right-hand side, world top 20 countries according to subgraph centrality rankings

knowledge that drives their growth. The higher is the ECI, the more advanced and diversified is an economy. In particular, countries whose economic complexity is greater than expected (on the basis of their global income) tend to grow faster than rich countries with a low ECI. In this perspective, ECI represents a suitable tool for comparing countries in the WTN independently of their total output and it has been extensively validated as a relevant economic measure by showing its capability to predict future economic changes and to explain international differences in countries incomes.

Although the network we analysed in the present work is based on the total normalized output and this fact prevents us from comparing directly their values with the ECI for a given country, there is a positive correlation between them. All the top 20 countries in Fig. 11 (right-hand side) show a positive and high value of ECI. More specifically, they kept a high value of ECI during the years preceding the year to which the network refers (2016) and this can justify the high value in the aforementioned centrality measures.

Finally, from the point of view a single country, it is worth to look for the closest trade partners, that is, the nearest nodes in terms of communicability distance. Figure 12 shows the distance profiles for China and Germany, respectively. For instance, looking at Fig. 12 (right-hand side), we can notice countries, as Austria, Poland, Czech Republic that are characterized by a condition of strong dependence on Germany, that is, a major player in the network. Similarly, Fig. 12 (left-hand side) shows how strong is the commercial relationship between China and Hong Kong, also as a result of the trade agreements between the two countries, like CEPA (Closer Economic Partnership Arrangement) aimed at eliminating duties on large categories of products. Indeed, it is well known that, for the Chinese trade market, Hong Kong plays a crucial role since foreign companies use Hong Kong as a springboard to invest in China thanks to its infrastructure network that has no equal in the world, investor protection, transparent and efficient judicial system, legal certainty.
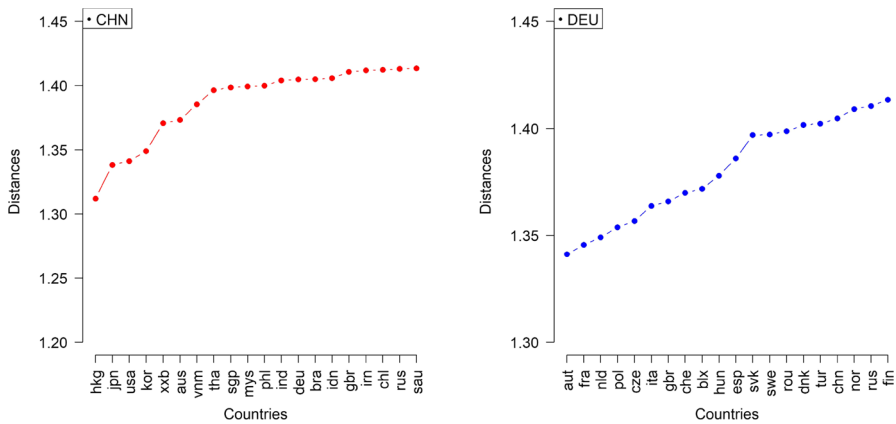
Fig. 12 Top 20 nearest countries for China (left) and Germany (right)

6.3.2 Results in terms of resistance metric

The methodology described in Sect. 6.2 has also been applied using the resistance distance ω. In this case, we consider the total trade of a given country as flow of the global wealth that has been produced during a year. Therefore, the gross domestic product (GDP) is the attribute of interest on each node. In this regard, the effective resistance of an edge expresses how easily (or not) a unit flow moves from a country to another one, i.e. how easily two countries trade a unit of wealth, independently of its nature. It is noteworthy that, according to Formula 8, the resistance distance between a pair of countries depends on the values of the vibrational centralities of both countries (the more central these countries are in the network, the less is the resistance distance between them) and on the value of their mutual correlation (the more correlated they are and again the less is their distance).

Therefore, we construct the vibrational communicability matrix $\mathbf{G}^v$ on the normalized network $\mathscr{G}_2$, and the corresponding resistance distance matrix $\mathbf{\Omega}$. Using this metric, we find that the nearest countries are, again, the USA and Canada with a distance $\omega_{\min} = 1.238$ and the farthest countries are the USA and Germany with a distance $\omega_{\max} = 1.497$. For each value of the threshold distance between minimum and maximum, we obtain the corresponding partition in communities. The maximum partition quality index Q corresponds to 15 communities plus isolated nodes. In Fig. 13, we plot the value of Q in red and the number of communities, counting each isolated node as an independent one, in blue as functions of the threshold ω_h. The maximum quality index Q is reached at a threshold distance $\omega_h = 1.365$. The main characteristic of this partition is the presence of a giant component of 127 nodes e 14 other components with few nodes.

Main results in terms of geographical distribution are displayed in Fig. 14, and as in the previous section, we summarize in Table 3 main composition of top communities in terms of number of constituents.

With respect to results based on communicability, we have that the first community has a larger number of countries (equal to 127). Additionally, the larger community
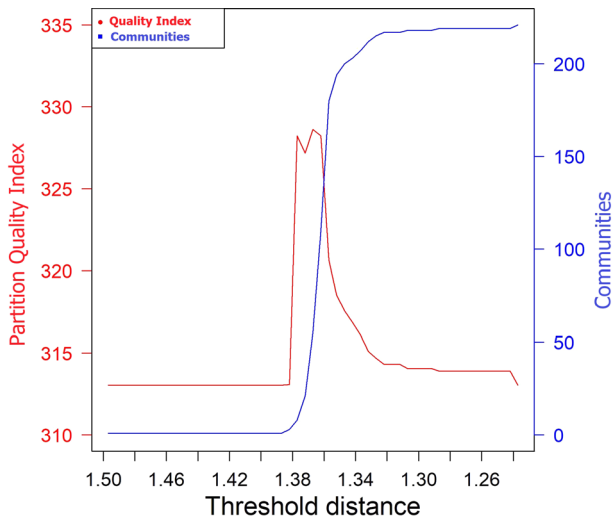
Fig. 13 Partition quality index Q (red line) and number of communities (blue line) as functions of the threshold resistance distance. Maximum Q is observed for $\omega_h = 1.365$

