IMT
SCHOOL
FOR ADVANCED
STUDIES
LUCCA

#08
2017

ISSN 2279-6894
IMT LUCCA EIC WORKING PAPER SERIES 08
November 2017

RA **Economics and institutional change**

The Organization of Global Supply Networks

Loredana Fattorini
Armando Rungi
Zhen Zhu

Research Area
Economics and institutional change

The Organization of Global Supply Networks

Loredana Fattorini
IMT School for Advanced Studies Lucca

Armando Rungi
IMT School for Advanced Studies Lucca

Zhen Zhu
IMT School for Advanced Studies Lucca

ISSN 2279-6894
IMT LUCCA EIC WORKING PAPER SERIES #08/2017

Piazza San Ponziano 6, 55100 Lucca

The Organization of Global Supply Networks

Loredana Fattorini* Armando Rungi† Zhen Zhu‡

November 2017

Abstract

In this contribution, we introduce a network approach for the organization of global production across national borders, beyond the sequential industry-level metrics proposed in the previous literature. First, we show and argue that several characteristics of global production processes would be lost in the analysis when assuming that they could be proxied as linear sequences. Hence, we propose an index that assesses the relevance of any input for the target output, including its role as an *input of inputs*. Thereafter, we exploit an own-built firm-level dataset of about 20,489 U.S. parent companies integrating more than 154,000 affiliates worldwide. Results show that the technological relevance of an input in a directed supply network is also a good predictor for: i) the probability that an input industry is actually integrated within a firm boundary; ii) the number of affiliates that are controlled by the parent company and active in that input industry.

Keywords: global value chains, supply networks, vertical integration, upstreamness, firm theory.

JEL Classification Numbers: F23, L23, L22, D57, F14.

*Corresponding author: mail to loredana.fattorini@imtlucca.it. IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy.

†Mail to: armando.rungi@imtlucca.it. IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy.

‡Mail to: zhen.zhu@imtlucca.it. IMT School for Advanced Studies Lucca, Piazza San Francesco 19, 55100 Lucca, Italy.

1 Introduction

Production processes are more than ever spatially fragmented, after a process of unbundling occurred globally, separating the production of intermediates and final goods across national borders (Hummels et al., 2001; Baldwin, 2006; Baldwin and Lopez-Gonzalez, 2015). Eventually, networks of firms can be established, where each stage of production can be vertically integrated or not within firm boundaries. If a production stage is vertically integrated, inputs will be exchanged intra-firm. Otherwise, buyers and suppliers will exchange arm's length after signing supply contracts. Yet, emerging literature on the determinants and consequences of the global supply chain studies the phenomenon of unbundling assuming that an ordered and linear sequence exists, from the conception of the product to its distribution and final use. Albeit an advancement with respect to previous literature, where each stage of production was considered separately, we argue that a linear organization of production is not realistic and a network approach allows catching the complexity of actual sourcing strategies. Production processes are characterized by a multiplicity of linkages and feedback loops more elaborated than simple circuits or linear flows, as pointed out in Hudson (2004).

The same inputs can be used at several stages of production, not in linear progression, before reaching the final demand, and the sourcing of an industry can assume different topological structures. Hence, we measure the distance between each input and the target output in a directed production network, building an industry-pair measure that takes into account the contribution of that input at several stages of manufacturing before reaching the final output. Our *Relevance Index* is based on the methodology of the personalised *PageRank* in Haveliwala (2002), Jeh and Widom (2003), White and Smyth (2003).[1] In fact, the relative input industry position and the industry elasticity of substitution together can determine the choice between *making or buying* a task by a parent company (Antràs and Chor, 2013). Differently from *downstreamness* metrics, as in Antràs and Chor (2013) and Alfaro et al. (2017), our index captures the different importance of inputs occupying a similar position, considering the topology of the technological network, which can be output-specific.

To prove our point, we make use of a firm-level dataset of about 20,489 U.S. parent companies integrating more than 154,000 affiliates worldwide. Results from our firm-level analysis show that the relevance of an input in oriented technological networks is also a good predictor for: i) the probability that an input industry is actually integrated within a firm boundary; ii) the number of affiliates that are controlled by the parent company and active in that input industry.

[1]The *Relevance Index* ranks all the inputs used in the production of a particular output according to their the technological importance as elaborated from input-output tables. All the details about the computation of the index are provided in the Appendix A.1.

This paper is organized as follows. The next section introduces our reference framework. Then, section 3 discusses the value of considering supply networks instead of supply chains, introducing our *Relevance Index*. Section 4 describes the data construction. Section 5 presents the empirical strategy and results. Concluding remarks are offered in Section 6.

2 The Framework

Since the 1980s, many empirical studies have attempted to model the choice of vertical integration that leads to the formation of firm boundaries, based on the degree of contractibility of that input and on the institutional environment of the market where companies operate.

The pioneering work of Coase (1937) helped to build a theory of the firm. In fact, new approaches to understanding what determines a firm's boundaries borrow from the theoretical literature on incomplete contracts (Williamson, 1971, 1975, 1979; Grossman and Hart, 1986) and incorporate these frameworks in general equilibrium models. Acemoglu et al. (2007) for the first time investigate how the degree of contractual incompleteness and the extent of technological complementarities between intermediate inputs affect the choice of technology by headquarters.[2] More recently, Harms et al. (2012) analyse the offshoring decision of firms whose production process is characterized by a particular sequence of steps and a non-monotonic variation of transportation costs. Costinot et al. (2012) build a theory of global supply chains, in which the key feature is that production is sequential and standardized in structure, and offer a first look at how vertical specialization shapes the interdependence of countries. On the country dimension, Antràs and de Gortari (2017) study how trade barriers shape the location of production along global value chains. They show that it is optimal to locate downstream stages of production in relatively central countries.

First of all, we aim at filling the gap in the empirical literature to quantify the position of production processes in technological networks, providing a complete understanding of the complex role that inputs play in the entire production network. In this, we improve on previous measures of positioning along supply chains, as started in contributions by Fally (2012), Antràs et al. (2012), and more recently for bilateral industry-pairs by Alfaro et al. (2017). For our purpose, we build a *Relevance Index* based on the methodology of the personalised *PageRank* in Haveliwala (2002), Jeh and Widom (2003), White and Smyth (2003).

Antràs and Chor (2013) develop a property-rights model of firms' boundaries choice along the value chain, and introduce two measures of industry's production line position, named *DUse_ TUse* and *DownMeasure*. They collapse to the [0,1] range the technological process

[2]For a detailed review of firms' organization strategies and trade, see Antràs and Yeaple (2014).

of production, where 0 indicates the beginning of the production line and 1 the proximity to the final demand, and construct the two novel measures using the U.S. input-output tables. The main prediction that emerges from the model is that the position of an input in the value chain is an important determinant of the ownership structure decisions related to that input. Moreover, such dependence is established by the size of the elasticity of demand faced by the final-good producer relative to the elasticity of inputs across production stages. Production processes are sequential in nature, i.e. downstream stages cannot commence until upstream stages are completed, therefore the organizational decisions along what most assume a production line relate to the contract setting more appropriate to secure each stage input. Usually, firms operate in an environment of incomplete contracts and intermediate producers along with the final producer bargain over the surplus associated with a particular stage. Owning a supplier is a source of power for the firm because it enhances its bargaining power through the residual control rights, but at the same time, it reduces the incentive of suppliers to invest in the relationship. Another important feature of the analysis is that along the value chain there is a spillover effect of relation-specific investments made by upstream suppliers on the incentives to invest of downstream suppliers. As a result, the Antràs and Chor (2013) model suggests that when the average demand elasticity faced by the final-good producer is high (low) relative to the input substitutability, input stages are sequential complements (substitutes), and it is optimal for the firm to vertically integrate relatively downstream (upstream) stages and outsource production stages more upstream (downstream). Exploiting the theoretical framework of Antràs and Chor (2013), Del Prete and Rungi (2017) test at the firm-level the optimal allocation of property rights along the supply chain and explore the correlation between the average downstreamness of the integrated affiliates and the one of the parent company (relative to the final demand), taking into account the output demand elasticity. An important insight in Del Prete and Rungi (2017) is that the decision-making center, i.e. the parent company, can generally be located at any of the production stages, meaning that it could also be far from the final consumer. Under these circumstances, what really matters is the relative position, as opposed to the absolute position, of each parent company with respect to its affiliates. When considering the sequence, they find that the position of the integrated stage of production is correlated with the position of the parent output.

