Under consideration for publication in Network Science

The Similarity of Global Value Chains: A Network-Based Measure

ZHEN ZHU
IMT School for Advanced Studies Lucca, Italy
(*e-mail:* zhen.zhu@imtlucca.it)

GREG MORRISON
University of Houston, USA
(*e-mail:* gcmorris@central.uh.edu)

MICHELANGELO PULIGA
IMT School for Advanced Studies Lucca, Italy
(*e-mail:* michelangelo.puliga@imtlucca.it)

ALESSANDRO CHESSA
IMT School for Advanced Studies Lucca, Italy
(*e-mail:* alessandro.chessa@imtlucca.it)

MASSIMO RICCABONI
IMT School for Advanced Studies Lucca, Italy
and
KU Leuven, Belgium
(*e-mail:* massimo.riccaboni@imtlucca.it)

Abstract

International trade has been increasingly organized in the form of global value chains (GVCs) where different stages of production are located in different countries. This recent phenomenon has substantial consequences for both trade policy design at the national or regional level and business decision making at the firm level. In this paper, we provide a new method for comparing GVCs across countries and over time. First, we use the World Input-Output Database (WIOD) to construct both the upstream and downstream global value networks, where the nodes are individual sectors in different countries and the links are the value-added contribution relationships. Second, we introduce a network-based measure of node similarity to compare the GVCs between any pair of countries for each sector and each year available in the WIOD. Our network-based similarity is a better measure for node comparison than the existing ones because it takes into account all the direct and indirect relationships between country-sector pairs, is applicable to both directed and weighted networks with self-loops, and takes into account externally defined node attributes. As a result, our measure of similarity reveals the most intensive interactions among the GVCs across countries and over time. From 1995 to 2011, the average similarity between sectors and countries have clear increasing trends, which are temporarily interrupted by the recent economic crisis. This measure of the similarity of GVCs provides quantitative answers to important questions about dependency, sustainability, risk, and competition in the global production system.

Keywords: *Networks, Node Similarity, Input-Output Analysis, Global Value Chains, Vertical Specialization, International Trade*

1 Introduction

International trade has been increasingly characterized by the content of intermediate inputs (Johnson & Noguera, 2012; Johnson, 2014a) and by the formation of *global value chains* (GVCs) (Feenstra & Hanson, 1999; Hummels *et al.*, 2001; Grossman & Rossi-Hansberg, 2008; Costinot *et al.*, 2013; Amador & Cabral, 2014; Baldwin & Lopez-Gonzalez, 2014; Koopman *et al.*, 2014; Los *et al.*, 2014). Thanks to the development of transportation, information, and communications technologies, different stages of production can be allocated and coordinated across borders. For instance, merely 3% of the total value-added of China's exports of iPhones and laptop computers in 2009 is sourced from China itself, while the remaining 97% is from other countries such as the United States, Japan, and South Korea (Xing, 2014).

The notion of GVCs has been useful in capturing the fact that different stages of production are organized across multiple countries, but the global production sharing at micro level (e.g., for a certain product such as iPhone) can be performed in a wide range of configurations, including a chain (or 'snake'), star (or 'spider'), or any network topology in between (Baldwin & Venables, 2013). More importantly, at the aggregated sector level the GVCs are necessarily embedded in a global production network, where significant value-added contributions flow between sectors located in different countries.

With *global multi-regional input-output* (GMRIO) tables becoming available (Tukker & Dietzenbacher, 2013), the phenomenon of GVCs has been explored extensively in recent years by both theoretical modeling (Grossman & Rossi-Hansberg, 2008; Baldwin & Venables, 2013; Costinot *et al.*, 2013) and empirical measurements (Feenstra & Hanson, 1999; Hummels *et al.*, 2001; Johnson & Noguera, 2012; Baldwin & Lopez-Gonzalez, 2014; Johnson, 2014a; Los *et al.*, 2014; Timmer *et al.*, 2014; Zhu *et al.*, 2015). Previous studies can tell us how *global* the GVCs are by measuring the foreign value-added content of exports for a given sector or country, but not how *similar* they are. The answer to the latter question is essential for our understanding of the dependency and competition between the GVCs. And a good measure of the similarity between the GVCs should take into account the intricate connections between sectors and adopt a network perspective.

To fill the gaps in the literature, we introduce a network-based measure of similarity between the GVCs. Decades of literature has implemented measures of *structural equivalence* between nodes, with equivalent nodes strongly connected to the same neighbors (Lü & Zhou, 2011; Lerner, 2005). More recent work has focused on the concept of *role equivalence*, which relaxes the constraint that equivalent nodes depend on the identical neighbors and requires instead that they depend on other equivalent nodes (Lü & Zhou, 2011; Jeh & Widom, 2002; Leicht *et al.*, 2006; Everett & Borgatti, 1988). Role equivalence gives a more generalized sense of the relationship between nodes by defining equivalence in a self-consistent fashion, but many of these approaches are defined only for undirected or unweighted networks and do not incorporate externally-defined node attributes (e.g., country or sector information that is available in the WIOD). In this paper, we develop a measure to identify the most intensive interactions among the GVCs across countries and over time incorporating the full network topology.

Our paper is also related to the longstanding literature on export similarity. Since the seminal work of Finger and Kreinin (Finger & Kreinin, 1979), multiple measures of similarity have been introduced in the empirical study of international trade to calculate the overlap between the distributions of exports or imports by commodity groups of two countries to the market of third countries (Lloyd, 2004). However, traditional measures of export similarity do not take into account the fragmentation of global production, which accounts for two-thirds of international trade (Johnson & Noguera, 2012). Therefore, our work contributes to recent efforts to advancing the construction of globalization indices (Martens *et al.*, 2015). In particular, we go beyond traditional measures of trade openness and diversification by considering both direct and indirect links and both domestic and foreign transactions in the global system of trade and production.

To the best of our knowledge, our paper is the first attempt to measure and compare the GVCs at the sector level from a network-based approach. First, from a complex networks perspective, we map the World Input-Output Database (WIOD) (Timmer *et al.*, 2015) into both the upstream and downstream global value networks (GVNs), where the nodes are the individual sectors in different countries and the links are the value-added contribution relationships. Second, we introduce a network-based measure of node similarity to compare the GVCs between any pair of countries for each sector and each year available in the WIOD. Unlike the previous methods, we take into account all the direct and indirect relationships to calculate the GVCs similarity, which provides a more accurate and systemic comparison between the GVCs in space and time.

We find that on average sectors have become more similar over time and the trend was temporarily interrupted by the 2008 crisis where the upstream GVCs experienced a much larger reduction than the downstream ones. Among all the similarity measures considered in the paper, ours is the only one that captures that the upstream and downstream GVCs tend to more closely follow the same growth path and that both experience a reduction after the crisis. Therefore, we contribute to a better understanding of what happened during the *great trade collapse* in the aftermath of the 2008 crisis (Ferrantino *et al.*, 2014). In particular, we detect not only the overall reduction of the GVCs similarity, but also the difference in magnitude between the upstream and downstream GVCs. Moreover, we find that manufacturing sectors tend to be more similar with each other than the services sectors and that countries like China has increased its average similarity over time. As a case study, we also find that the sector of electrical equipment in China has become upstream-similar to the one in Czech Republic and downstream-similar to the one in Taiwan. By measuring the increasing similarity between the GVCs, our method has nonetheless led to a warning about the systemic risk of the global production system (Acemoglu *et al.*, 2012). For instance, the effect of trade on business cycle sychronization must take into account the ubiquity of GVCs (Kose & Yi, 2001). Finally, in the field of complex networks, our measure of similarity may also provide valuable insights into node clustering or community detection (Morrison & Mahadevan, 2012; Girvan & Newman, 2002; Zhou, 2003; Piccardi *et al.*, forthcoming), link prediction (Lü & Zhou, 2011; Lü *et al.*, 2009; Zhou *et al.*, 2009), and block modeling (Lerner, 2005; Richeardt & White, 2007; Guimerá & Amaral, 2005).

The rest of the paper is structured as follows. Section 2 describes the WIOD and constructs both the upstream and downstream GVNs and introduces the network-based measure of GVCs similarity. Section 3 summarizes and discusses the results and Section 4 concludes the paper.

2 Data and Methods

A network can be broadly defined as a set of items (nodes) and the connections between them (edges) (Albert & Barabási, 2002; Newman, 2003). Recent years have witnessed a burgeoning body of research exploring topics in economics and finance from a network perspective (Pammolli & Riccaboni, 2002; Fagiolo *et al.*, 2009; Schweitzer *et al.*, 2009; Kitsak *et al.*, 2010; Riccaboni & Schiavo, 2010; Caldarelli *et al.*, 2013; Riccaboni *et al.*, 2013; Elliott *et al.*, 2014; Zhu *et al.*, 2014). The set of sectors and the input-output relationships between them can also be considered as an interdependent network (Blöchl *et al.*, 2011; Acemoglu *et al.*, 2012; Cerina *et al.*, 2015). In this section we first map the WIOD into both the upstream and downstream global value networks (GVNs), where the nodes are the individual sectors in different countries and the links are the value-added contribution relationships. Notice that the GVNs are both directed (i.e., links going from value-added provider sectors to receiver sectors) and weighted (i.e., the share of value-added contribution varies from one link to another). As a result, the upstream (or downstream) value system of a sector can be obtained by searching for all the direct and indirect incoming (or outgoing) neighbors of the given sector in the upstream (or downstream) GVN. We then propose a measure of GVC similarity that is applicable to the both directed and weighted GVNs with externally defined node attributes (the country and sector of the node) so that we can quantify how similar the GVCs are between any pair of countries for each sector and each year available in the WIOD.

2.1 Data

We use the recently available GMRIO database, the WIOD, to investigate the GVCs at the sector level (Tukker & Dietzenbacher, 2013). At the time of writing, the WIOD input-output tables cover 35 sector classifications for each of the 41 countries (one of them is actually the rest of the world, “RoW”) and the years from 1995 to 2011. Hence, for each year, we define $N_s = 35$ as the number of sector classifications and $N_c = 41$ as the number of countries and totally we have $N_c \times N_s = 1435$ sectors (or country-sector pairs). For the rest of the paper, we use *country-sector pair* and *sector* interchangably to mean an individual node in the network. Table A1 and Table A2 list the countries and sector classifications covered in the WIOD. For each year, there is a harmonized international input-output table listing the input-output relationships between any two country-sector pairs. The numeric data in the WIOD are in current basic (producers’) prices and are expressed in millions of US dollars.