includes again main Asian and Oceanian countries as well as several African countries. It is noteworthy that North America behaves as a separate cluster. This result is in line with the literature that emphasizes the interesting economic relation between Asia and Oceania. Several works showed that the Asia-Oceania community collapsed after China entered the WTO in 2001 and built strong trade relationships with other communities, especially with the external cores (i.e. the USA and Germany). China then became regionally attractive and restored the Asia-Oceania community as the community leader after it gained a significant portion of trade globally (see, for example, Zhu et al. 2014).

Significant differences are also observed for the European community (see community 2 in Table 3). Norway and Sweden and Great Britain and Ireland provide indeed two separate groups with respect to main European economic groups.

It is worth pointing out that communities detected above represent groups of countries showing a positive correlation in their trade strength, whereas members of different clusters show a negative correlation. Being strongly anti-correlated means that when the total trade deficit of a country grows, the total trade surplus of a second country grows too. For instance, Japan and the USA have been classified by the methodology in different communities. Indeed, in the literature, empirical analyses show a negative correlation coefficient between normalized trade strengths of these countries (see, for example, Kozmetsky and Yue 2012; del Rio-Chanona et al. 2017). Similar arguments can be extended also to other pairs of countries. For instance, Germany is negatively correlated with the USA (see Kozmetsky and Yue 2012) and shows a high positive correlation with Belgium and France (see del Rio-Chanona et al. 2017) that belong to the same community.

If we disentangle communities characterized by very tight relationships between countries, the results seem strictly related to the ECI index. We may expect that if two

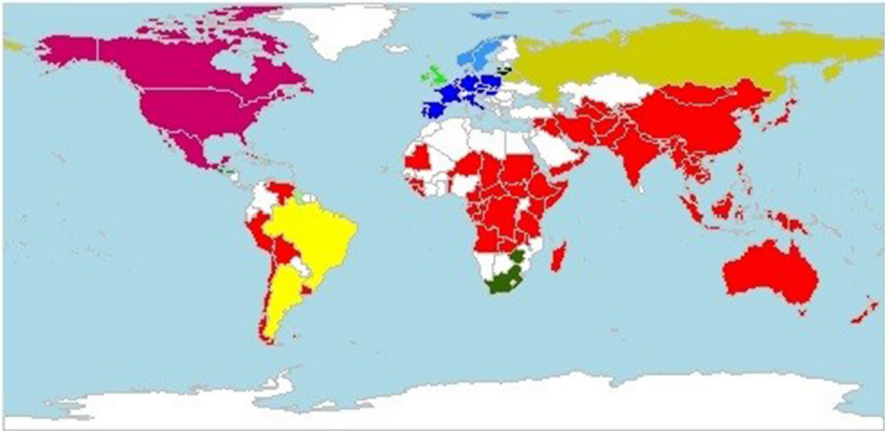## Maximum Quality Index Community Structure

Fig. 14 Communities detected by using the procedure based on the resistance matrix and considering the threshold ω_h that maximize the partition quality index Q

Table 3 Members for the seven main communities in terms of number of countries

	Size	Members
Community 1	127	AUS CHN HKG IDN IND IRN IRQ JPN KOR LAO PHL THA and others
Community 2	11	AUT BLX CZE DEU ESP FRA HUN ITA NLD POL PRT SVK
Community 3	3	CAN MEX USA
Community 4	2	NOR SWE
Community 5	2	GBR IRL
Community 6	2	BLR RUS
Community 7	2	ARG BRA

The optimal partition has been obtained by applying the procedure based on the maximum partition quality index Q and a threshold depending on the resistance distance

countries communicate well, then their ECIs could be similar. That is, if their mutual distance is small, both in terms of communicability metric and resistance metric, then they display similar values of ECI. In fact, the existence of multiple channels of trade exchange between them would result in a similar diversification of their output. This means that countries inside each community (could) share homogeneous values of ECI. Concerning Table 3, we notice small clusters whose components show homogeneous values of the ECI index. For instance, community 6 is formed by Russia (with an ECI of 0.855 in 2016) and Belarus (with an ECI of 0.744 in the same year). Similarly, Canada (1.084), Mexico (1.160) and the USA (1.781); Norway (1.199) and Sweden (1.862); the UK (1.549) and Ireland (1.409); and Brazil (0.648) and Argentina (0.380) constitute communities 3, 4, 5 and 7, respectively.

Table 4 Intercluster and intracluster characteristics of the distributions of resistance distances

	Intracluster Community 1	Community 2	Intercluster Community 1 vs 2
Number of nodes	127	12	—
Mean distance	1.414	1.391	1.418
Min distance	1.290	1.324	1.395
Closest countries	CHN-HKG	AUT-DEU	SMR-DEU
Max distance	1.429	1.424	1.466
Furthest countries	ARE-HKG	AUT-PRT	JPN-DEU
Standard deviation	0.008	0.026	0.010

Columns community 1 and community 2 refer to the intracluster properties of the two main detected communities, in terms of number of nodes. Last column reports the corresponding intercluster properties computed between the same two communities

As in the previous section, we explore main characteristics of two most relevant communities (see Table 4). It is noticeable that although the two groups show a very similar mean distance, European countries are characterized by a higher heterogeneity. Focusing on intercluster indicators, we notice also a lower similarity between the two communities with respect to Table 2 based on communicability.