In our contribution, we find evidence for the main prediction of Antràs and Chor (2013), according to which the ownership decisions of a firm depend on the position of an input industry in the entire production process. However, sourcing strategies are more complex than how they are described in the model, in fact, different configurations are possible across industries rather than linear sequences. Consider for example the case of two different industries present in our data: Electronic computer manufacturing (334111) and Automobile manufacturing (336111).[3] According to the *DownMeasure* from Antràs and Chor (2013), their positioning on the supply chain with respect to final use are respectively, .9589 for the Electronic computer manufacturing (334111) and .9997 for the Automobile manufacturing (336111). Although they are both close to the final consumer, we may observe heterogeneous shapes of the technological networks that lead to their final production.[4] Our results show that demand elasticities of final-good producers are not significant determinants of integration choices if we remove the assumption of a unique linear sequence of stages of production where the final-output producer is located at the end of the chain. All inputs are complements in nature, and often they enter the production processes of several outputs at different moments, so the effort choices of stage suppliers might not be driven by the demand elasticity of a single final-output producer. Instead, following a network perspective, we argue that the position of an input in the production network of a target output explains better the firm boundary.

Our paper also contributes to another strand of research set by Carvalho (2014), who describes the intersectoral linkages of industrial networks in U.S., and who shows how aggregate fluctuations can be generated and magnified by individual industry shocks when the latter hit a central industry. Similarly, Acemoglu et al. (2012) show the impact of the economy's architecture on macroeconomic fluctuations. At the firm-level, Atalay et al. (2011) theoretically and empirically characterize the buyer-supplier network overcoming the scale-free framework that does not fit actual firm-level buyer-supplier relationships. Further, Oberfield (2012) proposes a theory of the formation and the evolution of an economy's production network and studies the speed of transmission of productivity shocks in relation to the centrality of that industry in the sourcing strategies of other industries. However, none of the previous works consider firm's boundaries as an alternative mode of organization of supply networks.

3 From supply chains to networks

3.1 Spiders or snakes?

The intuition behind metrics of supply chains (Fally, 2012; Antràs et al., 2012; Alfaro et al., 2017) relies on the linearity of the technological production process, on a sequence, whereas

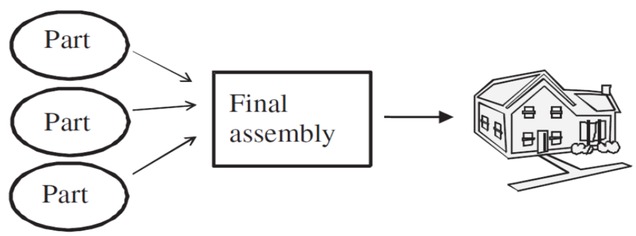
[3]For each sector, we report in parenthesis the IO2002 code from U.S. Bureau of Economic Analysis Input-Output tables.

[4]A portion of the technological networks of the Electronic computer manufacturing (334111) and Automobile manufacturing (336111) is showed in the graphs presented in section 3.3.

the connectivity of modern economies suggests that a network perspective fits better the actual organization of global production processes. We argue that the mutually interactive and recursive nature of the relationships involved in these processes are better understood as sophisticated global supply networks.

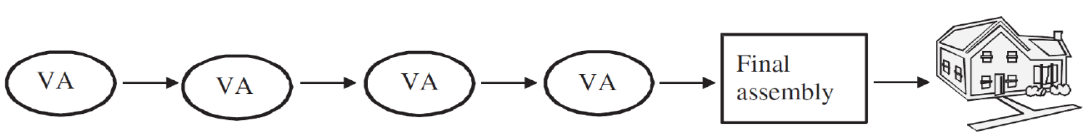
The engineering of the manufacturing process, typical for each industry, dictates the way in which different stages of production links, and it can be unbundled (Baldwin and Venables, 2013). There are two extreme simplified configurations commonly used to describe the supply chain: a spider and a snake. A spider is a production process in which different components come together to be assembled in the final output (Fig. 1). The snake structure requires the process to be a sequence of stages until the final assembly (Fig. 2). The latter relies on the predetermined order in which operations are performed. Measures of industry position along the supply chain proposed in the literature consider only the snake dimension. Instead, all production processes are better thought of as a combination of the snake and spider structure.

Figure 1: A spider global value chain.



Source: Baldwin and Venables (2013). Each cell is a part, component or final product itself. Each arrow is the physical movement of parts, components or the good itself. Movements can be within a plant in a country, or between plants located in different countries.

Figure 2: A snake global value chain.



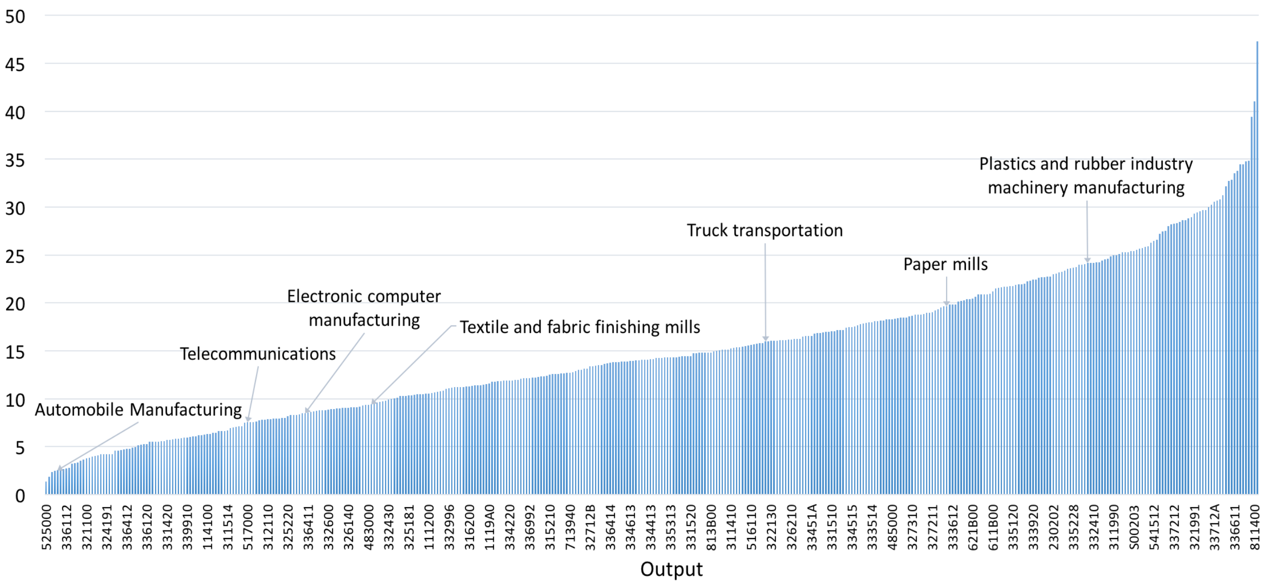
Source: Baldwin and Venables (2013). Each cell is a production stage at which value is added to a product for final consumption. Each arrow is the physical movement of parts, components or the good itself. Movements can be within a plant in a country, or between plants located in different countries.