A GMRIO table can be compactly presented as

$$\begin{bmatrix} \mathbf{Z} & \mathbf{f} & \mathbf{x} \\ \mathbf{v}^T & & \\ \mathbf{x}^T & & \end{bmatrix}$$

where uppercase boldface letters denote matrices, lowercase boldface letters denote vectors, and the uppercase letter T denotes the transpose operator. Below we explain each component in detail.

The input-output flows between sectors is called the transactions matrix and is often denoted by

$$\mathbf{Z} = \begin{bmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & \ddots & \vdots \\ Z_{n1} & \cdots & Z_{nn} \end{bmatrix}$$

where Z_{ij} denotes the amount of material flow supplied from sector i to sector j, and $n = N_c \times N_s = 1435$ is the total number of sectors in the WIOD. In other words, each row of $\mathbf{Z}$ is the distribution of a sector's outputs throughout the world, while each column of $\mathbf{Z}$ is the distribution of a sector's inputs from the world. Note that sectors often buy a significant portion of inputs from themselves due to the highly aggregated sector classification. As a result, the diagonal of $\mathbf{Z}$ typically has positive and large numbers. Besides intermediate sector use, the remaining outputs are absorbed by an additional column of final demand, which includes household consumption, government expenditure, etc., and is defined as $\mathbf{f}^T = \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix}$, where f_i denotes the final demand of sector i's output. Similarly, production necessitates not only inter-sectoral transactions but also other value-added items such as labor, management, depreciation of capital, and taxes, which is defined as $\mathbf{v}^T = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$, where v_i denotes the value-added supplied to sector i. Finally, the total output vector is defined as $\mathbf{x}^T = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}$, where x_i denotes the total output produced by sector i.

2.2 Construct the Global Value Networks

For a given sector i, its total output equals both the sum of its sales to the sectors in the world and to final demand, and to the sum of its expenditures on inputs from the sectors in the world and on other value-added items. That is

$$x_i = Z_{i1} + \cdots + Z_{in} + f_i = Z_{1i} + \cdots + Z_{ni} + v_i.$$

In matrix form, it is easily seen that $\mathbf{Ax} + \mathbf{f} = \mathbf{x}$, where the so-called technical coefficient matrix is defined as

$$\mathbf{A} = \mathbf{Z}\hat{\mathbf{x}}^{-1} = \begin{bmatrix} Z_{11} & \cdots & Z_{1n} \\ \vdots & \ddots & \vdots \\ Z_{n1} & \cdots & Z_{nn} \end{bmatrix} \begin{bmatrix} \frac{1}{x_1} & & \\ & \ddots & \\ & & \frac{1}{x_n} \end{bmatrix} = \begin{bmatrix} \frac{Z_{11}}{x_1} & \cdots & \frac{Z_{1n}}{x_n} \\ \vdots & \ddots & \cdots \\ \frac{Z_{n1}}{x_1} & \cdots & \frac{Z_{nn}}{x_n} \end{bmatrix}$$

and a 'hat' $\wedge$ over a vector is defined as the operation to produce a diagonal matrix with the elements of the vector on its diagonal.

Solving $\mathbf{x}$ in terms of $\mathbf{f}$ and $\mathbf{A}$, we get $\mathbf{x} = (\mathbf{I}-\mathbf{A})^{-1}\mathbf{f}$, where $\mathbf{I}$ is an identity matrix. The so-called Leontief inverse (Leontief, 1936; Miller & Blair, 2009) is $\mathbf{L} = (\mathbf{I}-\mathbf{A})^{-1}$, where its element L_{ij} measures the change of output in sector i due to a one-unit change in final demand for sector j.

We further define the value-added share vector as

$$\mathbf{w} = \hat{\mathbf{x}}^{-1}\mathbf{v} = \begin{bmatrix} \frac{1}{x_1} & & \\ & \ddots & \\ & & \frac{1}{x_n} \end{bmatrix} \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \frac{v_1}{x_1} \\ \vdots \\ \frac{v_n}{x_n} \end{bmatrix}$$

so that the value-added contribution matrix can be computed as

$$\mathbf{G} = \hat{\mathbf{w}}\mathbf{L}\hat{\mathbf{f}} \tag{1}$$

where $\mathbf{G}$ is the value-added contribution matrix and its element G_{ij} is sector i's value-added contribution to sector j's total final demand, f_j. The upstream value-added share matrix, $\mathbf{U}$, is defined as the column-normalized version of $\mathbf{G}$,

$$\mathbf{U} = \mathbf{G}(\widehat{\mathbf{G}^T\mathbf{1}})^{-1} \tag{2}$$

where the element U_{ij} is sector i's share of value-added contribution out of sector j's total final demand, f_j, and $\mathbf{1}$ is a vector of all 1's of conformable length. The downstream value-added share matrix, $\mathbf{D}$, is similarly defined as the row-normalized version of $\mathbf{G}$:

$$\mathbf{D} = (\widehat{\mathbf{G}\mathbf{1}})^{-1}\mathbf{G} \tag{3}$$

where the element D_{ij} is sector j's share out of sector i's total value-added contribution. Note that the sum of each column of $\mathbf{U}$ is 1 while the sum of each row of $\mathbf{D}$ is 1. $\mathbf{U}$ contains the upstream information because it identifies the shares of the value-added providers for any given sector. And $\mathbf{D}$ contains the downstream information because it identifies the shares of the value-added receivers for any given sector. Finally, the upstream GVNs are constructed by using $\mathbf{U}$ as the weight matrix while the downstream GVNs can be constructed with $\mathbf{D}$ as the weight matrix. Notice that the GVNs are directed, weighted, and contain self-loops.

2.3 A Network-Based Measure of Node Similarity

A wide range of similarity measures between nodes in a complex network have been developed recently (Lü & Zhou, 2011) that could potentially be used to determine similar nodes in the GVNs. The simplest of these that are applicable to weighted networks include those defined by a comparison of the overlap of direct providers, with prominent examples including the weighted Jaccard coefficient (Ioffe, 2010) or cosine similarity (Lü & Zhou, 2011) between a pair of nodes P and Q (with each node representing a country-sector pair). These measures are respectively defined as

$$J_{PQ} = \sum_{cs} \min(p_{cs}, q_{cs}) / \sum_{cs} \max(p_{cs}, q_{cs}) \tag{4}$$

and

$$C_{PQ} = \sum_{cs} p_{cs} q_{cs} / [(\sum_{cs} p_{cs}^2)(\sum_{cs} q_{cs}^2)]^{1/2} \tag{5}$$

where p_{cs} is the dependence of P on the country-sector cs (and a similar definition for q_{cs}) and the summation runs over all countries c and all sectors s that either P or Q depend on (i.e., all c and s for which $p_{cs} > 0$ or $q_{cs} > 0$). In this paper, we will focus on a different but related similarity measure defined by

$$S_{PQ}^{(0)} = \frac{\sum_{cs} [p_{cs}^2 + q_{cs}^2 - (p_{cs} - q_{cs})^2]}{\sum_{cs} [p_{cs}^2 + q_{cs}^2 + (p_{cs} - q_{cs})^2]}. \tag{6}$$

J_{PQ}, C_{PQ} and $S_{PQ}^{(0)}$ share a number of desirable properties in common: they are all strictly bounded between 0 and 1, with the value 0 attained iff P and Q have no providers in common and the value 1 attained iff P and Q receive from the same nodes by an identical amount. We further show in the Appendix this definition is strongly related to the definition of the weighted Jaccard Coefficient and differs from the cosine similarity only by a different normalization. The general characteristics of these local measures of similarity are schematically diagrammed in Fig. 1 (A) for a hypothetical dependency network of German Construction (node P) and Italian Construction (node Q). For all the three measures, only identical dependencies provide a contribution to the similarity between P and Q. In this hypothetical example, $J_{PQ} = 0.25$, $C_{PQ} \approx 0.647$, and $S_{PQ}^{(0)} = 4/9$.

While purely local measures of similarity have been implemented in a wide range of studies, they are too limited to fully understand the relationship between national production systems because upstream providers that are ‘similar’ but not identical contribute nothing to the measure of similarity between P and Q. More meaningful information about the similarity between two production systems can be extracted by defining a measure of role equivalence (Lü & Zhou, 2011; Jeh & Widom, 2002; Leicht *et al.*, 2006) which implements a more self-consistent measure of similarity.

Many of these self-consistent measures of similarity are applied to undirected networks (Leicht *et al.*, 2006; Jeh & Widom, 2002). These approaches define the similarity between two nodes in terms of the similarities of their neighbors, with recursive methods for computing the similarities between nodes. A similar definition for similarity can be applied to directed networks (Blondel *et al.*, 2004) which takes the similarity between the country-sector pairs P and Q as the sum of the similarities of their upstream and downstream neighbors. The algorithm was originally defined in terms of an unweighted directed network, and we implement their algorithm in a weighted network as $X_{PQ} \propto \sum_{P'Q'}(G_{PP'}G_{QQ'} + G_{P'P}G_{Q'Q})X_{P'Q'}$, with the normalization that $\sum_{PQ} X_{PQ}^2 = 1$. This equation can be solved iteratively by computing

$$\mathbf{X}^k = \frac{\mathbf{G}^T\mathbf{X}^{k-1}\mathbf{G} + \mathbf{G}\mathbf{X}^{k-1}\mathbf{G}^T}{||\mathbf{G}^T\mathbf{X}^{k-1}\mathbf{G} + \mathbf{G}\mathbf{X}^{k-1}\mathbf{G}^T||_F} \tag{7}$$

for an even number of times (Blondel *et al.*, 2004), with $||\cdot||_F$ the Frobenius norm and $\mathbf{G}$ the value-added contribution matrix defined above, and we implement the recursive algorithm in Eq. 7 modified for weighted networks in each year, halting after 100 iterations for each computation. This approach to measuring the similarity between nodes differs from ours defined in Eq. 9 below in a number of ways, including the fact that the self-similarity of each country is not normalized (i.e. $X_{PP} \neq 1$), there is no distinction between a country and sector in the measure of similarity, and that upstream and downstream links are incorporated simultaneously.