The relevance of a country can be now assessed in terms of vibrational centrality. To this end, we display in Fig. 15, the top 20 countries, calculated over the whole network. China, the USA and Germany are again in the top 3, with China playing as the best spreader node. Also in this case, almost all the top 20 has a positive ECI. A comparison between Figs. 11 and 15 confirms the different role played by the USA and China in the global network. As confirmed by WTO (2017), the USA is the leading commercial service provider and in such a way, it is widespread well integrated in the global market; on the other side, China plays the role of hub for goods and represents the leading merchandise trader and this gives to the country a very robust position which makes it less vulnerable to market turmoil.

Finally, from the point of view a single country, it is worth to look for the closest trade partners, that is, the nearest nodes in terms of resistance distance. Figure 16 shows the distance profiles for the most central country of communities 1 and 2, respectively. These plots can be interpreted as the list, in decreasing order, of countries that are most positively correlated with the selected centre, China or Germany. For instance, while in terms of communicability distance, China is well communicating with the USA (third position in Fig. 12), the USA does not belong to the top 20 most correlated countries with China. Rather, the left-hand side in Fig. 16 clearly shows a driving and synchronizing effect of the Chinese giant in the entire South-East Asia area. Similarly, Fig. 16 (right-hand side) confirms the role of Germany in the European Union and the strong correlation with Austria, Czech Republic and Poland.

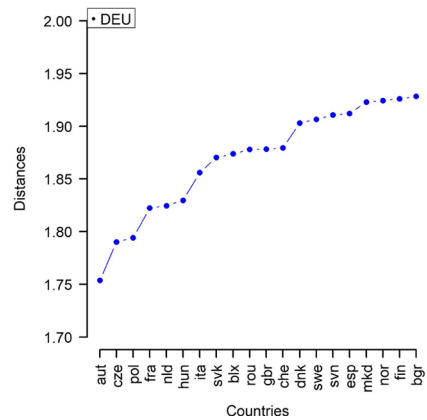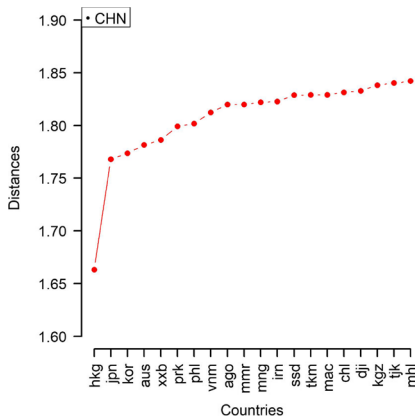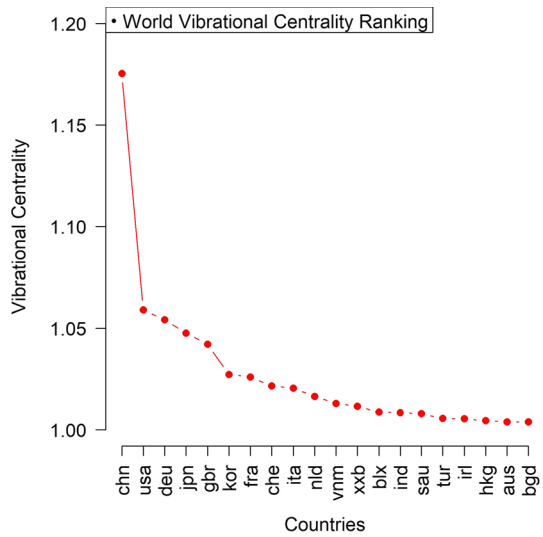**Fig. 15** World top 20 countries according to vibrational centrality rankings

Fig. 16 Top 20 nearest countries for China (left) and Germany (right)

6.4 Comparison with different approaches applied to the same network

It is worth briefly comparing our results with those obtained by other methodologies on the same network (see Barigozzi et al. 2011; Piccardi and Tajoli 2012). In particular, in Piccardi and Tajoli (2012), several approaches are proposed to analyse the community structure of the WTN at different times. The authors showed that the recognition of mesoscale structures is increasingly difficult also because the world is becoming increasingly global over time. This makes even more compelling the search for a method that forces even slight deviations from a random structure to emerge. Both directed and undirected networks have been tested, although no significant differences have been found. As in our case, results reported in Piccardi and Tajoli (2012) show that geographical proximity still matters for international trade, jointly with trade

agreements, common language or religion and traditional partnerships. In particular, focusing on the application of a classical maximum modularity criterion, the authors find in 2008 (the most recent year of their analyses) three big communities containing 68, 66 and 47 countries, with the largest cluster associated with Asia and Oceania. This is partially in line with our result in which a large relevant community including China, Oceania and North America is observed. On the other hand, by using either communicability or resistance distance, we found a higher level of granularity. Additionally, our approach provides a higher flexibility allowing to emphasize stronger connections when the threshold decreases.

Piccardi and Tajoli (2012) also adopt a notion of distance among nodes based on random walks by row-normalizing the weighted matrix. Modelling the WTN by stochastic matrix corresponds to moving from absolute to relative trade values. That distance between nodes is defined by complementing a similarity measure. A dendrogram is computed initially by defining groups containing single nodes and then by iteratively linking pairs of groups with minimal distance. This approach looks similar to ours being based on a varying threshold. They choose to maximize the so-called cophenetic correlation coefficient, which is defined as the linear correlation between the distances and the cophenetic distances, which are the heights of the link joining (directly or indirectly) nodes in the dendrogram.