Before we introduce our metric, we show here how different production networks are from sequential supply chains. We elaborate the following analysis starting from the U.S. Bureau of Economic Analysis (BEA) Input-Output tables.[5] A simple way to account for the spider

[5]In fact, we follow Antràs and Chor (2013) and Carvalho (2014) in the data choice. We use the

dimension, i.e. the width of the production process, is to count the number of effective inputs that are put together for assembling the final output.[6] The higher number of effective inputs used to produce a target output, the more complex the mechanism of production. In Fig. 3, we report the distribution of inputs needed by each industry, where the most complex industry is General state and local government service (S00700) and the simplest is Funds, Trust, and other financial vehicles (525000).

Figure 3: The 'spiderness' of the supply chain.



Source: Own elaboration. Reported number of inputs used by the 426 IO2002 industries from U.S. 2002 I-O tables.

Industries along the supply chain might be very different in their input structures; therefore when measuring their positioning, it should also be taken it into account. In Fig. 4, we measure the similarity between pairs of industries, as indicated by the dominance of the orange colour.[7] The most similar industries, with a score of about 0.876 are Automobile

Commodity-by-Industry Direct Requirements tables after Redefinitions. The input-output account reports the amount of the commodity required to produce one dollar of the industry's output of 426 different industries in U.S.. We use the publicly-available 2002 I-O tables at 6-digit IO2002 code level because it provides a level of industry detail close to NAICS rev. 2012 industry classification codes at 6-digit.

[6]We compute the number of effective inputs borrowing the formula of the Herfindahl index and taking the inverse, as follows:

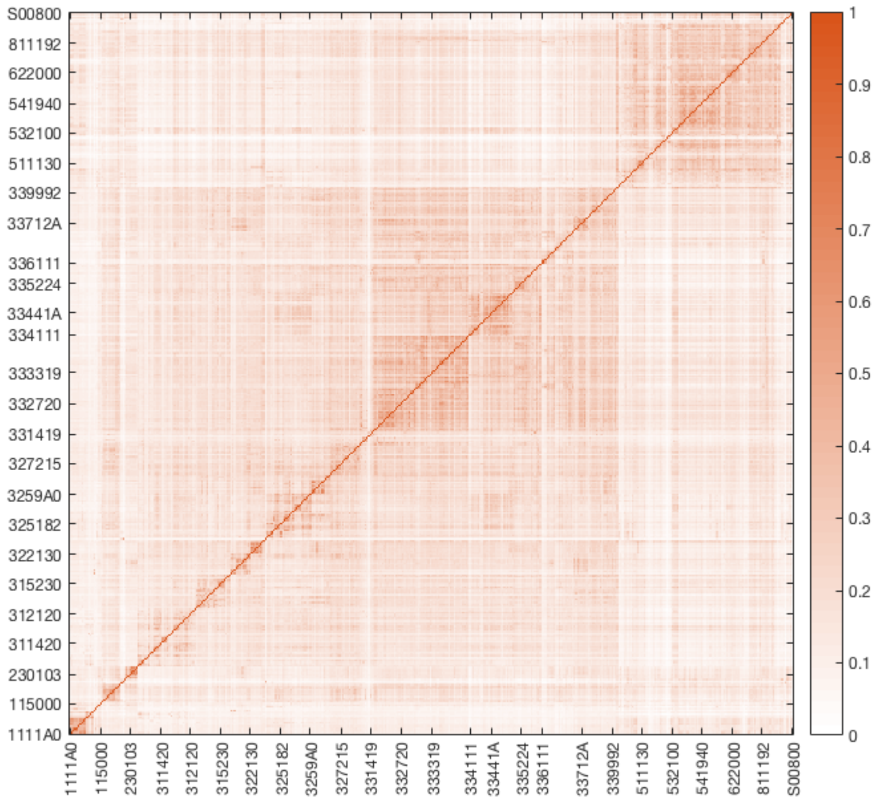
$$C_i = \frac{1}{H_i} \tag{1}$$

with $H_i = \sum_i d_{ij}^2$, and d_{ij} is the direct requirement coefficient from industry providing input i to the industry producing the final output j, in the 2002 input-output tables from the U.S. Bureau of Economic Analysis.

[7]We use the Jaccard similarity index which compares elements for two sets to see which members are shared and which are distinct. In our case, we compare the input's structure of pairwise industries k_i and k_j.

manufacturing (336111) and Light truck and utility vehicle manufacturing (336112), while the least, with a score of 0.006 are Funds, trusts, and other financial vehicles (525000) and Light truck and utility vehicle manufacturing (336112).

Figure 4: Pairwise similarity between industries.



Source: Own elaboration. Reported similarity input structure for 426 pairwise industries at 6-digit IO2002 codes. Labels of only 25 industries. The intense orange colour indicates that two industries are very similar regarding the inputs used.

3.2 Supply Networks

In a broader perspective, the economy can be seen as a thick network where heterogeneous nodes are industries, and heterogeneous and directed edges are economic transactions between these industries.

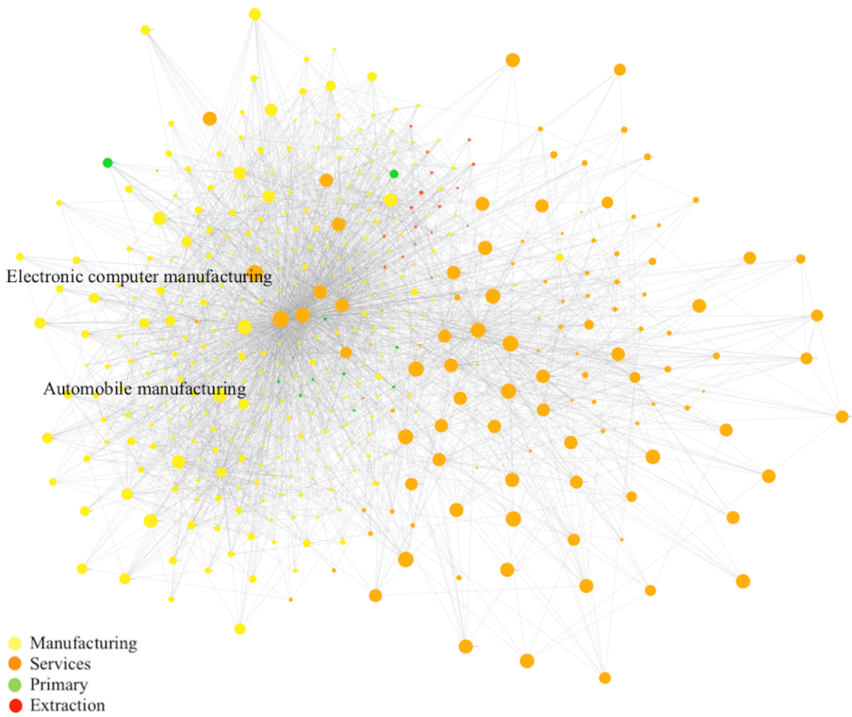
The network visualization in Fig. 5 helps us catching the actual complexity and all the dimensions of the sourcing strategies, at the industry-level. We assume that input-output

It ranges from 0 to 1. The higher the value, the more similar the two sets of inputs P_{s_i} and P_{s_j}, respectively. The formula is:

$$J_{k_i,k_j} = \frac{\sum_s min(P_{s_i}, P_{s_j})}{\sum_s max(P_{s_i}, P_{s_j})} \tag{2}$$

where s are the 426 IO2002 industries in the U.S. Bureau of Economic Analysis classification.

Figure 5: Input-Output Network.