Therefore, existing methods of measuring role equivalence may not be appropriate for the study of the GVCs, because the attributes of each node in the network cannot be treated on an equal footing. One might expect that a country-sector pair could change the nationality of its provider (for example, German construction exchanging its direct input from French construction to the construction sector in another country), but not change the sector of the input (German construction could not replace its French construction input to another industrial sector, regardless of the country of origin). The differing economic meanings behind the node attributes suggest that we develop a measure of similarity that explicitly takes these attributes into account (as in S_{PQ} defined below).

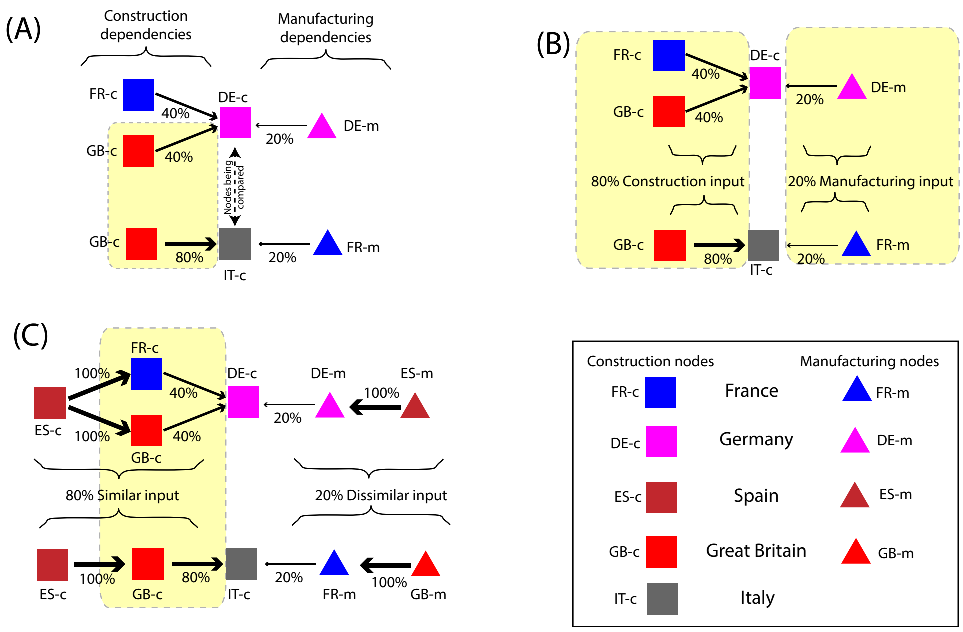

Fig. 1. Schematic diagrams of the methods of measuring the similarity between two nodes in the GVNs, with hypothetical dependencies of the German and Italian Construction sectors (P = DE-c and Q = IT-c respectively) shown. Construction sectors are shown as squares and manufacturing sectors as triangles, while countries are represented by color (France is blue, Germany cyan, Spain brown, Britain red, and Italy gray). Dependency links that provide a significant contribution to the similarity between DE-c and IT-c are highlighted in yellow. In (A), we diagram structural similarity using purely local dependency information (as in J_{PQ}, C_{PQ}, and $S^{(0)}_{PQ}$), with the similarity between DE-c and IT-c due solely to the overlap between the identical provider of British Construction (GB-c). In (B), we show the sectoral dependency of the nodes are assumed identical (captured in $S^{(1)}_{PQ}$), so all links contribute to the similarity if national differences are ignored. (C) shows an interpolation between these two extremes, where all upstream construction links for both DE-c and IT-c have the same provider (ES-c), making these providers similar, but the manufacturing links for DE-c and IT-c have different providers.

The definition of $S^{(0)}_{PQ}$ in Eq. 6 represents a lower bound on any meaningful definition of role equivalence between country-sector pairs, because it treats each distinct national production system as *completely* different. We can define an upper bound for similarity in a related manner by assuming that national systems of production are all completely identical instead of being completely distinct. This approximation is schematically diagrammed in Fig. 1 (B), where sectors of production are considered distinct (Construction and Manufacturing are different fields) but national identities are treated as irrelevant. A measure of similarity equivalent to that in Eq. 6 can be developed in this approximation, with

$$\begin{aligned} S^{(1)}_{PQ} &= \frac{\sum_s \left\{ \left(\sum_c p_{cs}\right)^2 + \left(\sum_c q_{cs}\right)^2 - \left[\sum_c \left(p_{cs} - q_{cs}\right)\right]^2 \right\}}{\sum_s \left\{ \left(\sum_c p_{cs}\right)^2 + \left(\sum_c q_{cs}\right)^2 + \left[\sum_c \left(p_{cs} - q_{cs}\right)\right]^2 \right\}} \\ &= \frac{\sum_{cs} \left[p^2_{cs} + q^2_{cs} - (p_{cs} - q_{cs})^2\right] + \sum_s T^-_{PQ}(s)}{\sum_{cs} \left[p^2_{cs} + q^2_{cs} + (p_{cs} - q_{cs})^2\right] + \sum_s T^+_{PQ}(s)} \end{aligned} \quad (8)$$

where we have defined $T^{\pm}_{PQ}(s) = \sum_{c \neq c'} \left[p_{cs} p_{c's} + q_{cs} q_{c's} \pm (p_{cs} - q_{cs})(p_{c's} - q_{c's})\right]$. In Fig. 1 (B) it is straightforward to see that $S^{(1)}_{PQ} = 1$, because the inputs on the sectoral level are identical between German construction and Italian construction. We note that Eq. 8 is identical to Eq. 6 in the absence of the terms $T^{\pm}_{PQ}(s)$ (a fact that is the primary reason for our choice in using this measure of similarity).

The difference between perfect national similarity (Eq. 8) and perfect national dissimilarity (Eq. 6) is entirely contained within the sector-dependent terms $T^{\pm}_{PQ}(s)$, and we note that $T^{\pm}_{PQ}(s)$ is a sum over terms involving the direct relationship between P and Q to the countries c and c' in sector s. In the context of a role equivalence calculation, these terms should not all be treated equally: country-sector pairs that are role-equivalent should contribute significantly to the similarity of P and Q, while country-sector pairs that are not role-equivalent should not contribute (diagrammed schematically in Fig. 1 (C)). This can be accomplished by weighting each term in the sum by the similarity between country c and c' in sector s, and we thus write the self-consistent relation

$$S_{PQ} = \frac{\sum_s \sum_{c,c'} \left\{ \left[p_{cs} p_{c's} + q_{cs} q_{c's} - (p_{cs} - q_{cs})(p_{c's} - q_{c's})\right] \times S_{cs,c's} \right\}}{\sum_s \sum_{c,c'} \left\{ \left[p_{cs} p_{c's} + q_{cs} q_{c's} + (p_{cs} - q_{cs})(p_{c's} - q_{c's})\right] \times S_{cs,c's} \right\}} \quad (9)$$

as our final expression for the similarity between two country-sectors P and Q. It is straightforward to verify that the diagonal elements identically satisfy $S_{PP} \equiv 1$ for all country sector pairs P, and that $S^{(0)}_{PQ} \leq S_{PQ} \leq S^{(1)}_{PQ}$ for all P and Q. If all countries are treated as different (with $S_{PQ} = 0$ for $P \neq Q$) Eq. 9 reduces to Eq. 6, whereas Eq. 9 reduces to Eq. 8 if all countries are assumed identical (with $S_{PQ} = 1$ for all countries). In the Appendix, we discuss some additional numerical properties of Eq. 9 and the algorithm we use to determine the numerical values of the similarity. Eq. 9 incorporates a comparison between each of the direct providers of P and Q, but by weighting each term by the similarity implicitly includes a comparison between the *indirect* suppliers of P and Q (those that are providers of the providers). Two different direct providers of P and Q that themselves have similar inputs will have a large contribution to the similarity S_{PQ}, while direct providers who themselves have very different value chains will give a small contribution. This can be clearly seen by computing the similarity in Fig. 1 (C), where we numerically find $S_{PQ} \approx 0.889$ (in comparison to $S_{PQ}^{(0)} \approx 0.444$ and $S_{PQ}^{(1)} = 1$). This shows that Eq. 9 captures our expectation that the similarities in the direct construction inputs due to the shared indirect link (Spanish construction) increases the similarity between German and Italian construction, but the dissimilarities in the direct manufacturing suppliers prevent a perfect role-similarity between them.

The magnitude of S_{PQ} by itself cannot distinguish between similarity due to P and Q sharing *direct* identical providers versus sharing *indirect* role-equivalent providers, therefore we further define the rescaled similarity

$$R_{PQ} = \frac{S_{PQ} - S_{PQ}^{(0)}}{S_{PQ}^{(1)} - S_{PQ}^{(0)}} \tag{10}$$

which indicates how close S_{PQ} is to its upper bound $S_{PQ}^{(1)}$ with respect to its lower bound $S_{PQ}^{(0)}$. Because the upper bound $S_{PQ}^{(1)}$ completely ignores the national difference, if R_{PQ} is very close to 1, it means that there is a significant network effect offsetting the national difference. In other words, if $S_{PQ}^{(0)}$ measures the *direct* similarity between P and Q, the rescaled version R_{PQ} allows us to measure the *indirect* similarity between them.

In this section we have only discussed the similarity based on the upstream GVNs, whose adjacency matrix $\mathbf{U}$ is both asymmetrical (directed) and real-valued between 0 and 1 (weighted) and with non-zero diagonal elements (self-loops). Measuring a downstream similarity using the methods in this section can be equivalently accomplished by applying the same methodologies to the transposed downstream networks (reversing the direction of the links, so that receiver sectors become provider sectors).

3 Results

3.1 General Patterns of Similarity

We compute the pairwise similarity across countries for each sector and each year available in the WIOD. It is worthwhile to examine how strongly correlated our measure of similarity is with other alternative measures. Our measure tends to be highly correlated with other local measures of similarity (all the correlation coefficients are above 0.96 as shown in Fig. 2 for the upstream GVNs). Even though the correlation is high, it must be noticed that, unlike the local measures, ours takes into account both direct and indirect relationships along the value chain. In Fig. 2, we see that when we include indirect value-added providers in the computation of the upstream similarity, country-sector pairs become more similar to one another (more dots above the 45 degree line). In Fig 2, we also include the comparison with directed similarity (DS), which also takes into account the network topology. However, it is less correlated with ours (with the correlation coefficient of 0.61) because it differs in a number of ways from ours as mentioned above. Finally, as a complementary measure capturing the *indirect* effect, the rescaled version of our similarity is much less correlated with other measures of similarity (see Fig. A2 in the Appendix).