Some common evidences are noticeable also in this case. The USA and Canada form one of the strongest partnerships: their distance in the dendrogram stays constantly very small over time. France is strongly connected to some of its former colonies, as we also pointed out above, whereas Germany is close to other European countries. Main differences are related to the behaviour of very small countries. While, in our case, small countries are often classified as isolated nodes, in Piccardi and Tajoli (2012), very small countries are connected to much larger ones as an effect of the disassortativity observed in the WTN. These links tend to be small in absolute terms, given the small economic size of the countries, but they appear as relevant in relative terms, because of the strong preference for a given partner.

Piccardi and Tajoli (2012) also used stability and persistence to confirm their results. A random walker starting in a community is likely to remain for quite a long time within that community, before leaving it to enter another one. The analysis of the persistence probabilities induced in a network by a given partition has recently been proven to be an effective tool for testing the existence and significance of communities. Also in this case, we observe that communities with high persistence probability have common features with our results. Indeed, the top communities identified in Piccardi and Tajoli (2012) consider the entire set of European countries, plus a number of minor non-European partners, that is, in line with the top community selected by the communicability approach. Similarly, the second large community with a high persistence probability includes the entire North America and most of Central and South America, plus China, Australia and many others. Although less granular, this community is fully comparable with community 2 detected by the communicability approach.

A quantitative correlation between the world partition in communities obtained by a modularity criterion and geographical distances has been investigated in Barigozzi et al. (2011). The authors, both at an aggregate level and at a number of commodity-

specific levels, compare the two maximum modularity partitions of the input–output network and of the weighted network of the geographical closenesses. They find a high similarity between aggregate trade and geography-based communities, greater than, for instance, communities determined by regional trade agreements. They conclude that geographically related factors explain the patterns of global trade more than political determinants. Although a positive correlation is present between monetary flows and geographical closenesses, we noticed that the geographical distances are less relevant when indirect relationships are also considered via either communicability or resistance distances.[6] As a consequence, the community structure we find appears more granular than the groups found in Barigozzi et al. (2011) and the composition cannot be explained only by geographical patterns. Other factors are involved as historical relationships, trade agreements and strategic economic alliances.

To conclude, although some common results with Barigozzi et al. (2011) and Piccardi and Tajoli (2012) are observed, our methodology has the advantage of clearly highlighting even small differences and forcing the emergence of very strong ties between different countries through the use of a distance threshold. Furthermore, the partition quality index Q we applied turns out to be a simple and flexible tool, more homogeneous to the context of a network interpreted as a metric space.

7 Conclusions and further research

Community detection is a key topic in the analysis of complex systems, where discovering the inner structure plays a relevant role. In particular, the centrality of countries and the relationships between them assume specific relevance in the World Trade Network, where economical and geopolitical phenomena affect over time the structure of the global network. In this framework, this work aimed at detecting different levels of clustered communities in the network on the basis of both communicability and resistance distances. The proposed methodology allows to discover the hidden hierarchical structure of the network, as it presents a degree of flexibility highlighting very tight relationships by varying the threshold parameter, and revealing in this way the clusters of nodes that more easily communicate. Moreover, it performs well also for weighted and extremely dense network, as the case of the WTN.

Features and properties of each community can be exploited in order to compare the characteristics of different clusters and to detect the most central countries inside the single community as well in the whole network.

Numerical results depict the structure of the economic trade detecting main relevant communities. In particular, main community sees the USA and China as main actors. Most flows are polarized around the exchange channel between China and the USA and all their satellite countries. However, focusing on the correlation between trades, the procedure emphasizes the different role of these two countries. In particular, it is worth mentioning the emerging of China-Oceania community when deep links emerge. Furthermore, it is confirmed that Germany plays a key role in Europe and

---

[6] The rank correlation between these distances and the geographical distance between capital cities is lower than 0.15.

preferential channels of internal exchanges are observed in the European market. In line with Zhu et al. (2014), emphasizing tight links, we obtain that although the strong trade relationships with USA and Germany, China became regionally attractive and restored the leadership of Asia-Oceania community. European community is highly centralized around founding members of the European Economic Community with the central role of Germany. High-income countries in Northern Europe are instead in a separate community with a less relevant role in the network.

Appendix A

We report here thorough computations and proofs of Formulae 3, 4 and 5 in Sect. 3.3. The expression 3 of the total potential energy U can be handled in the following way:

$$U = \frac{1}{4}\mathcal{K}\sum_{i,j} A_{ij}[z_i^2 - 2z_i z_j + z_j^2]$$

$$= \frac{1}{4}\mathcal{K}\left[\sum_i z_i^2 \sum_j A_{ij} - 2\sum_{i,j} A_{ij}z_i z_j + \sum_j z_j^2 \sum_i A_{ij}\right]$$

$$= \frac{1}{4}\mathcal{K}\left[\sum_i z_i^2 k_i - 2\sum_{i,j} A_{ij}z_i z_j + \sum_j z_j^2 k_j\right] = \frac{1}{2}\mathcal{K}\left[\sum_i z_i^2 k_i - \sum_{i,j} A_{ij}z_i z_j\right]$$

$$= \frac{1}{2}\mathcal{K}\left[\sum_{ij} z_i (\mathbf{K\text{-}A})_{ij} z_j\right] = \frac{1}{2}\mathcal{K}\left[\sum_{ij} z_i L_{ij} z_j\right].$$