Source: Own elaboration on the U.S. 2002 I-O tables, BEA. Reported input transactions above 1%. Visualization: Gephi software, Force Atlas layout (Jacomy et al., 2011).

linkages represent the backbones of standard technologies, such that correlation in input usage exists across countries, especially when countries share the same technology frontier.[8] The entire production network is reported, where industries are nodes and each non-zero entry for input requirements implies a directed edge, i.e. a flow of input is needed to produce an output. We end up with 51,768 edges. For the empirical analysis that follows, in particular for the construction of our metric, we retain both the information of inputs required to produce an output and the weights associated with each link. We consider all links, including those representing very small transactions to keep the analysis more accurate.[9] The whole set of industries can be sorted into four macro categories: manufacturing (with a density of 65.49%), services (27.46%), primary (4.46%), extraction (2.58%). Interestingly, the yellow

[8]Further assumptions when using I-O tables are the existence of a homogeneous production function for all firms in an industry, and the possibility to neglect economies of scale. Moreover, the extension of U.S. I-O tables to other countries different from U.S. requires assuming that cross country factor price equalization holds. See also Acemoglu et al. (2009), and Fan and Lang (2000) for previous works extending U.S. tables. For example, automobile makers will always require tires, glass, plastic and steel wherever, but size of firms and product quality may change also within the same industry.

[9]We point out that discarding some inputs with little real value prevents the analysis to keep track of those inputs that could have a relevant role in the supply network. For example, inputs appearing several times in the production process, or essential and inelastic one which still may have a high economic value.

nodes, which have the most industries are in the manufacturing sector, but the larger orange nodes are in the service sector.

To characterize this network, capturing graph-theoretic properties, we provide a quick overview of some basic network statistics. The first issue we address concerns the connectivity. More specifically, we explore the extent to which industries are more or less connected, regarding the number of links and interaction intensity. We have that the fraction of the actual links over the potential links, thus the density is 0.286, indicating a rather high degree of connectivity if we consider the overall set of links, meaning in our perspective that the supply network is rather tangled. The increasing fragmentation of production, as well as the reduction of transportation costs, make it easy to provide even small components from spatially farther suppliers. Furthermore, this is confirmed by the diameter of 4, which suggests that the length of the longest geodesic path between any pair of nodes is very short, and the average path length is even shorter, about 1.716, pointing out to a small-world nature of the supply networks. Indeed, the input-output network exhibits clustering patterns, i.e. there is a high probability that if industry k_i and k_j interact and, the latter interacts with industry k_z, then in turn industry k_i and k_z interact. The average clustering coefficient is 0.576. We now turn the attention to node-specific indicators, in particular, we consider the number of each industry's links. Node size in Fig. 5 gives an idea of the distribution across macro categories. The in-degree of an industry k_i, defined as the number of distinct input-demand transactions, ranges from 45 to 296. The highest numbers of incoming links supply a lot of services' industries, among others, Retail trade (4A0000), Scientific research and development services (541700), General state and local government services (S00700), Wholesale trade (420000). On the other hand, the out-degree, thus the number of distinct output-supply transactions, ranges from 1 to 425. As expected, Wholesale trade (420000) provides inputs to many industries, as well as all transportation's industries, Telecommunications (517000), Electric power transmission and generation (221100), among others. Besides, information on node strength is given by the weighted degree, which is based on the number of edges and their weights. In the input-output network, the average weighted degree equals 0.571. The higher is the node strength, the higher is the intensity of interactions mediated by that node. It is interesting to note that two nodes with the same degree can be associated with two different strength. For example, this is the case of Computer storage device manufacturing (334112) and Surgical and medical instrument manufacturing (339112), which both have a degree of 156, but the former has a weighted degree of 1.087, while the latter of 0.609. In other words, it means that the flow of inputs directed to the Computer storage device manufacturing (334112) is greater than the one received by Surgical and medical instrument manufacturing (339112). In our data, the industry Management of companies and enterprises (541610) is the one with the highest weighted degree, followed by Wholesale trade (420000).

Another important feature is the identification of the main industries in the global production network, through node centrality measures. We refer to the global notion of centrality, i.e. a node is central if it has a strategic position in the overall network structure. The *betweenness centrality*, defined as the proportion of all the shortest paths between any two nodes passing through a given node, can be regarded as a measure of a node's control over the flow between others. In sociological terms, it measures the extent to which a node plays the role of a 'broker' in the network. For instance, in the input-output network under examination, critical industries according to this measure are Scientific research and development services (541700), Retail trade (4A0000), Nonresidential maintenance and repair (S00203), Management of companies and enterprises (541610), Food services and drinking places (722000), Wholesale trade (420000), along with others.

The focal point of our analysis relies on the stage structure of each industry in directed technological networks. We argue that each node has a complex structure upstream. For instance, consider Electronic computer manufacturing (334111) and Automobile manufacturing (336111), we can pull out those nodes from the overall network and represent their particular input-output structure separately. In the next section, we clarify the potential of this more advanced analysis, and we provide a visualization of these two networks. Broadly speaking, we believe that the more realistic perspective of complex production networks, rather than simple sequences, calls for analytical tools that can characterize their properties, therefore pointing out to a richer characterization of evolving production structures.

3.3 A Relevance Index for the sourcing of inputs

We propose a new industry-pair measure, which reflects the position in the supply network of direct and indirect inputs for the production of an output, in a network perspective, based on input-output tables. Our aim is to catch the production staging position of each input, moreover, its relevance for the orientation of the technological process, when vertical structures of production may require its usage recursively before completing the process.

The recent developments in network science provide a rich toolbox for analyzing the relationship between a pair of nodes in a network.[10] The easiest way is to check if there exists a link between two nodes or evaluate the link weight between them if the network is weighted. To have a more sophisticated idea of how closely the two nodes are related by taking into account both direct and indirect connections, one may consider measuring the shortest path

[10] Several measures of centrality, catching the absolute position of a node in the network, are well-defined and implemented. See Freeman et al. (1979), Katz (1953), Brin and Page (2012) and Kleinberg (1999) for the most used.

length (Newman, 2010) or the communicability (Estrada and Hatano, 2008) between them. More explicitly, White and Smyth (2003) propose some algorithms to quantify the relative importance of one node with respect to another based on weighted paths and Markov chains. In this context, we build on the methodology of the personalized *PageRank*[11] in Haveliwala (2002), Jeh and Widom (2003), White and Smyth (2003), to estimate the relative importance of industries. Unlike the regular *PageRank*, which evaluates each node's global importance, the personalized *PageRank* enables us to evaluate and rank each node *with respect* to a root node.

The proposed industry-pair measure, we called *Relevance Index*, is constructed using the U.S. I-O tables issued by BEA in 2002. For each pair of IO2002 codes, we identify the direct requirement coefficient d_{ij}, i.e. the input from industry k_i used in the production of industry k_j's output.[12] The *Relevance Index*, $RI(i|r) \equiv \boldsymbol{\pi}_r$, is computed using the personalized *PageRank* algorithm with respect to the root node r, as follows:

$$\boldsymbol{\pi}_t = (1-\alpha)\mathbf{P}\boldsymbol{\pi}_{t-1} + \alpha\mathbf{e}_r \tag{3}$$

$\mathbf{P}$ is the normalized direct requirement matrix[13], $\mathbf{e}_r$ is a column vector with all its elements as 0s with the exception of the rth element which equals 1 and represents the selected root node. The parameter α which lies in [0,1] is the tuning parameter.[14] In our framework, α can be interpreted as a distance parameter and by choosing a rather high value, inputs closely (directly or indirectly) related to the target output will be more important in the

[11]The approach undelying the personalized *PageRank* is that of Markov Chain. A graph is viewed as a stochastic process, specifically a first-order Markov chain, where the edges are labeled by the probabilities of transitiong from one node at time t to another node at time $t+1$. For a detailed description see the Appendix A.1.

[12]By column-normalizing the direct requirement coefficients of the input-output table, after removing the final demand and the external value-added, the analogy with the transition probabilities in the Markov chains is easily employed. While excluding the value-added contribution of the final output, the intrinsic (second round) value-added in its intermediate inputs does not affect the analysis since we are interested in the relative importance of the direct and indirect inputs.