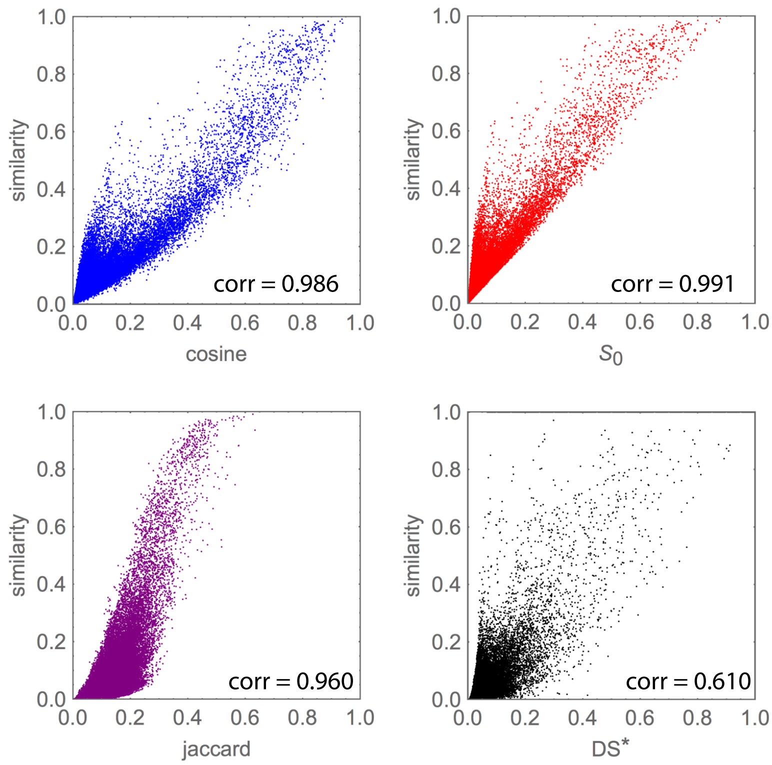

Fig. 2. The scatter plots of our measure of similarity versus other four measures of similarity, Cosine [C_{PQ}], Similarity(0) [$S^{(0)}_{PQ}$], Jaccard [J_{PQ}], and Directed Similarity (DS) [X_{PQ}] (the asterisk indicates that X is normalized by a constant term $\frac{<S>}{<X>}$ where $< \cdot >$ is the mean operator.), for all pairwise comparison of sectors across countries for the upstream GVNs and for all years.

We explore the evolution of the similarity between sectors by computing the mean similarity for all sectors and country pairs, $(\sum_s \sum_{c,c',c\neq c'} S_{cs,c's})/[N_s N_c(N_c-1)]$, with $N_c = 41$ the number of countries and $N_s = 35$ the number of sectors. Fig. 3 reveals that, on average sectors across the globe tend to be more similar over time, a fact that is consistently observed using all measures of similarity, except for directed similarity (DS) which shows fluctuations within a very narrow band (between 0.00057 and 0.0007). All the local measures also show that the upstream similarity is more volatile and less intense than the downstream similarity. However, when all network interdependences are taken into account, our original and rescaled measures of the upstream and downstream similarities tend to more closely follow the same path of growth and both exhibit a temporary reduction in the aftermath of the great recession in 2008 (the latter is also captured by Jaccard). Therefore, by taking into account the network topology, our measures are more sensitive to the 2008 crisis (especially for downstream) and show a clear pattern that the similarity between the GVCs was higher for upstream than for downstream right before the crisis.

The directed algorithm of ref. (Blondel *et al.*, 2004) is the only comparison we make to a more complex measure of similarity between nodes in a network, and it is worth emphasizing the advantages of our algorithm over that approach. The definition of the directed similarity combines upstream and downstream weights, and thus does not permit the evaluation of differing measures of upstream and downstream similarity. The directed algorithm is also computationally very expensive in comparison to our algorithm due to the fact that sector and country characteristics are treated on an equal footing, and there is no way to compute $X_{cs,c's}$ (different countries in the same sector) without computing the similarities between all pairs of countries *and* sectors. Our algorithm requires an evaluation of $(N_c^2 N_s) \times (N_c^2 N_s) \times k$ operations, with k the number of iterations, due to the fact that different sectors are taken to be fundamentally different and we do not need to compute $S_{cs,c's'}$ for $s \neq s'$. Eq. 7 requires two matrix multiplications of $(N_c N_s) \times (N_c N_s)$ matrices, implying $(N_c N_s)^2 \times (N_c N_s)^2 \times k$ computations. The ability to neglect the similarity between different sectors thus provides a speedup by a factor of $N_s^2 = 1225$ in our case. There are thus clear advantages to our approach to measuring similarity in the GVNs and in other networks where nodes can be identified as having different characteristics that are not on an equal footing.

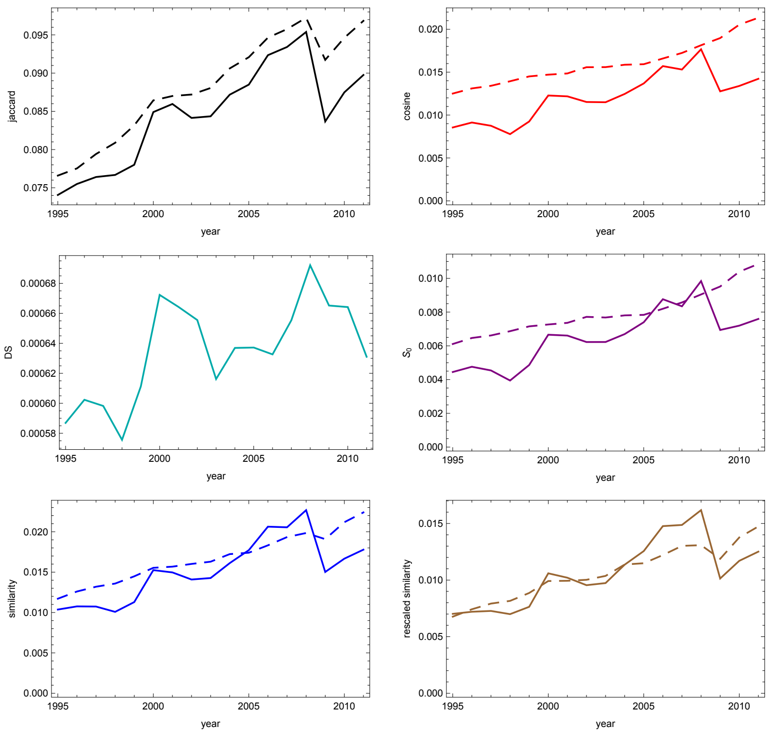

Fig. 3. The evolution of average similarity across countries and sectors over time, 1995-2011. We compare six different measures of similarity: Jaccard [J_{PQ}], Cosine [C_{PQ}], Directed Similarity (DS) [X_{PQ}], Similarity(0) [$S_{PQ}^{(0)}$], our Network Similarity [S_{PQ}], and its rescaled version [R_{PQ}]. Except for DS (which combines upstream and downstream), we report both upstream (solid lines) and downstream (broken lines) similarity.

For each year, we can average across countries to have the average similarity for each sector s, $(\sum_{c,c',c\neq c'} S_{cs,c's})/[N_c(N_c-1)]$. Fig. 4 shows both the average upstream and downstream similarities for all the sectors and for the years 1995 and 2011. It is straightforward to see that most sectors have increased their similarities over time as most 'arrows' are pointing to the northeast direction. Sectors like "Coke, Refined Petroleum and Nuclear Fuel (Cok)" have high average upstream similarity and relatively low average downstream similarity, which means that it is more likely to find country-sector overlap in their upstream value chains. This makes sense for the sector "Cok": energy providers tend to be concentrated in only a few countries. More generally, the manufacturing sectors tend to be more similar across countries than the services sectors as the former is clustered in the top right of Fig. 4 and the latter is clustered in the lower left of Fig. 4. The rescaled similarity has qualitatively similar results (see Fig. A3 in the Appendix).

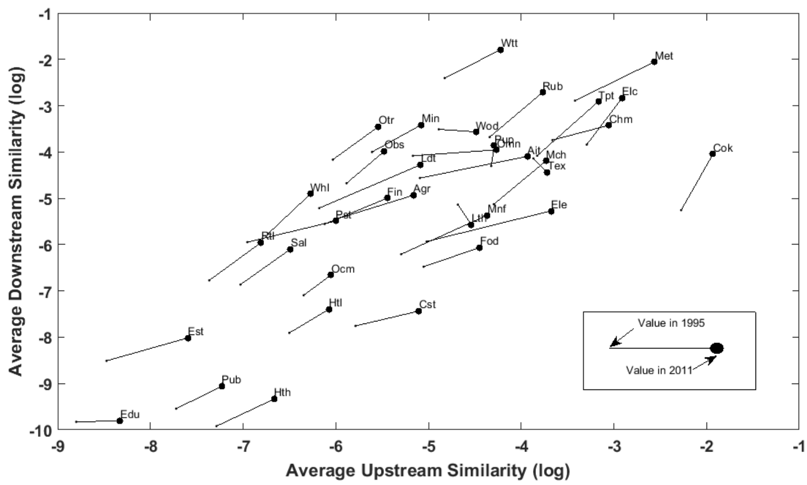

Fig. 4. The average upstream and downstream similarities of sectors for the years 1995 and 2011 using a logarithmic scale.

For each year and each country c, we can also average across sectors and its foreign countries $(\sum_s \sum_{c',c'\neq c} S_{cs,c's})/[N_s(N_c-1)]$ to define an mean similarity for each country. Fig. 5 shows both the average upstream and downstream similarities for all the countries and for the years 1995 and 2011. Again, we observe a general increasing trend of the similarities (see the change of the axis range over time). Over time, countries such as Brazil and India have relatively low average similarity while countries such as Germany have relatively high average similarity. As in the study of Ref. (Dean *et al.*, 2011), we also find that China has been increasingly involved in the vertical specialization and has made a dramatic move over time that it has joined the other "Asian miracle" countries such as South Korea and Taiwan in terms of the similarities. The rescaled similarity has qualitatively similar results (see Fig. A4 in the Appendix).