We compute now the expression of the partition function $\mathcal{Z}$ in Formula 4. Using the spectral decomposition of the Laplacian matrix $\mathbf{L} = \mathbf{M\Lambda M}^T$, where $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues and $\mathbf{M}$ is the corresponding matrix of the eigenvectors, we

have the following chain of equalities:

$$\mathcal{Z} = \int e^{-\frac{1}{2}\beta\mathcal{K}\sum_{ij}z_i(\mathbf{M}\mathbf{\Lambda}\mathbf{M}^T)_{ij}z_j} \prod_k dz_k = \int e^{-\frac{1}{2}\beta\mathcal{K}\mathbf{z}^T(\mathbf{M}\mathbf{\Lambda}\mathbf{M}^T)\mathbf{z}} \prod_k dz_k$$

$$= \int e^{-\frac{1}{2}\beta\mathcal{K}(\mathbf{M}^T\mathbf{z})^T\mathbf{\Lambda}(\mathbf{M}^T\mathbf{z})} \prod_k dz_k = \int e^{-\frac{1}{2}\beta\mathcal{K}\mathbf{x}^T\mathbf{\Lambda}\mathbf{x}} \prod_k dx_k$$

$$= \int e^{-\frac{1}{2}\beta\mathcal{K}\sum_k \mu_k x_k^2} \prod_k dx_k$$

where we set $\mathbf{x} = \mathbf{M}^T\mathbf{z}$ and $d\mathbf{z} = |\det \mathbf{M}|d\mathbf{x} = d\mathbf{x}$.

As usual in the literature, we remove the contribution from $\mu_n = 0$, providing the modified partition function we still call $\mathcal{Z}$, (see Estrada 2012b on page 117); this yields:

$$\mathcal{Z} = \prod_{k=1}^{n-1} \int e^{-\frac{1}{2}\beta\mathcal{K}\mu_k x_k^2} dx_k = \prod_{k=1}^{n-1} \sqrt{\frac{2\pi}{\beta\mathcal{K}\mu_k}}$$

the last equality being valid because all integrals are Gaussian with $\mu_k > 0$, $k = 1, \ldots, n-1$.

Finally, we compute $G_{ij}^v(\beta)$ (Formula 5):

$$G_{ij}^v(\beta) = \frac{1}{\mathcal{Z}} \int z_i z_j e^{-\beta U} d\mathbf{z}$$

$$= \frac{1}{\mathcal{Z}} \int z_i z_j e^{-\frac{1}{2}\beta\mathcal{K}\sum_{ij}z_i L_{ij}z_j} \prod_k dz_k$$

$$= \frac{1}{\mathcal{Z}} \int (\mathbf{M}\mathbf{x})_i (\mathbf{M}\mathbf{x})_j e^{-\frac{1}{2}\beta\mathcal{K}\sum_k \mu_k x_k^2} \prod_k dx_k$$

$$= \frac{1}{\mathcal{Z}} \int \left(\sum_{k=1}^{n} \psi_k(i)x_k\right) \left(\sum_{k=1}^{n} \psi_k(j)x_k\right) \prod_{k=1}^{n} e^{-\frac{1}{2}\beta\mathcal{Z}\mu_k x_k^2} dx_k$$

Notice that, computing the product of the two sums inside the integral above, all the integrals involving mixed terms are null, as the integrand is an odd function and the integral is extended to $\mathbb{R}$ for each x_k. Then, only the squared terms remain inside the integral, so that:

$$G_{ij}^v(\beta) = \frac{1}{\mathcal{Z}} \int \left(\psi_1(i)\psi_1(j)x_1^2 + \cdots + \psi_n(i)\psi_n(j)x_n^2\right) \prod_{k=1}^{n} e^{-\frac{1}{2}\beta\mathcal{K}\mu_k x_k^2} dx_k$$

$$= \frac{1}{\mathcal{Z}}\psi_1(i)\psi_1(j) \int x_1^2 e^{-\frac{1}{2}\beta\mathcal{K}\mu_1 x_1^2} dx_1 \cdot \int e^{-\frac{1}{2}\beta\mathcal{K}\mu_2 x_2^2} dx_2 \cdot \cdots \cdot \int e^{-\frac{1}{2}\beta\mathcal{K}\mu_n x_n^2} dx_n + \cdots +$$

$$\frac{1}{\mathcal{Z}}\psi_n(i)\psi_n(j) \int e^{-\frac{1}{2}\beta\mathcal{K}\mu_1 x_1^2} dx_1 \cdot \int e^{-\frac{1}{2}\beta\mathcal{K}\mu_2 x_2^2} dx_2 \cdot \cdots \cdot \int x_n^2 e^{-\frac{1}{2}\beta\mathcal{K}\mu_n x_n^2} dx_n$$

We remove once again the contribution from $\mu_n = 0$; then, computing the integrals, we have:

$$
G_{ij}^v(\beta) = \frac{1}{\mathcal{Z}} \psi_1(i)\psi_1(j) \frac{\sqrt{2\pi}}{\sqrt{(\beta\mathcal{K}\mu_1)^3}} \cdot \frac{\sqrt{2\pi}}{\sqrt{\beta\mathcal{K}\mu_2}} \cdot \cdots \cdot \frac{\sqrt{2\pi}}{\sqrt{\beta\mathcal{K}\mu_{n-1}}} + \cdots +
$$

$$
\frac{1}{\mathcal{Z}} \psi_{n-1}(i)\psi_{n-1}(j) \cdot \frac{\sqrt{2\pi}}{\sqrt{\beta\mathcal{K}\mu_1}} \cdot \frac{\sqrt{2\pi}}{\sqrt{\beta\mathcal{K}\mu_2}} \cdot \cdots \cdot \frac{\sqrt{2\pi}}{\sqrt{(\beta\mathcal{K}\mu_{n-1})^3}}
$$

$$
= \frac{1}{\mathcal{Z}} \prod_{k=1}^{n-1} \sqrt{\frac{2\pi}{\beta\mathcal{K}\mu_k}} \left[\frac{\psi_1(i)\psi_1(j)}{\beta\mathcal{K}\mu_1} + \cdots + \frac{\psi_{n-1}(i)\psi_{n-1}(j)}{\beta\mathcal{K}\mu_{n-1}} \right] = \sum_{k=1}^{n-1} \frac{\psi_k(i)\psi_k(j)}{\beta\mathcal{K}\mu_k}.
$$

Appendix B

The expression of the pseudo-inverse of the Laplacian $\mathbf{L}^+ = \left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right)^{-1} - \frac{1}{n}\mathbf{J}$ allows an interesting interpretation of the resistance distance ω_{ij} in an economic, or financial, networked system.