[13]Elements of matrix $\mathbf{P}$ are the transition probabilities of passing from a node i to a node j.

[14]The parameter α, defined as the 'back' probability in White and Smyth (2003), represents the probability to go back to the root node at each time-step. Higher values of α mean that the 'random surfer', which can be interpreted as a 'purchasing agent' in our context, can travel short distances, implying that further nodes are seldom visited, so at the end, the average time spent on those nodes, which proxies their relative importance, should be little. The choice of this parameter is empirical, and in this study we show that we can choose an α_i peculiar to each input, according to one of its characteristic, such as the contractibility. For instance, a more contractible input (higher α_i), i.e. not sold on an exchange nor reference priced, is more likely to be considered by the 'purchasing agent' for possible integration within firm's boundaries.

final ranking.[15] Imposing the steady-state condition $\boldsymbol{\pi}_t = \boldsymbol{\pi}_{t-1} \triangleq \boldsymbol{\pi}_r$, the solution of (3) is:

$$RI(i|r) \equiv \boldsymbol{\pi}_r = \alpha[\mathbf{I} - (1-\alpha)\mathbf{P}]^{-1}\mathbf{e}_r \tag{4}$$

where $\boldsymbol{\pi}_r$ is the vector of probabilities, ranking all the inputs used in the production of a target output. The *Relevance Index* lies in the [0,1] range by construction, where values are assigned to each node proportionally to its economic distance from the root node.[16]

The new metric, which proxies the technological importance of inputs against the producer of the final output, has several desirable properties that guarantee its use as a measure of production position in global supply networks. Firstly, it allows capturing the different strength of inputs at the same distance from the target output, considering their upstream structure. Each input can be seen as the output of its production process embedded in a network, therefore inputs with the same distance to the final output have a different history behind, which shapes the future relationship with the final output. Secondly, inputs enter the production process directly and indirectly, and possibly more than once. Therefore the positioning of an input also rests on recursive power dynamics. We argue that our measure is able to capture an important feature of the complex structure of production processes. The index varies for each input-output pair and can be directly mapped to parent-affiliate pairs.

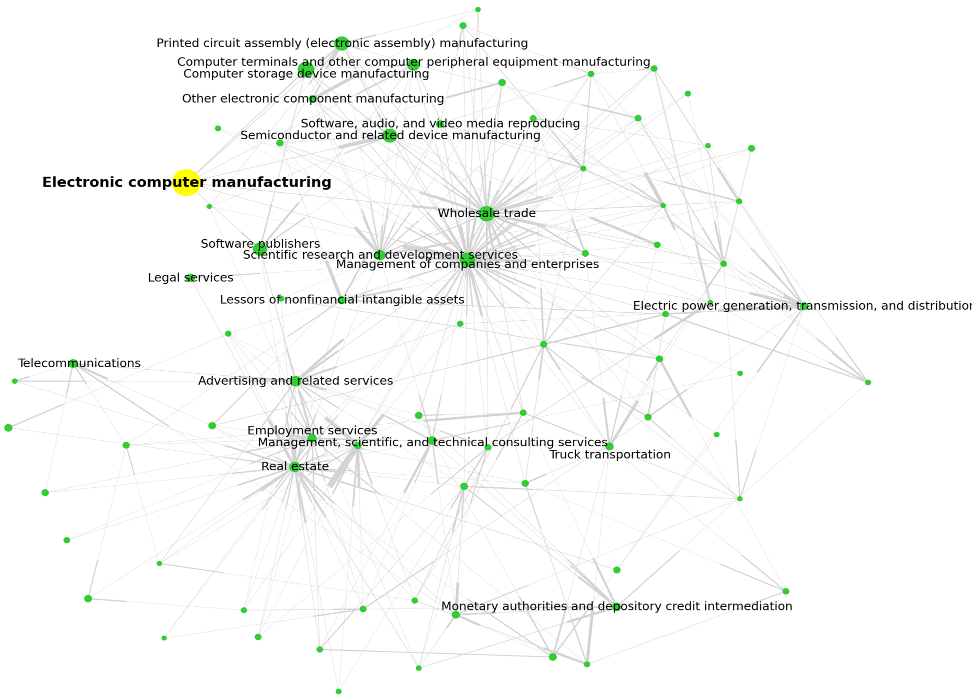
In Fig. 6, we report a portion of two extracted supply networks, namely the Electronic computer manufacturing (334111) and the Automobile manufacturing (336111) networks. According to the downstreamness measure (Antràs and Chor, 2013), their positioning on the supply chain is .9589 for the Electronic computer manufacturing (334111) and .9997 for the Automobile manufacturing (336111). Although they are both close to final demand, a network perspective captures the heterogeneous shape of the directed technological networks. In Fig. 6a, the yellow node represents the computer sector, while green nodes are some of the industries producing its inputs. The dimension of each node is given by the *Relevance Index*, which measures the distance between each input and the target output, taking into account the contribution of that input at several stages before reaching the final output, and its role as an input of inputs. The industries featuring the highest values of the metric for the Electronic computer manufacturing (334111) are Computer storage device manufacturing (334112), Printed circuit assembly manufacturing (334418), Semiconductor and related device manufacturing (334413), Software publishers (511200), Computer

[15] The distance factor is important to take into account the 'dispersion' in usage of the metric, although flexibility is provided to employ several assumptions on the magnitude of the dispersion process (e.g., 'knowledge dispersion' as in the case of innovation diffusion and technological spillover).

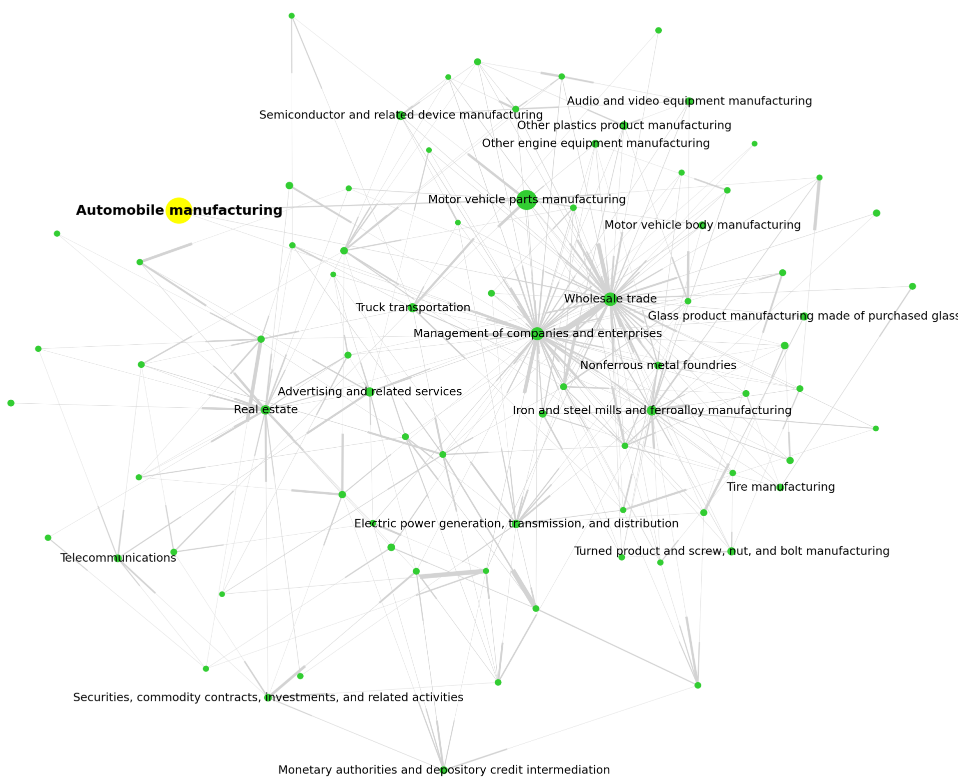
[16] Details about the construction and the computation of the *Relevance Index* are provided in the Appendix A.1.