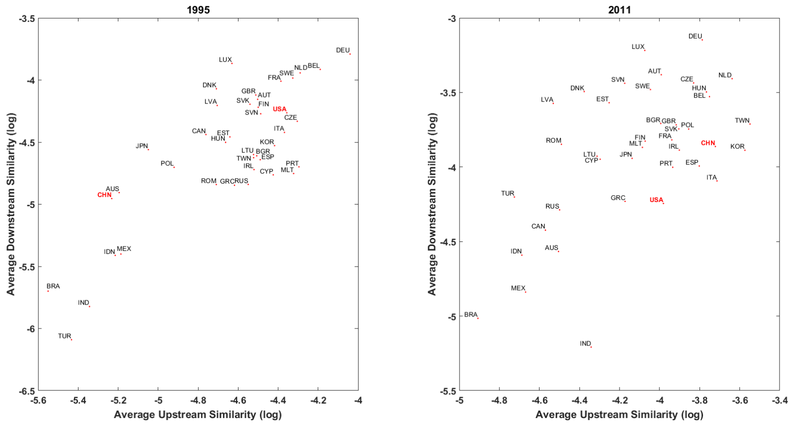

Fig. 5. The average upstream and downstream similarities of countries in years 1995 and 2011 and using logarithmic axes.

3.2 Specific Case Studies

A convenient way to organize our results is to show the country-by-country matrix of pairwise similarities for specific sectors and years. In the specific comparison of the sectors, we are interested in knowing how the production network makes them similar. Therefore, we use the rescaled similarity to capture the network (or indirect) effect. Fig. 6 is an example for the upstream rescaled similarity and the downstream rescaled similarity for the electrical engineering sector, "Elc" (see (De Backer & Miroudot, 2013; Ferrarini, 2011) for a recent analysis of the same sector). Notice that, by our definition of similarity, the matrix is symmetrical and has all 1's in its diagonal. As a result, we only show the lower triangular part of the matrix. As the rescaled similarity goes from 0 to 1, the color changes from red to white (around 0.5) and to blue. There is a visually clear increase in the similarity between most countries in "Elc" between 1995 and 2011, and many countries that were very dissimilar in 1995 became very similar in 2011 (with China being a prominent example). In 1995, China is neither upstream-similar nor downstream-similar to any other countries as its corresponding rows or columns are mostly deep red. In 2011, however, China becomes fairly upstream-similar to Czech Republic, Hungary, Mexico, Slovak, Taiwan, etc, with Czech Republic as its most upstream-similar country. On the other hand, China becomes highly downstream-similar to South Korea and Taiwan, with Taiwan as its most downstream-similar country.

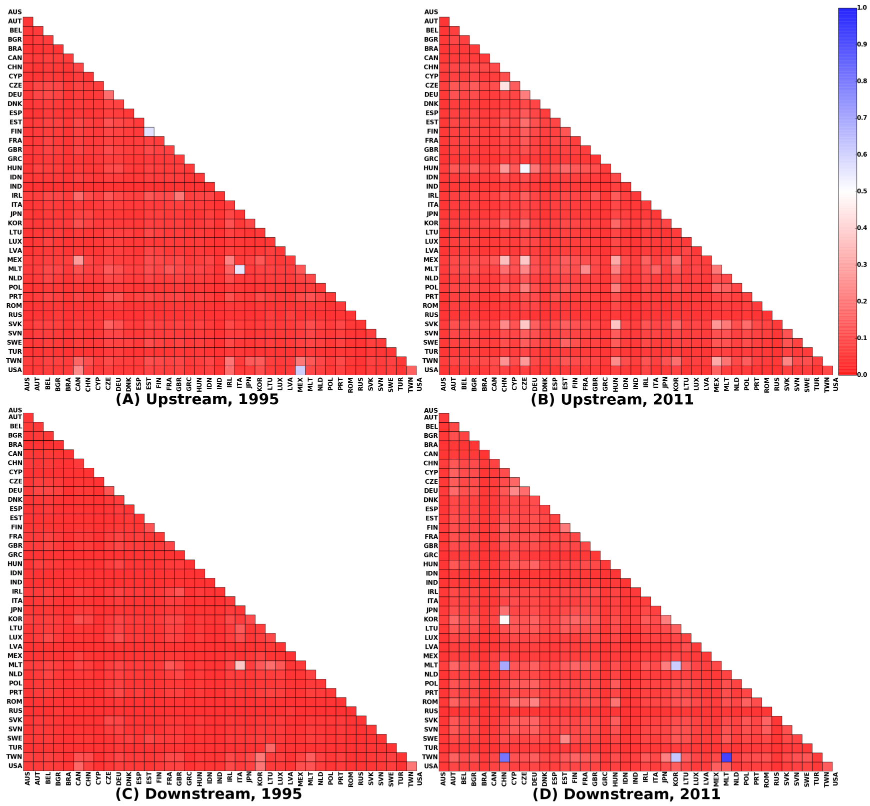

Fig. 6. The pairwise upstream and downstream similarity across countries of the electrical equipment sector in 1995 and 2011. From low to high values, the color changes from red to white (around 0.5) to blue. In 1995, China is not very similar to any other countries. In 2011, the most similar countries to China are Czech Republic (upstream) and Taiwan (downstream).

To see the dynamics at a finer resolution, we show the significant first-degree neighbors (i.e., those with link weight no less than 0.005) of the electrical equipment sector in China and Czech Republic in the upstream GVNs in 1995 and 2011 in Fig. 7, and in Fig. 8 the significant first-degree neighbors of "Elc" in China and Taiwan in the downstream GVNs in 1995 and 2011. Note that while our measure of similarity takes into account all the indirect neighbors, we only show the first-degree neighbors in Figs. 7-8 for better visualization. Over time, the number of shared value-added providers between China and the Czech Republic has increased, and a direct interaction between the two sectors becomes significant as a new link is formed between them. Likewise, the number of shared value-added receivers increases between China and Taiwan over time.

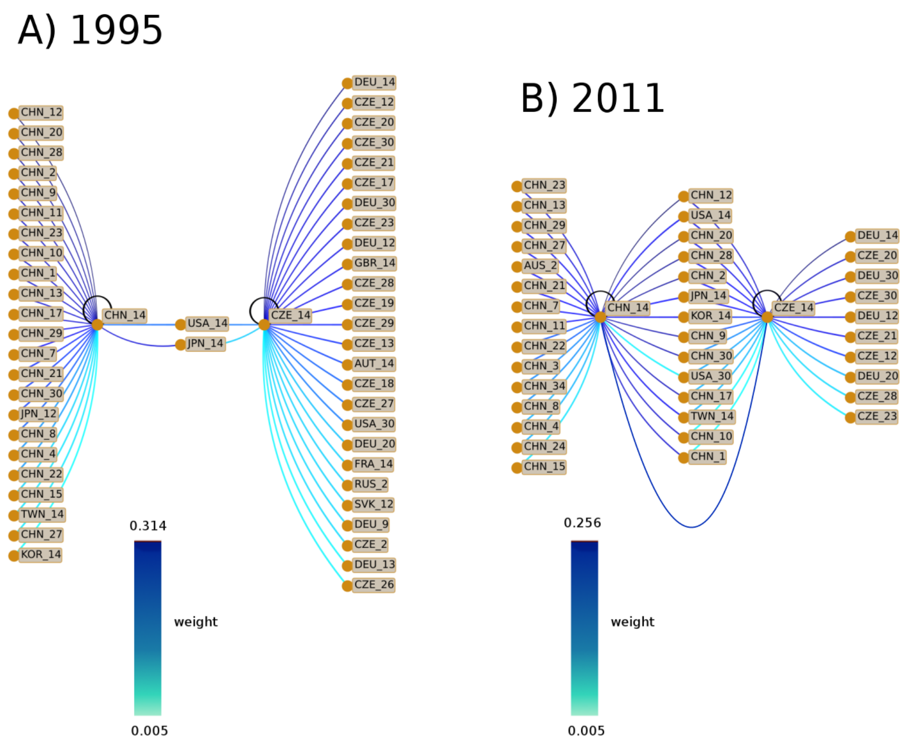

Fig. 7. The first-degree neighbors of the electrical equipment sector in China and Czech Republic in the upstream GVNs in 1995 and 2011. Any incoming links to the two sectors with weight greater than or equal to 0.005 are shown. Over time, the number of shared value-added providers increases for the two sectors, and the direct interaction between the two sectors becomes significant as a new link is formed between them in 2011.

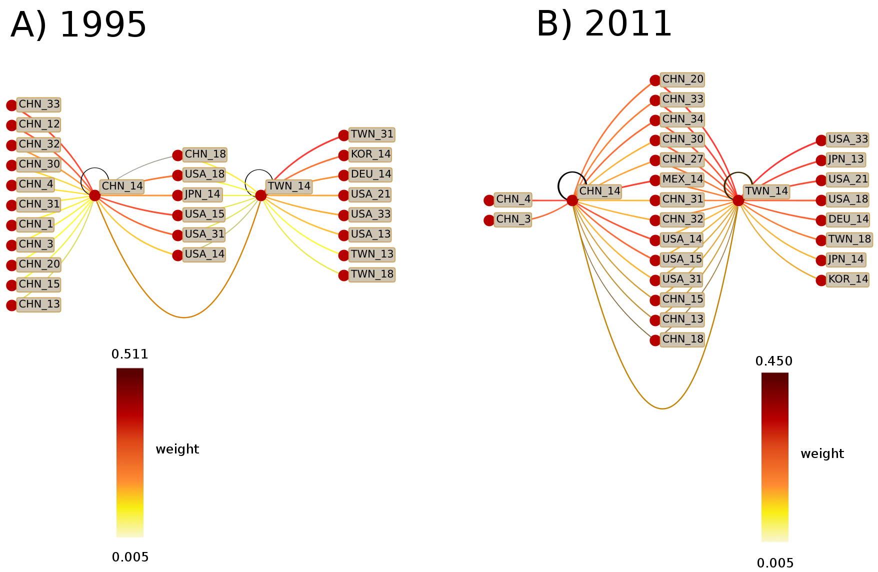

Fig. 8. The first-degree neighbors of the electrical equipment sector in China and Taiwan in the downstream GVNs in 1995 and 2011. Any outgoing links from the two sectors with weight greater than or equal to 0.005 are shown. Over time, the number of shared value-added receivers has increased for the two sectors.