Suppose that, to each node, a value of a given attribute is assigned through a state vector $\mathbf{v} = [v_1, v_2, \ldots, v_n]^T$ (such an attribute could be, for instance, the GDP of a country or the assets of a financial institution), and let $I_{ij} = v_i - v_j$ be the flow of such an attribute from node i to node j. We denote by I_i the total outgoing flow from the node i to its adjacent nodes, i.e. $I_i = \sum_{j=1}^{n} a_{ij}(v_i - v_j)$.

In matrix form, the total outgoing flow of the nodes attribute is then

$$
\mathbf{I} = (\mathbf{K} - \mathbf{A})v = \mathbf{L}v.
$$

The Laplacian matrix transforms nodes attributes $v_i, i = 1, \ldots, n$ into outgoing flows from nodes I_i, under the assumption that a flow I_{ij} along a given edge is equal to the gradient $\Delta v_{ij} = v_i - v_j$. This assumption is equivalent to choose an effective resistance equal to 1 along all edges. Of course, we may have both outgoing and ingoing currents according to the sign of Δv_{ij}: positive for outgoing flows from i and negative for ingoing flows into i.

A similar meaning can be given to $\left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right) \mathbf{v}$. Indeed,

$$
\left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right) \mathbf{v} = \mathbf{L}\mathbf{v} + \frac{1}{n}\mathbf{J}\mathbf{v} = \mathbf{I} + \bar{v}\mathbf{u},
$$

where $\bar{v} = \frac{1}{n} \sum_{k=1}^{n} v_k$, that is, the operator $\mathbf{L} + \frac{1}{n}\mathbf{J}$ adds to the flows a constant term given by the mean value of all the attributes of the nodes. Then, the matrix $\left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right)$ transforms nodes attributes $\mathbf{v}$ into total outgoing flows $\mathbf{I}$ in the network, up to an additive constant.

In a similar way, the inverse $\left(\mathbf{L} + \frac{1}{n}\mathbf{J} \right)^{-1}$ acts on a current vector $\mathbf{I}$ and produces a state vector $\mathbf{v}$, which can be interpreted as the cause of such currents in the network.

Specifically,

$$\mathbf{v} = \mathbf{L}^+\mathbf{I} = \left[\left(\mathbf{L} + \frac{1}{n}\mathbf{J}\right)^{-1} - \frac{1}{n}\mathbf{J}\right]\mathbf{I} = \left(\mathbf{L} + \frac{1}{n}\mathbf{J}\right)^{-1}\mathbf{I} - \bar{\mathbf{I}}\mathbf{u}$$

where, once again, the term $\frac{1}{n}\mathbf{JI} = \bar{\mathbf{I}}\mathbf{u}$ is the average value of the outgoing currents coming from every node.

Suppose now that in the system, there are an outgoing flow equal to 1 from a node (node 1, for instance), an ingoing flow equal to -1 into another node (for instance, node 2), whereas for all the other nodes, the flow is zero. This is equivalent to a current vector equal to $\mathbf{I} = [1, -1, 0, \ldots, 0]^T = \mathbf{e}_1 - \mathbf{e}_2$. Loosely speaking, a unit information is coming out from node 1 and goes entirely into node 2. To produce these flows, we have to start from an initial attributes vector on nodes given by

$$\mathbf{v} = \mathbf{L}^+(\mathbf{e}_1 - \mathbf{e}_2) = \left[\left(\mathbf{L} + \frac{1}{n}\mathbf{J}\right)^{-1} - \frac{1}{n}\mathbf{J}\right]\begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \left(\mathbf{L} + \frac{1}{n}\mathbf{J}\right)^{-1}\begin{pmatrix} 1 \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where the last equality holds because $\mathbf{J}(\mathbf{e}_1 - \mathbf{e}_2) = \mathbf{0}$, that is $\bar{\mathbf{I}} = 0$. Thus, the resistance distance between nodes 1 and 2 is given by

$$\omega_{12} = (\mathbf{e}_1 - \mathbf{e}_2)^T \left(\mathbf{L} + \frac{1}{n}\mathbf{J}\right)^{-1} (\mathbf{e}_1 - \mathbf{e}_2) = v_1 - v_2 = \Delta v_{12}.$$

If Δv_{12} is small, a small gradient is enough to transmit such a unit flow from node 1 to node 2, whereas if $v_1 - v_2$ is big, a high gradient is needed in order to produce the same unit flow. More in general, let's imagine that in the node i, the value v_i is positive. Then, the fact that another attribute v_j with $j \neq i$ is positive means that node i and node j are strongly correlated since it is enough a low attribute difference to subtract from node i a unit flow. This means that these two nodes communicate a lot. Whereas if for another node k with $k \neq i$, the corresponding component v_k is negative, this implies that node i and node k are strongly anti-correlated since, in order to produce a unit flow from node i, node k has to be at a negative attribute, i.e. the attribute difference between i and k must be high. This means that the two nodes do not communicate well. The signs of the components of the vector $\mathbf{v}$ indicate nodes that are positively or negatively correlated with node i according to the fact these components have the same sign as v_i or not. Let us observe that, in general, $\mathbf{v} = \mathbf{L}^+\mathbf{I} = \mathbf{L}^+(\mathbf{e}_i - \mathbf{e}_j) = L_i^+ - L_j^+$ with L_i^+ ith column of the matrix L^+. That is, if we want to decrease by 1 the attribute of node i and increase by 1 the attribute of node j, we have to take an initial distribution of attributes on nodes equal to the difference between ith column of $\mathbf{L}^+$ and jth column of $\mathbf{L}^+$, and these columns are also the values of vibrational communicability $\mathbf{G}^v$ between nodes, as defined in the text.