terminals and other computer peripheral equipment manufacturing (33411A), Telecommunications (517000), Management, scientific, and technical consulting services (541610), Other electronic component manufacturing (334419), Software, audio, and video media reproducing (33461A), and Scientific research and development services (541700) among others. On the other hand, some industries which result more important for the Automobile manufacturing (336111) sector are Motor vehicle parts manufacturing (336300), Iron and steel mills and ferroalloy manufacturing (331110), Semiconductor and related device manufacturing (334413), Truck transportation (484000), Other plastics product manufacturing (32619A), Turned product and screw, nut, and bolt manufacturing (332720), Glass product manufacturing (327215), Other engine equipment manufacturing (333618), Motor vehicle body manufacturing (336211), along with others. In both cases, the 'nearest' inputs with possibly a high technological relevance are amongst the basic components to final outputs. In both extracted networks, we observe the presence of Wholesale trade (420000), which highlights the fact that these industries resort to an intermediate process in the distribution of merchandise. Additionally, they significantly interact with Real estate (531000), Management of companies and enterprises (550000), and Monetary authorities and depository credit intermediation (52A000) industries.

Figure 6: A visual comparison of two extracted networks.



(a) Electronic computer manufacturing (334111).



(b) Automobile manufacturing (336111).

Source: Own elaboration. Reported the first 80 nodes, sized according to the *Relevance Index* value ($\alpha = 0.5$). Edge threshold of 0.02. Labels of the 20 highest values nodes. Visualization: Phyton NetworkX, Graphviz interface, Neato layout (North, 2004).

4 Data and Sample Construction

Our firm-level data are sourced from the Orbis database, by Bureau Van Dijk (BvDEP), gathering financial and ownership information on millions of firms located worldwide. In particular, we collect information on U.S. parent companies and their affiliates operating in 210 countries. We end up with a sample of 20,489 U.S. parents controlling at least 154,836 firms active in a variety of industries at the end of the year 2015.[17] The selection of U.S. parents is made coherently with the choice of using the U.S. Input-Output tables. Among them, 11.4% of parents lead a multinational group with at least one affiliate located abroad. On average, a representative U.S. parent controls 9.57 affiliates. In Table 1, we provide descriptives of the size distribution of U.S. parent companies included in the sample.[18] Small and medium-sized parents account for 83.73% of the dataset.

Table 1: Size distribution of U.S. parent companies.

Size category	Number	%
Small	9,927	48.45
Medium	7,229	35.28
Medium-large	2,001	9.77
Large	939	4.58
Very large	393	1.92
Total	20,489	100

Affiliates from our sample can be active in any industry: manufacturing (28.86%), services (69%), primary (0.29%), extraction (1.85%). Approximately 19% of companies integrated by U.S. parents are located outside the United States. In Table 2, we provide the geographic coverage of our sample by main hosting countries/areas. Not surprisingly, the global supply networks originate mainly in OECD economies, where the 96% of affiliates are based. European Union economy hosts the largest number of firms after the United States, where U.S. parent companies still invest the most, possibly from one state to another. Within Europe,

[17]The Orbis database enables the identification of an *ultimate owner* and its linkages with affiliates worldwide. An *ultimate owner* is broadly defined as a shareholder (individual, family, public authority) that is on top of any ownership path, i.e. it can not be owned by any other shareholder in nature or by law (Rungi et al., 2017). To build our sample identifying firm boundaries, we follow international standards (Altomonte and Rungi, 2013; UNCTAD, 2016; Rungi et al., 2017), in which the observation unit is the link between an affiliate company and a parent, coming out from the direct or indirect equity participation of the latter when the absolute majority of votes ($\geq$ 50.01%) is reached. We adopt a network approach which allows taking into account both the direct control at the parent-level and the indirect control through affiliates controlling sub-affiliates at different hierarchical levels. Cross-participations are neglected.

[18]We classify firms by size based on a combination of criteria: revenues, or total assets, or number of employees, or capitalization, or listed on the stock exchange.

Germany, United Kingdom, and Netherlands attract a significant share of foreign affiliates providing especially service activities. Whereas, several Canadian companies supply U.S. multinationals with final and intermediate goods. The less active geographic areas of the supply networks in which U.S. parents are embedded are Russia, Middle East, and Africa.

Table 2: Sample geographic coverage of affiliates.

Country	Final goods		Intermediate goods		Services		Total	
	Affiliates	%	Affiliates	%	Affiliates	%	Affiliates	%
United States	20,571	16.34	24,590	19.53	80,729	64.13	125,890	100
European Union	1,934	11.45	2,084	12.34	12,872	76.21	16,890	100
of which:								
Germany	273	13.17	306	14.76	1,494	72.07	2,073	100
France	171	11.03	213	13.73	1,167	75.24	1,551	100
United Kingdom	563	11.44	624	12.68	3,734	75.88	4,921	100
Italy	136	19.37	139	19.80	427	60.83	702	100
Netherlands	158	6.77	171	7.33	2,005	85.90	2,334	100
Canada	980	30.36	923	28.59	1,325	41.05	3,228	100
Russia	18	11.69	30	19.48	106	68.83	154	100
Asia	251	15.02	312	18.66	1,109	66.32	1,672	100
of which:								
Japan	87	11.52	76	10.07	592	78.41	755	100
China	92	12.06	66	8.65	605	79.29	763	100
India	122	15.66	149	19.13	508	65.21	779	100
Africa	67	14.17	93	19.66	313	66.17	473	100
Middle East	82	18.22	80	17.78	288	64	450	100
South America	221	12.10	395	21.63	1,210	66.27	1,826	100
of which:								
Argentina	24	8.08	70	23.57	203	68.35	297	100
Brazil	137	14.59	219	23.32	583	62.09	939	100
Mexico	98	23.28	154	36.58	169	40.14	421	100
Australia	123	14.20	157	18.13	586	67.67	866	100
Rest of the world	489	16.49	585	19.72	1,892	63.79	2,966	100
Total	24,834	16.04	29,403	18.99	100,599	64.97	154,836	100

For the purpose of our analysis, we map parents' and affiliates' industry affiliation at 4-digit NAICS rev.2012 into industries from 2002 U.S. I-O tables produced by BEA.[19] After

[19]We use and re-elaborate U.S. BEA correspondence tables from NAICS to 2002 I-O codes at 6-digit

matching the primary activities, we combine firm-level data with industrial metrics well-grounded in the literature of firm boundaries. Demand elasticity for the industry that sells or buys an input is sourced from Broda and Weinstein (2006); metrics of sector position on supply chains based on U.S. I-O tables (*DUse_ TUse* and *DownMeasure*) are sourced from Antràs and Chor (2013); a measure of contractibility, ranking industries by the easiness to contract with a supplier is retrieved from Antràs and Chor (2013), based on the methodology of Nunn (2007). Finally, we complement the dataset with our industry-pair metric, namely the *Relevance Index*, by matching each affiliate-parent observation corresponding activities.

5 Empirical strategy and results

Our aim is twofold. First, we test whether the relevance of an input in supply network can explain the 'make or buy' choice by a parent company. Second, we investigate the economic weight of each sector proxied by the number of integrated affiliates for a relevant input industry.