Eq. 9 can give valuable insights into the details behind the increased similarity. We can decompose the numerator of Eq. 9 into individual terms and examine exactly how much each pair of countries contributes to its magnitude. If denoting the country where sector P is located as c_P and the country where sector Q is located as c_Q and $\{c_P, c_Q\} \equiv \Omega_{PQ}$, we can rewrite the numerator of Eq. 9 as

$$\sum_s \left\{ \left[\sum_{\substack{\{c\in\Omega_{PQ},\\ c'\in\Omega_{PQ}\}}} T_{PQ}(s,c,c') \right] + \left[\sum_{\substack{\{c\in\Omega_{PQ},\\ c'\notin\Omega_{PQ}\},or\\ \{c\notin\Omega_{PQ},\\ c'\in\Omega_{PQ}\}}} T_{PQ}(s,c,c') \right] + \left[\sum_{\substack{\{c\notin\Omega_{PQ},\\ c'\notin\Omega_{PQ}\}}} T_{PQ}(s,c,c') \right] \right\} \tag{11}$$

where $T_{PQ}(s,c,c') = \left\{ [p_{cs}p_{c's} + q_{cs}q_{c's} - (p_{cs} - q_{cs})(p_{c's} - q_{c's})] \times S_{cs,c's} \right\}$. According to Eq. 11 we can divide the country pairs into three categories: internal (both countries either China or Czech Republic for the upstream case), external (neither country China nor Czech Republic for the upstream case), and mixed (one either China or Czech Republic and the other a different country for the upstream case). Fig. 9 (A) shows the three components by share of Eq. 11 between the electrical equipment sector in China and the one in Czech Republic over time. On the other hand, Fig. 9 (B) shows the three components by share of Eq. 11 between the electrical equipment sector in China and the one in Taiwan over time.

It is straightforward to see the internal share has been increasing over time between China and Czech Republic (upstream), which implies an increasingly intensive direct interaction between China and the Czech republic. Unlike the upstream case, the internal share has always been high over time between China and Taiwan (downstream), suggesting a persistently strong direct interaction between China and Taiwan.

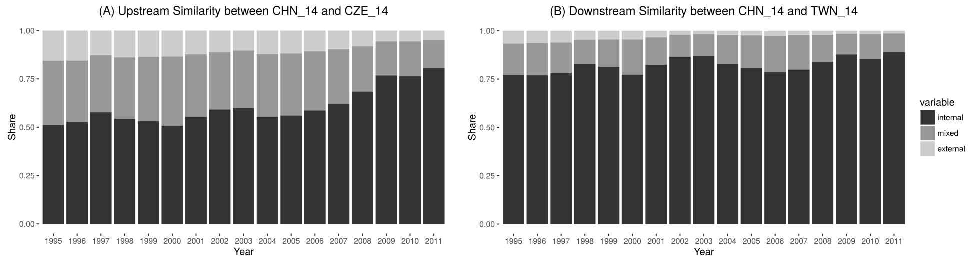

Fig. 9. (A) The three components by share of the numerator of the upstream similarity between the electrical equipment sector ("Elc") in China and the Czech Republic from 1995-2011. (B) The three components by share of the numerator of the downstream similarity between the electrical equipment sector in China and Taiwan from 1995-2011. The three components are internal, mixed, and external (from dark to light).

4 Concluding remarks

In recent decades, international trade has been marked by the spatial fragmentation of production, which is captured by the notion of global value chains (GVCs). A good understanding of the evolution of the GVCs is of vital importance for the macro decision makers to design proper and timely policies and for the micro decision makers to engage in and benefit from the evolution. For example, the detailed knowledge of how much foreign and domestic value-added is contained in gross exports and imports may significantly alter trade policy (e.g., lowering trade barriers such as tariff) and affect local firms' participation in the GVCs (Johnson, 2014b). A method of measuring and comparing the GVCs in a systematic way is necessary for informed decisions on both scales, but about which the existing literature remains silent. This paper has aimed to fill this gap in the literature. First, we use the World Input-Output Database (WIOD) to construct both the upstream and downstream global value networks where the nodes are the individual sectors in different countries and the links are the value-added contribution relationships. Second, to systematically compare the GVCs, we define a network-based measure of role equivalence that takes the differing types of attributes of each node into account. Our measure of similarity assumes that while it is possible to exchange the nationality of a direct provider in a particular sector, the sectors themselves are not interchangeable. Coupling this expectation with naturally-defined lower and upper bounds on similarity permitted the self-consistent definition of similarity.

We have found that on average sectors have had a clear increasing trend of similarity over time and experienced a temporary reduction after the 2008 crisis. However, the reduction was much larger for the upstream GVCs than for the downstream ones. Moreover, manufacturing sectors tend to be more similar across countries than the services sectors while countries like China has increased its average similarity over time. As a case study, we found that the sector of electrical equipment in China has become upstream-similar to the one in Czech Republic and downstream-similar to the one in Taiwan. Our measure of similarity enables us to identify the most intensive interactions among the GVCs across countries and over time. However, the driving forces behind the interactions can be either internal or external, which can be interpreted as value-chain integration or value-chain competition accordingly. Identifying and quantifying these differences will be left for future work.

Regarding the potential uses and policy implications of our measure of GVCs similarity, we expect that the GVC similarity will be a better measure than the traditional export similarity (measured without reference to the topology of the global network). The latter has been largely used in the trade literature as a proxy for competition and trade diversion between countries. However, the gross trade statistics can be seriously flawed (by double counting) as the global production sharing has become a norm. In addition, the trade diversification measured by the export similarity has become a less reliable indicator of a country's competitiveness because similar GVCs are compatible with very dissimilar export outputs (as was the case for China). Our work is therefore in line with the recent efforts to advancing the construction of globalization indices (Martens *et al.*, 2015). In particular, we go beyond traditional measures by considering both direct and indirect links and both domestic and foreign transactions in the global system of trade and production. Furthermore, since the GVCs tend to become more similar over time and countries tend to become more vertically specialized, there are concerns about the systemic risk of the global production system (Kose & Yi, 2001). Integration and diversification are two important features for the stability of input-output systems (Acemoglu *et al.*, 2012; Elliott *et al.*, 2014). Our results suggest that effective diversification is lower than expected due to the increasing overlap of trading partners along value chains, and hence increases the risk of instability. Finally, in the field of complex networks, our measure may also be useful for community detection (Morrison & Mahadevan, 2012; Girvan & Newman, 2002; Zhou, 2003) and as a predictor for future link formation (Lü & Zhou, 2011; Lü *et al.*, 2009; Zhou *et al.*, 2009). For instance, the high similarity between the country-sector pairs identified by our measure may suggest an increasingly intense value-added relationships in the future.

Some possible future extensions to this paper include quantifying the driving forces behind the dynamics of similarity, as mentioned above. Our approach can be generalized to networks with more than two types of node attributes, and so long as it is possible to meaningfully define the lower and upper bounds on the similarity given the constraints of the differing attributes. This approach can also be modified to incorporate other economically relevant information. For example, the great reliance that a sector typically has on itself and the domestic economy at large (in comparison to foreign sectors) may suggest that differentiating between domestic and foreign sectors and treating self-loops differently may be appropriate. In these cases, adapting the upper and lower bounds found in Eq. 8 and 6 to meaningfully capture the differences between foreign and domestic or between self- and non-self-dependence should naturally give rise to an alternative self-consistent measure of similarity.

Acknowledgments

All authors acknowledge support from the PNR project CRISIS Lab. MR and ZZ acknowledge funding from the MIUR (FIRB project RBFR12BA3Y).

5 Appendix

5.1 WIOD Coverage

Table A1. **List of WIOD countries.**

Euro-Zone		Non-Euro EU		NAFTA		East Asia		BRIIAT	
Country	**3L Code**	**Country**	**3L Code**	**Country**	**3L Code**	**Country**	**3L Code**	**Country**	**3L Code**
Austria	AUT	Bulgaria	BGR	Canada	CAN	China	CHN	Australia	AUS
Belgium	BEL	Czech Rep.	CZE	Mexico	MEX	Japan	JPN	Brazil	BRA
Cyprus	CYP	Denmark	DNK	USA	USA	South Korea	KOR	India	IND
Estonia	EST	Hungary	HUN			Taiwan	TWN	Indonesia	IDN
Finland	FIN	Latvia	LVA					Russia	RUS
France	FRA	Lithuania	LTU					Turkey	TUR
Germany	DEU	Poland	POL						
Greece	GRC	Romania	ROM						
Ireland	IRL	Sweden	SWE						
Italy	ITA	UK	GBR						
Luxembourg	LUX								
Malta	MLT								
Netherlands	NLD								
Portugal	PRT								
Slovakia	SVK								
Slovenia	SVN								
Spain	ESP								

Table A2. **List of WIOD sector classifications.**

Full Name	**ISIC Rev. 3 Code**	**WIOD Code**	**3-Letter Code**
Agriculture, Hunting, Forestry and Fishing	AtB	c1	Agr
Mining and Quarrying	C	c2	Min
Food, Beverages and Tobacco	15t16	c3	Fod
Textiles and Textile Products	17t18	c4	Tex
Leather, Leather and Footwear	19	c5	Lth
Wood and Products of Wood and Cork	20	c6	Wod
Pulp, Paper, Paper , Printing and Publishing	21t22	c7	Pup
Coke, Refined Petroleum and Nuclear Fuel	23	c8	Cok
Chemicals and Chemical Products	24	c9	Chm
Rubber and Plastics	25	c10	Rub
Other Non-Metallic Mineral	26	c11	Omn
Basic Metals and Fabricated Metal	27t28	c12	Met
Machinery, Nec	29	c13	Mch
Electrical and Optical Equipment	30t33	c14	Elc
Transport Equipment	34t35	c15	Tpt
Manufacturing, Nec; Recycling	36t37	c16	Mnf
Electricity, Gas and Water Supply	E	c17	Ele
Construction	F	c18	Cst
Sale, Maintenance and Repair of Motor Vehicles and Motorcycles; Retail Sale of Fuel	50	c19	Sal
Wholesale Trade and Commission Trade, Except of Motor Vehicles and Motorcycles	51	c20	Whl
Retail Trade, Except of Motor Vehicles and Motorcycles; Repair of Household Goods	52	c21	Rtl
Hotels and Restaurants	H	c22	Htl
Inland Transport	60	c23	Ldt
Water Transport	61	c24	Wtt
Air Transport	62	c25	Ait
Other Supporting and Auxiliary Transport Activities; Activities of Travel Agencies	63	c26	Otr
Post and Telecommunications	64	c27	Pst
Financial Intermediation	J	c28	Fin
Real Estate Activities	70	c29	Est
Renting of M&Eq and Other Business Activities	71t74	c30	Obs
Public Admin and Defence; Compulsory Social Security	L	c31	Pub
Education	M	c32	Edu
Health and Social Work	N	c33	Hth
Other Community, Social and Personal Services	O	c34	Ocm
Private Households with Employed Persons	P	c35	Pvt