References

Abeysinghe T, Forbes K (2005) Trade linkages and output-multiplier effects: a structural var approach with a focus on asia. Rev Int Econ 13(2):356–375

Barigozzi M, Fagiolo G, Mangioni G (2011) Identifying the community structure of the international-trade multi-network. Physica A 390(11):2051–2066

Blöchl F, Theis FJ, Vega-Redondo F, Fisher EO (2011) Vertex centralities in input-output networks reveal the structure of modern economies. Phys Rev E 83(4):046127

Bozzo E (2013) The Moore–Penrose inverse of the normalized graph Laplacian. Linear Algebra Appl 439(10):3038–3043. https://doi.org/10.1016/j.laa.2013.08.039

Cepeda-López F, Gamboa-Estrada F, León C, Rincón-Castro H (2019) The evolution of world trade from 1995 to 2014: a network approach. J Int Trade Econ Dev 28(4):452–485

Cerqueti R, Ferraro G, Iovanella A (2018) A new measure for community structure through indirect social connections. Expert Syst Appl 114:196–209

Cerqueti R, Clemente GP, Grassi R (2019) A network-based measure of the socio-economic roots of the migration flows. Soc Indic Res. https://doi.org/10.1007/s11205-018-1883-6

Chang C, Liao W, Chen Y, Liou L (2016) A mathematical theory for clustering in metric spaces. IEEE Trans Netw Sci Eng 3(1):2–16

Clemente GP, Cornaro A (2019) A novel measure of edge and vertex centrality for assessing robustness in complex networks. Soft Comput 24:3687–13704

Clemente GP, Fattore M, Grassi R (2018) Structural comparisons of networks and model-based detection of small-worldness. J Econ Interact Coord 13(1):117–141

De Benedictis L, Tajoli L (2011) The world trade network. World Econ 34(8):1417–1454

De Benedictis L, Tajoli L (2016) Comparative advantage and centrality in the world network of trade and value added: an analysis of the Italian Position. Riv Polit Econ 66(3):537–554

Dées S, Saint-Guilhem A (2011) The role of the united states in the global economy and its evolution over time. Empir Econ 41(3):573–591

Del Rio-Chanona RM, Grujic J, Jeldtoft Jensen H (2017) Trends of the world input and output network of global trade. PLoS ONE 12(1):1–14. https://doi.org/10.1371/journal.pone.0170817

Ellens W, Spieksma F, Van Mieghem P, Jamakovic A, Kooij R (2011) Effective graph resistance. Linear Algebra Appl 435:2491–2506

Estrada E (2012) Complex networks in the euclidean space of communicability distances. Phys Rev E 85:066122. https://doi.org/10.1103/PhysRevE.85.066122

Estrada E (2012) The structure of complex networks: theory and applications. Oxford University Press, Oxford

Estrada E, Hatano N (2008) Communicability in complex networks. Phys Rev E 77:036111. https://doi.org/10.1103/PhysRevE.77.036111

Estrada E, Hatano N (2009) Communicability graph and community structures in complex networks. Appl Math Comput 214(2):500–511. https://doi.org/10.1016/j.amc.2009.04.024

Estrada E, Hatano N (2010) Resistance distance, information centrality, node vulnerability and vibrations in complex networks. Springer, London

Estrada E, Hatano N (2010) A vibrational approach to node centrality and vulnerability in complex networks. Phys A Stat Mech Appl 389(17):3648–3660. https://doi.org/10.1016/j.physa.2010.03.030

Estrada E, Rodriguez-Velazquez JA (2005) Subgraph centrality in complex networks. Phys Rev E Stat Nonlinear Soft Matter Phys 71:056103. https://doi.org/10.1103/PhysRevE.71.056103

Fagiolo G (2007) Clustering in complex directed networks. Phys Rev E. https://doi.org/10.1103/physreve.76.026107

Fagiolo G, Reyes J, Schiavo S (2008) On the topological properties of the world trade web: a weighted network analysis. Physica A 387(15):3868–3873

Fagiolo G, Reyes J, Schiavo S (2010) The evolution of the world trade web: a weighted-network analysis. J Evolut Econ 20(4):479–514

Fagiolo G, Squartini T, Garlaschelli D (2013) Null models of economic networks: the case of the world trade web. J Econ Interact Coord 8(1):75–107

Fagiolo G, Victor JN, Lubell M, Montgomery A (2015) The international trade network: empirics and modeling. In: The Oxford handbook of political networks, pp. 173–193

Fan Y, Ren S, Cai H, Cui X (2014) The state's role and position in international trade: a complex network perspective. Econ Model 39:71–81

Ferraz de Arruda G, Luiz Barbieri A, Rodríguez PM, Rodrigues FA, Moreno Y, da Fontoura Costa L (2014) The role of centrality for the identification of influential spreaders in complex networks. Phys Rev E 90:032812