5.1 The choice of vertical integration

To test vertical integration, we adopt a parent-level fixed effects conditional logit model, as in Del Prete and Rungi (2017), as follows [20]:

$$\begin{aligned} d_{i(j)k_i} &= \beta_0 + \beta_1 RI_{k_i k_j} + \beta_2 D_{k_i} + \beta_3 \mathbf{1}(\rho_{k_j} > \rho_{med}) \cdot D_{k_i} + \beta_4 \mathbf{1}(\rho_{k_j} > \rho_{med}) \cdot RI_{k_i k_j} \\ &\quad + \beta_5 C_{k_i} + \gamma_j + \varepsilon_{i(j)k_i k_j} \end{aligned} \tag{5}$$

where $d_{i(j)}$ is a binary variable equal to 1 if the ith affiliate active in input industry k_i is integrated by the jth parent company mainly operating in output industry k_j and 0 otherwise. We also consider input industries that could have been integrated, but which eventually were not, relying on the methodology of Fan and Lang (2000).[21] $RI_{k_i k_j}$ is our industry-pair *Relevance Index*, described in section 3.3, reflecting the technological position of the input i with respect to the target output j, i.e. its technological relevance. D_{k_i} is the absolute downstreamness measure from Antràs and Chor (2013) indexing the position of an input industry k_i in the value chain, with larger values corresponding to stages further downstream (closer

(IO2002) to merge firm and industry-level information at different levels of disaggregation.

[20] See McFadden (1974) for more details and Head et al. (1995) for an application.

[21] We combine information on parents' reported primary activity at 4-digit NAICS rev.2012 with 2002 U.S. I-O tables. In particular, we derive all possible industries' combinations using the 425 IO2002 sectors (IO2002 industry 814000 'Private Households' has been excluded) and we match them with the main industry in which each parent is active.

to final end product). $\mathbf{1}(\rho_j > \rho_{med})$ is a latent variable based on the elasticity of demand faced by the parent company, taking the value 1 in the case of complements' industries and 0 in the case of substitutes' industries according to the theory of Antràs and Chor (2013). C_{k_i} is the contractibility for actual and potential affiliates' activities following the methodology of Nunn (2007) which proxies the average specificity of the input being transacted. Finally, γ_j are a full set of parent-level fixed effects. Standard errors are clustered by parent.

Table 3: Fixed effects conditional logit baseline estimation results.

Dep. Var.: Yes/No Input i integrated by parent j	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	CL	CL	CL	CL	CL	CL	CL
Relevance Index$_{k_i k_j}$	12.8***	12.9***	11.3***	11.0***	10.2***	9.81***	9.56***
	(0.029)	(0.030)	(0.051)	(0.052)	(0.070)	(0.12)	(0.12)
Complements$_{k_j}$*Relevance Index$_{k_i k_j}$					1.63***	0.34*	2.57***
					(0.11)	(0.18)	(0.17)
Downstreamness$_{k_i}$		-0.81***	-1.18***	-1.40***	-1.37***	-0.10	-2.06***
		(0.030)	(0.055)	(0.073)	(0.069)	(0.099)	(0.16)
Complements$_{k_j}$*Downstreamness$_{k_i}$			0.18**	0.15	0.044	-0.42***	0.51**
			(0.082)	(0.11)	(0.11)	(0.15)	(0.22)
Contractibility$_{k_i}$				-0.87***	-0.94***	-1.49***	-0.39***
				(0.035)	(0.034)	(0.066)	(0.046)
Observations	8,564,068	8,402,090	2,856,445	2,173,242	2,173,242	867,531	975,114
Pseudo R-squared (McFadden's)	0.405	0.409	0.234	0.294	0.297	0.215	0.324
Parent-level FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent BEC category	All	All	All	All	All	Final goods	Intermediates

Errors clustered by parent in parentheses * p-value<0.1, ** p-value<0.05, *** p-value<0.01.

We present nested results in Table 3. Results show that all else equal, with a unit increase in the *Relevance Index*, the odds of suppliers being integrated versus being potentially outsourced increase (by a factor e^{β_1}). That is to say, a more relevant supplier, i.e. in a more strategic position with a high technological relevance, as identified from the relative distance from its output in the supply network, is more likely to be integrated by the parent company. When we control for the role of the elasticity of substitution in interaction with the *Relevance Index*, we find that its marginal effect is positive and significant in all specifications. Therefore, it seems that when the output demand is elastic or inputs are not particularly substitutable, for a firm could be optimal to incorporate strategic suppliers which may not honour their contractual commitment while contracting with the others more prone to undertake relation-specific investments. The effect is positive and significant in all specifications, and it is stronger for final goods parents, i.e. *downstream* parents, when compared to intermediate goods parents, i.e. *midstream* parents. Then, we find that the absolute positioning of an input with respect to final use on a supply chain, i.e. the downstreamness, seems to have a weaker explanatory power in integration choices. In particular, we find that overall parents tend to integrate fewer production stages closer to final demand, but this is particularly true for intermediate goods producer, while the coefficient loses significance restricting the sample to *midstream* parents. Further, the marginal effect of the parent demand elasticity combined with the downstreamness is overall not significant. However, when we differentiate in columns (6) and (7) the output of parents between final and intermediate goods, we find higher propensity toward integration of suppliers further downstream from *midstream* parents when the output demand is somewhat elastic. On the contrary, a negative marginal effect of $Complements_{k_j}$ on $Downstreamness_{k_i}$ is observed for *downstream* parents, which is not in line with the model's prediction of Antràs and Chor (2013). Interestingly, the coefficient of $Contractibility_{k_i}$ is negative and significant at 1% level. As expected, the degree of contractibility plays a central role in integration decisions (Broda and Weinstein, 2006). In particular, restricting to our analysis, the more input is customized to a specific output production, the lower should be the relative contractual frictions by suppliers, hence establishing arm's length contracts should be the optimal choice.

5.2 Technological relevance and multiple affiliates

We can observe from our sample that more than one affiliate performs a production stage within a firm boundary. That is, a parent can establish more than one affiliate in an input industry. We test here whether the number of affiliates active in an input industry depends on the technological relevance of that input in the production network. We use a negative binomial regression model to estimate the following equation:

$$\begin{aligned} N_{i(j)k_i} = \beta_0 + \beta_1 RI_{k_i k_j} + \beta_2 D_{k_i} + \beta_3 \mathbf{1}(\rho_{k_j} > \rho_{med}) + \beta_4 \mathbf{1}(\rho_{k_j} > \rho_{med}) \cdot D_{k_i} \\ + \beta_5 \mathbf{1}(\rho_{k_j} > \rho_{med}) \cdot RI_{k_i k_j} + \beta_6 Size_j + \beta_7 C_{k_i} + \varepsilon_{i(j)k_i k_j} \end{aligned} \tag{6}$$

where now, we model $N_{i(j)k_i}$ as a count variable indicating the number of integrated affiliates operating in input's industry k_i. In addition to explanatory variables as in (5), we include in $Size_j$ the parent size categories (small, medium, medium-large, large, very large). Standard errors are clustered by parent.

Results in Table 4, show that a unit change in the *Relevance Index* increases the expected number of integrated affiliates in an industry k_i (by a factor e^{β_1}), holding all other variables constant. In other words, as suppliers provide directly or indirectly important inputs for the production of the final output, thus they have a high technological relevance, parents internally establish more than one producer of those inputs. Moreover, we find that the demand elasticities are not a significant driver of intensity integration as also pointed out in Rungi et al. (2017). The marginal effect of demand elasticity on $Downstreamness_{k_i}$ and the $Relevance\ Index_{k_i k_j}$ are both not significant. Interestingly, the coefficient of input con-

Table 4: Negative binomial baseline estimation results.