5.2 Relationship with Jaccard and Cosine Similarities

There are many possible ways of measuring the similarity between nodes in a weighted network using information involving only their nearest neighbors, with the Jaccard (Ioffe, 2010) and Cosine (Lü & Zhou, 2011) similarities being often used. We have chosen to use Eq. 6, and in this section we show its relationship to both the Jaccard and Cosine similarities. It is a mathematical identity that

$$J_{PQ} = \frac{\sum_{cs} \min(p_{cs}, q_{cs})}{\sum_{cs} \max(p_{cs}, q_{cs})} = \frac{\sum_{cs} [p_{cs} + q_{cs} - |p_{cs} - q_{cs}|]}{\sum_{cs} [p_{cs} + q_{cs} + |p_{cs} - q_{cs}|]} \tag{12}$$

with the numerator and denominator differing only in a change of sign on the terms involving the absolute value of $p_{cs} - q_{cs}$. J_{PQ} satisfies the useful property that $0 \leq J_{PQ} \leq 1$ with the equalities occurring iff P and Q have either no weight to identical nodes or all identical weights. Many other functional forms satisfy this requirement, though, with a family of examples being

$$J_{PQ}^{(\alpha)} = \frac{\sum_{cs}[p_{cs}^{\alpha} + q_{cs}^{\alpha} - |p_{cs} - q_{cs}|^{\alpha}]}{\sum_{cs}[p_{cs}^{\alpha} + q_{cs}^{\alpha} + |p_{cs} - q_{cs}|^{\alpha}]} \tag{13}$$

for all $\alpha > 0$, with Eq. 6 coinciding with the choice of $\alpha = 2$. Due to the convenient link between Eqs. 6 and 8 that could exist only with the choice of $\alpha = 2$, there is utility in selecting this specific value of α. We further note that for $\alpha = 2$, the numerator of Eq. 13 is $\sum_{cp} p_{cp}^2 + q_{cp}^2 - (p_{cp} - q_{cp})^2 = 2\sum_{cp} p_{cp} q_{cp}$, exactly twice the numerator in the definition of Cosine similarity. While $S_{PQ}^{(0)}$ and C_{PQ} have differing normalizations, we naturally expect that these measures of similarity will be highly correlated. The high degree of similarity between the definitions of $S_{PQ}^{(0)}$, J_{PQ}, and C_{PQ} suggests that the usage of $S_{PQ}^{(0)}$ is reasonable as a measure of similarity.

5.3 Computational Algorithm

The definition of similarity in Eq. 9 is not analytically tractable due to its nonlinearity, and approximate methods for determining the similarity between countries in specific sectors. We use an iterative method to solve for S_{PQ}, by defining the $(k+1)^{th}$ iteration of the similarity as

$$S_{PQ;k+1} = \frac{\sum_s \sum_{c,c'} \left\{ [p_{cs}p_{c's} + q_{cs}q_{c's} - (p_{cs} - q_{cs})(p_{c's} - q_{c's})] S_{cs,c's;k} \right\}}{\sum_s \sum_{c,c'} \left\{ [p_{cs}p_{c's} + q_{cs}q_{c's} + (p_{cs} - q_{cs})(p_{c's} - q_{c's})] S_{cs,c's;k} \right\}}. \tag{14}$$

In the results presented in this paper, we set $S_{PQ;0} = S_{PQ}^{(0)}$ as the initial value of the similarity. This iteration is continued until $\max_{PQ}(|S_{PQ;k+1} - S_{PQ;k}|) \leq 0.001$, at which point the algorithm is assumed to have converged. This relatively high convergence tolerance is due to the computational complexity of the similarity: there are $\sim N_s \times N_c^2$ (each sector and each pairing of countries for each year) similarities that must be computed, and each requires at on the order of $N_s \times N_c^2$ operations (the number of terms in the sums in Eq. 9). This leads to a computational time scaling as $N_s^2 N_c^4 (\approx 3 \times 10^9$ operations for $N_s = 35$ and $N_c = 41)$ to compute one iteration of the of the algorithm. Convergence to the threshold occurred after ~ 30 minutes on a desktop computer (with the algorithm written in C++), and was evaluated on 17 years of data.

The method does converge exponentially fast as a function of the iteration (shown in Fig. A1), and the similarities can be computed after a few hours on a single desktop. We also compared the values of similarity generated using the initial condition $S_{PQ;0} = S_{PQ}^{(0)}$ with that using the initial condition $S_{PQ;0} = S_{PQ}^{(1)}$ (defined in Eq. 8), and found that the largest difference between the two measured similarities was on the order of 0.001, the convergence threshold. This is consistent with the expectation that the algorithm converges to a unique solution.

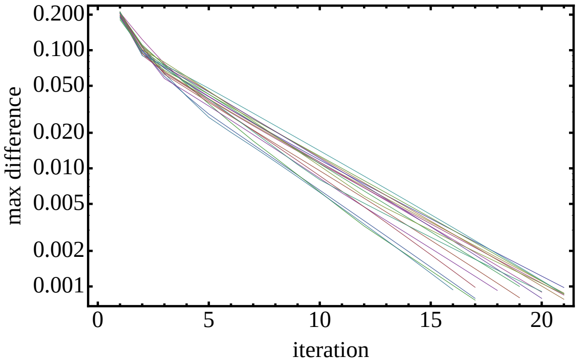

Fig. A1. The convergence of the algorithm as a function of the iteration. Each line denotes the maximum difference $\max_{PQ}(|S_{PQ;k+1} - S_{PQ;k}|)$ as a function for the 17 years (1995-2011) on log-linear axes.

5.4 General Patterns of Rescaled Similarity

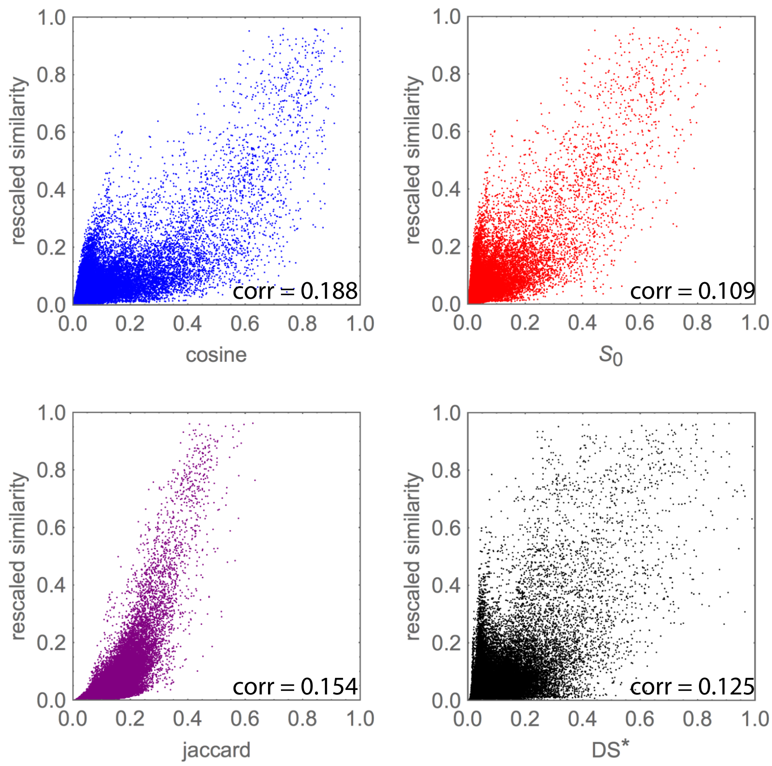

Fig. A2. The scatter plots of the rescaled similarity versus other four measures of similarity, Cosine [C_{PQ}], Similarity(0) [$S^{(0)}_{PQ}$], Jaccard [J_{PQ}], and Directed Similarity (DS) [X_{PQ}] (the asterisk indicates that X is normalized by a constant term $\frac{<S>}{<X>}$ where $< \cdot >$ is the mean operator.), for all pairwise comparison of sectors across countries for the upstream GVNs and for all years.

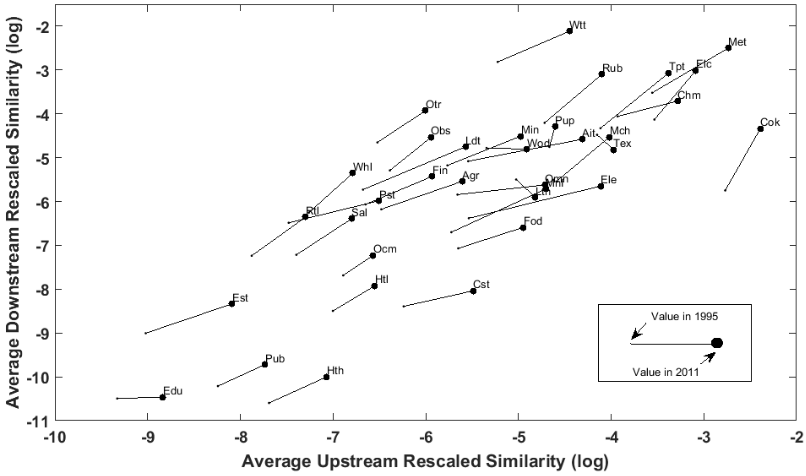

Fig. A3. The average upstream and downstream rescaled similarities of sectors for the years 1995 and 2011 using a logarithmic scale.