Fortunato S (2010) Community detection in graphs. Phys Rep 486(3–5):75–174

Fortunato S, Hric D (2016) Community detection in networks: a user guide. Phys Rep 659:1–44

Garlaschelli D, Loffredo MI (2004) Fitness-dependent topological properties of the world trade web. Phys Rev Lett 93(18):188701

Garlaschelli D, Loffredo MI (2005) Structure and evolution of the world trade network. Physica A 355(1):138–144

Garlaschelli D, Di Matteo T, Aste T, Caldarelli G, Loffredo MI (2007) Interplay between topology and dynamics in the world trade web. Eur Phys J B 57(2):159–164

Giudici P, Spelta A (2016) Graphical network models for international financial flows. J Bus Econ Stat 34(1):128–138

Gutman I, Xiao W (2004) Generalized inverse of the Laplacian matrix and some applications. Bulletin (Académie serbe des sciences et des arts. Classe des sciences mathématiques et naturelles. Sciences mathématiques), pp 15–23

Hausmann R, Hidalgo CA, Bustos S, Coscia M, Simoes A, Yildirim MA (2014) The atlas of economic complexity: mapping paths to prosperity. MIT Press, Cambridge

Kali R, Reyes J (2007) The architecture of globalization: a network approach to international economic integration. J Int Bus Stud 38(4):595–620

Klein D, Randic M (1993) Resistance distance. J Math Chem 12:81–95. https://doi.org/10.1007/BF01164627

Kozmetsky G, Yue P (2012) Global economic competition: today's warfare in global electronics industries and companies. Springer, New York

Kuznetsova NV, Kocheva EV, Matev NA (2016) The analysis of foreign trade activities of Russia and Asia-Pacific region. Int J Econ Financ Issues 6(2):736–744

Lancichinetti A, Fortunato S (2009) Community detection algorithms: a comparative analysis. Phys Rev E 80(5):056117

Lee C, Tenneti S, Eun DY (2019) Transient dynamics of epidemic spreading and its mitigation on large networks. CoRR arXiv:1903.00167

Li X, Jin YY, Chen G (2003) Complexity and synchronization of the World trade web. Physica A 328(1–2):287–296

Lidth Van, de Jeude J, Di Clemente R, Caldarelli G, Saracco F, Squartini T (2019) Reconstructing mesoscale network structures. Complexity. https://doi.org/10.1155/2019/5120581

Newman ME (2004) Fast algorithm for detecting community structure in networks. Phys Rev E 69(6):066133

Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. Phys Rev E 69(2):026113

Nicolini C, Bordier C, Bifone A (2017) Community detection in weighted brain connectivity networks beyond the resolution limit. Neuroimage 146:28–39

Piccardi C, Tajoli L (2012) Existence and significance of communities in the world trade web. Phys Rev E 85:066119. https://doi.org/10.1103/PhysRevE.85.066119

Piccardi C, Tajoli L (2018) Complexity, centralization, and fragility in economic networks. PLoS ONE 13(11):1–13. https://doi.org/10.1371/journal.pone.0208265

Rattigan MJ, Maier M, Jensen D (2007) Graph clustering with network structure indices. In: Proceedings of the 24th international conference on Machine learning. ACM, pp 783–790

Reyes J, Schiavo S, Fagiolo G (2008) Assessing the evolution of international economic integration using random walk betweenness centrality: the cases of East Asia and Latin America. Adv Complex Syst 11(05):685–702

Schiavo S, Reyes J, Fagiolo G (2010) International trade and financial integration: a weighted network analysis. Quant Finance 10(4):389–399

Serrano MA, Boguñá M (2003) Topology of the world trade web. Phys Rev E 68(1):015101

Serrano MA, Boguñá M, Vespignani A (2007) Patterns of dominant flows in the world trade web. J Econ Interact Coord 2(2):111–124

Smith DA, White DR (1992) Structure and dynamics of the global economy: network analysis of international trade 1965–1980. Soc Forces 70(4):857–893

Snyder D, Kick EL (1979) Structural position in the world system and economic growth, 1955–1970: a multiple-network analysis of transnational interactions. Am J Sociol 84(5):1096–1126

Traag VA, Aldecoa R, Delvenne J (2015) Detecting communities using asymptotical surprise. Phys Rev E 92(2):022816

Tzekina I, Danthi K, Rockmore DN (2008) Evolution of community structure in the world trade web. Eur Phys J B 63(4):541–545

Van Mieghem P, Devriendt K, Cetinay H (2017) Pseudoinverse of the Laplacian and best spreader node in a network. Phys Rev E. https://doi.org/10.1103/PhysRevE.96.032311

Van Berkum S (2013) Trade effects of the EU-Morocco Association Agreement. 2013-070. LEI, onderdeel van Wageningen UR

Varela LM, Rotundo G, Ausloos M, Carrete J (2015) Complex network analysis in socioeconomic models. In: Complexity and geographical economics. Springer, pp 209–245

Wang X, Pournaras E, Kooij R, Van Mieghem P (2014) Improving robustness of complex networks via the effective graph resistance. Eur Phys J B 87(9):221

Ward MD, Ahlquist JS, Rozenas A (2013) Gravity's rainbow: a dynamic latent space model for the world trade network. Netw Sci 1(1):95–118

Wilhite A (2001) Bilateral trade and small-world networks. Comput Econ 18(1):49–64

WTO (2017) World trade statistical review. Technical report. World Trade Organizations

Zhu Z, Cerina F, Chessa A, Caldarelli G, Riccaboni M (2014) The rise of China in the international trade network: a community core detection approach. PLoS ONE 9(8):e105496