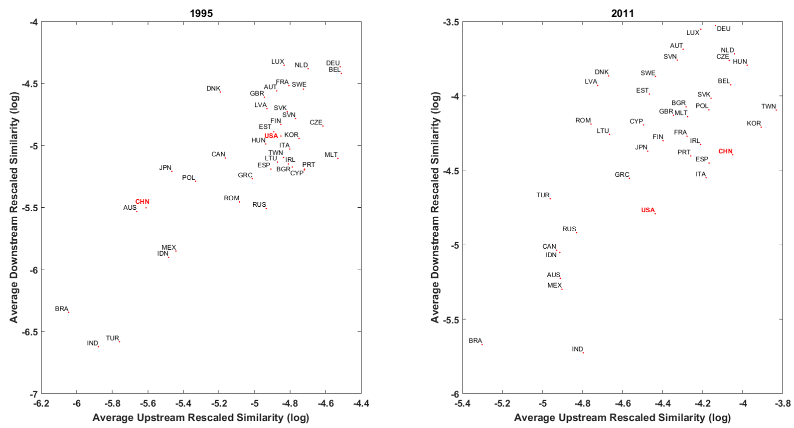

Fig. A4. The average upstream and downstream rescaled similarities of countries in years 1995 and 2011 and using logarithmic axes.

References

Acemoglu, Daron, Carvalho, Vasco M, Ozdaglar, Asuman, & Tahbaz-Salehi, Alireza. (2012). The network origins of aggregate fluctuations. *Econometrica*, **80**(5), 1977–2016.

Albert, R, & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, **74**(1), 47–97.

Amador, João, & Cabral, Sónia. (2014). Global value chains: A survey of drivers and measures. *Journal of economic surveys*.

Baldwin, Richard, & Lopez-Gonzalez, Javier. (2014). Supply-chain trade: A portrait of global patterns and several testable hypotheses. *The world economy*.

Baldwin, Richard, & Venables, Anthony J. (2013). Spiders and snakes: offshoring and agglomeration in the global economy. *Journal of international economics*, **90**(2), 245–254.

Blöchl, Florian, Theis, Fabian J, Vega-Redondo, Fernando, & Fisher, Eric ON. (2011). Vertex centralities in input-output networks reveal the structure of modern economies. *Physical review e*, **83**(4), 046127.

Blondel, Vincent D, Gajardo, Anahí, Heymans, Maureen, Senellart, Pierre, & Van Dooren, Paul. (2004). A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *Siam review*, **46**(4), 647–666.

Caldarelli, Guido, Chessa, Alessandro, Pammolli, Fabio, Gabrielli, Andrea, & Puliga, Michelangelo. (2013). Reconstructing a credit network. *Nature physics*, **9**, 125–126.

Cerina, Federica, Zhu, Zhen, Chessa, Alessandro, & Riccaboni, Massimo. (2015). World input-output network. *Plos one*, **10**(7), e0134025.

Costinot, Arnaud, Vogel, Jonathan, & Wang, Su. (2013). An elementary theory of global supply chains. *The review of economic studies*, **80**(1), 109–144.

De Backer, Koen, & Miroudot, Sébastien. (2013). Mapping global value chains. *Oecd publishing*.

Dean, Judith M, Fung, Kwok-Chiu, & Wang, Zhi. (2011). Measuring vertical specialization: The case of china. *Review of international economics*, **19**(4), 609–625.

Elliott, Matthew, Golub, Benjamin, & Jackson, Matthew O. (2014). Financial networks and contagion. *The american economic review*, **104**(10), 3115–3153.

Everett, MG, & Borgatti, S. (1988). Calculating role similarities: an algorithm that helps determine the orbits of a graph. *Soc. net.*, **10**, 77.

Fagiolo, Giorgio, Reyes, Javier, & Schiavo, Stefano. (2009). World-trade web: Topological properties, dynamics, and evolution. *Physical review e*, **79**(3), 036115.

Feenstra, Robert C, & Hanson, Gordon H. (1999). The impact of outsourcing and high-technology capital on wages: estimates for the united states, 1979-1990. *Quarterly journal of economics*, 907–940.

Ferrantino, Michael J, Taglioni, Daria, *et al.* . (2014). Global value chains in the current trade slowdown. *World bank economic premise*, **137**.

Ferrarini, Benno. (2011). Mapping vertical trade. *Asian development bank economics working paper series*.

Finger, Michael J, & Kreinin, Mordechai E. (1979). A measure of 'export similarity' and its possible uses. *The economic journal*, 905–912.

Girvan, M, & Newman, MEJ. (2002). Community structure in social and biological networks. *Proc. natl. acad. sci.*, **99**, 7821.

Grossman, Gene M, & Rossi-Hansberg, Esteban. (2008). Trading tasks: A simple theory of offshoring. *American economic review*, **98**(5), 1978–1997.

Guimerá, R, & Amaral, LAN. (2005). Cartography of complex networks: modules and universal roles. *J. stat. mech.*, **2005**, P02001.

Hummels, David, Ishii, Jun, & Yi, Kei-Mu. (2001). The nature and growth of vertical specialization in world trade. *Journal of international economics*, **54**(1), 75–96.

Ioffe, Sergey. (2010). Improved consistent sampling, weighted minhash and L1 sketching. *Page 246 of: 2010 ieee 10th int conf on data mining*.

Jeh, G, & Widom, J. (2002). Simrank: a measure of structural-context similarity. *Proc. 8th acm sigkdd knowl. disc. and data min.*, **8**, 538.

Johnson, Robert C. (2014a). Five facts about value-added exports and implications for macroeconomics and trade research. *The journal of economic perspectives*, **28**(2), 119–142.

Johnson, Robert C. (2014b). Five facts about value-added exports and implications for macroeconomics and trade research. *The journal of economic perspectives*, 119–142.

Johnson, Robert C, & Noguera, Guillermo. (2012). Accounting for intermediates: Production sharing and trade in value added. *Journal of international economics*, **86**(2), 224–236.

Kitsak, Maksim, Riccaboni, Massimo, Havlin, Shlomo, Pammolli, Fabio, & Stanley, H Eugene. (2010). Scale-free models for the structure of business firm networks. *Physical review e*, **81**(3), 036117.

Koopman, Robert, Wang, Zhi, & Wei, Shang-Jin. (2014). Tracing value-added and double counting in gross exports. *The american economic review*, **104**(2), 459–494.

Kose, M Ayhan, & Yi, Kei-Mu. (2001). International trade and business cycles: is vertical specialization the missing link? *The american economic review*, **91**(2), 371–375.

Leicht, EA, Holme, P, & Newman, MEJ. (2006). Vertex similarity in networks. *Phys. rev. e*, **73**, 026120.

Leontief, Wassily W. (1936). Quantitative input and output relations in the economic systems of the united states. *The review of economic statistics*, 105–125.

Lerner, J. (2005). Role assignments. *Pages 216–252 of: Network analysis*. Springer.

Lloyd, Peter J. (2004). Measures of similarity and matching in international trade. *Empirical methods in international trade, edward elgar: Cheltenham, uk*, 21–34.

Los, Bart, Timmer, Marcel P, & Vries, Gaaitzen J. (2014). How global are global value chains? a new approach to measure international fragmentation. *Journal of regional science*.

Lü, L, & Zhou, T. (2011). Link prediction in complex networks: A survey. *Physica a*, **390**, 1150.

Lü, L, Jin, CH, & Zhou, T. (2009). Similarity index based on local paths for link prediction of complex networks. *Phys. rev. e*, **80**, 046122.

Martens, Pim, Caselli, Marco, De Lombaerde, Philippe, Figge, Lukas, & Scholte, Jan Aart. (2015). New directions in globalization indices. *Globalizations*, **12**(2), 217–228.

Miller, Ronald E, & Blair, Peter D. (2009). *Input-output analysis: foundations and extensions*. Cambridge University Press.

Morrison, G, & Mahadevan, L. (2012). Discovering communities through friendship. *PLoS one*, **7**, e38704.

Newman, MEJ. (2003). The structure and function of complex networks. *SIAM rev.*, **45**, 167.

Pammolli, Fabio, & Riccaboni, Massimo. (2002). Technological regimes and the growth of networks: An empirical analysis. *Small business economics*, **19**(3), 205–215.

Piccardi, Carlo, Riccaboni, Massimo, Tajoli, Lucia, & Zhu, Zhen. (forthcoming). Random walks on the world inputoutput network. *Journal of complex networks*.

Riccaboni, Massimo, & Schiavo, Stefano. (2010). Structure and growth of weighted networks. *New journal of physics*, **12**(2), 023003.

Riccaboni, Massimo, Rossi, Alessandro, & Schiavo, Stefano. (2013). Global networks of trade and bits. *Journal of economic interaction and coordination*, **8**(1), 33–56.

Richeardt, J, & White, DR. (2007). Role models for complex networks. *Europhys. j. b*, **60**, 217.

Schweitzer, Frank, Fagiolo, Giorgio, Sornette, Didier, Vega-Redondo, Fernando, Vespignani, Alessandro, & White, Douglas R. (2009). Economic networks: The new challenges. *science*, **325**(5939), 422–425.

Timmer, Marcel P, Erumban, Abdul Azeez, Los, Bart, Stehrer, Robert, & de Vries, Gaaitzen J. (2014). Slicing up global value chains. *The journal of economic perspectives*, **28**(2), 99–118.

Timmer, Marcel P., Dietzenbacher, Erik, Los, Bart, Stehrer, Robert, & de Vries, Gaaitzen J. (2015). An illustrated user guide to the world inputoutput database: the case of global automotive production. *Review of international economics*, **23**(3), 575–605.

Tukker, Arnold, & Dietzenbacher, Erik. (2013). Global multiregional input–output frameworks: an introduction and outlook. *Economic systems research*, **25**(1), 1–19.

Xing, Yuqing. (2014). China's high-tech exports: The myth and reality. *Asian economic papers*, **13**(1), 109–123.

Zhou, H. (2003). Distance, dissimilarity index, and network community structure. *Phys. rev. e*, **67**, 061901.

Zhou, T, Lü, L, & Zhang, YC. (2009). Predicting missing links via local information. *Europhys j b*, **71**, 623.

Zhu, Zhen, Cerina, Federica, Chessa, Alessandro, Caldarelli, Guido, & Riccaboni, Massimo. (2014). The rise of china in the international trade network: A community core detection approach. *Plos one*, **9**(8), e105496.

Zhu, Zhen, Puliga, Michelangelo, Cerina, Federica, Chessa, Alessandro, & Riccaboni, Massimo. (2015). Global value trees. *Plos one*, **10**(5), e0126699.