Dep. Var.: N. integrated affiliates in industry k_i	(1) NBRM	(2) NBRM	(3) NBRM	(4) NBRM	(5) NBRM	(6) NBRM	(7) NBRM	(8) NBRM
Relevance Index$_{k_i k_j}$	0.39***	0.39***	0.55***	0.32*	0.47	1.13***	1.46***	1.19***
	(0.11)	(0.11)	(0.16)	(0.18)	(0.31)	(0.12)	(0.24)	(0.18)
Complements$_{k_j}$*Relevance Index$_{k_i k_j}$					-0.26	0.046	-0.35	0.14
					(0.37)	(0.16)	(0.30)	(0.23)
Downstreamness$_{k_i}$		0.29**	0.53*	1.11***	1.16***	0.75***	1.43***	0.017
		(0.14)	(0.27)	(0.31)	(0.29)	(0.15)	(0.22)	(0.41)
Complements$_{k_j}$			0.35*	0.11	0.18	-0.15	0.024	-0.35*
			(0.20)	(0.22)	(0.20)	(0.11)	(0.22)	(0.20)
Complements$_{k_j}$*Downstreamness$_{k_i}$			-0.38	-0.23	-0.29	0.018	-0.060	0.22
			(0.35)	(0.39)	(0.37)	(0.19)	(0.32)	(0.43)
Contractibility$_{k_i}$				0.81***	0.84***	0.47***	-0.021	0.63***
				(0.10)	(0.10)	(0.062)	(0.18)	(0.086)
Medium parent						0.77***	0.78***	0.76***
						(0.013)	(0.023)	(0.022)
Medium-large parent						1.55***	1.48***	1.62***
						(0.030)	(0.051)	(0.045)
Large parent						2.21***	2.11***	2.31***
						(0.040)	(0.055)	(0.063)
Very large parent						2.72***	2.79***	2.65***
						(0.056)	(0.085)	(0.072)
Constant	1.44***	1.27***	1.05***	0.74***	0.68***	-0.71***	-1.16***	-0.41**
	(0.038)	(0.084)	(0.16)	(0.18)	(0.17)	(0.085)	(0.15)	(0.18)
ln(α) (Dispersion)	0.45***	0.44***	0.39***	0.43***	0.43***	-0.15***	-0.090	-0.17***
	(0.025)	(0.025)	(0.029)	(0.034)	(0.033)	(0.037)	(0.058)	(0.054)
Observations	29,621	29,535	12,512	9,446	9,446	9,446	4,032	4,201
Pseudo R-squared	0.001	0.002	0.003	0.011	0.011	0.134	0.128	0.141
BEC category of parent	All	All	All	All	All	All	Final goods	Intermediates

Errors clustered by parent in parentheses * p-value<0.1, ** p-value<0.05, *** p-value<0.01
Small parent is set as base category.

tractibility is positive and significant, although mainly for *midstream* parents. One possible explanation could be that proposed by Nunn and Trefler (2008) in the framework of foreign supplier's inputs. That is, an improvement in the degree of contractibility has two effects: the 'standard' effect that encourages arm's length contract because it is easier to reach a deal with a supplier also abroad thanks to reduced contractual uncertainties; and the 'surprise' effect that causes the most productive arm's length relationships to be vertically integrated, hence increasing the number of integrated affiliates producing a certain input. Which one of the two effects dominates is an empirical question, and from our sample, the surprising effect seems to prevail. Finally, as expected, the larger the size of the parent the higher is the number of integrated affiliates in industry k_i. For instance, medium size parents have 2.16 ($e^{0.77}$) affiliates more in industry k_i than small parents, while very large parents have almost 15 ($e^{2.72}$) more.

5.3 Robustness checks

We explore the robustness of the results across different sample compositions. In Table 5 and Table 6, we repeat the exercise for the fixed effect conditional logit and the negative binomial models. In the first and second columns, we include only observations where the industry of the actual or potential affiliate is different from the industry of the parent; in the third and fourth columns, we consider just manufacturing actual or potential affiliates; in columns (5) and (6) we alternatively compute the variable $Complements_{k_j}$ as a latent variable equal to 1 if $\rho_j > \alpha_i$, where α_i is the median elasticity of input industries, a closer definition of the Antràs and Chor (2013) model; finally, in the last columns we restrict the sample to the top 100 inputs of each parent's output, as from the contribution indicated by the direct requirement coefficient in the U.S. 2002 I-O tables. The baseline results for the *Relevance Index* still hold, although the magnitude and the significance varies according to the different subsamples.

Table 5: Fixed effects conditional logit estimation results, different sample compositions.

Dep. Var.: Yes/No Input i integrated by parent j	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Industry k_i $\neq$ Industry k_j	Industry k_i $\neq$ Industry k_j	Only manuf affiliates	Only manuf affiliates	Output & input elast	Output & input elast	Top 100 inputs	Top 100 inputs
Relevance Index$_{k_ik_j}$	48.6***	8.31***	8.48***	5.82***	10.3***	8.38***	8.44***	8.89***
	(1.17)	(0.15)	(0.11)	(0.12)	(0.063)	(0.16)	(0.083)	(0.46)
Downstreamness$_{k_i}$	-0.70***	-1.01***	-0.42***	-0.14*	-1.52***	-0.65***	-0.28**	0.50***
	(0.079)	(0.051)	(0.082)	(0.082)	(0.076)	(0.062)	(0.14)	(0.11)
Complements$_{k_j}$*Downstreamness$_{k_i}$	0.0080	0.25***	-0.46***	-0.19	0.36***	0.50***	0.89***	0.96***
	(0.13)	(0.079)	(0.13)	(0.13)	(0.11)	(0.10)	(0.26)	(0.17)
Contractibility$_{k_i}$	-1.19***		-0.44***		-0.99***		-0.57***	
	(0.047)		(0.035)		(0.035)		(0.042)	
Complements$_{k_j}$*Relevance Index$_{k_ik_j}$	-11.2***	3.69***	2.40***	6.63***	2.11***	4.03***	2.09***	2.08**
	(1.93)	(0.52)	(0.13)	(0.70)	(0.11)	(0.56)	(0.14)	(0.82)
Observations	1,198,376	2,856,445	1,286,016	1,286,016	2,173,242	2,173,242	309,633	521,276
Pseudo R-squared (McFadden's)	0.029	0.113	0.228	0.112	0.298	0.137	0.426	0.158
Parent-level fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
alpha variable	No	Yes	No	Yes	No	Yes	No	Yes

Errors clustered by parent in parentheses * p-value<0.1, ** p-value<0.05, *** p-value<0.01.

Table 6: Negative binomial estimation results, different sample compositions.

Dep. Var.: N. integrated affiliates in industry k_i	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Industry k_i $\neq$ Industry k_j	Industry k_i $\neq$ Industry k_j	Only manuf affiliates	Only manuf affiliates	Output & input elast	Output & input elast	Top 100 inputs	Top 100 inputs
Relevance Index$_{k_ik_j}$	0.61	-1.73	1.18***	1.61***	0.83***	2.42***	1.14***	3.01***
	(3.34)	(1.73)	(0.11)	(0.19)	(0.14)	(0.60)	(0.13)	(0.48)
Downstreamness$_{k_i}$	1.01***	0.41**	0.76***	0.39***	0.55***	0.43***	0.76***	0.33
	(0.19)	(0.18)	(0.12)	(0.12)	(0.16)	(0.15)	(0.20)	(0.21)
Complements$_{k_j}$	-0.12	-0.041	-0.17	-0.18	-0.024	0.11	-0.36***	-0.082
	(0.14)	(0.15)	(0.12)	(0.12)	(0.12)	(0.11)	(0.13)	(0.13)
Complements$_{k_j}$*Downstreamness$_{k_i}$	0.025	-0.015	0.056	0.15	-0.0062	-0.096	-0.32	-0.050
	(0.25)	(0.25)	(0.20)	(0.20)	(0.19)	(0.19)	(0.25)	(0.26)
Contractibility$_{k_i}$	0.64***		0.49***		0.24***		0.61***	
	(0.11)		(0.063)		(0.078)		(0.054)	
Complements$_{k_j}$*Relevance Index$_{k_ik_j}$	-3.51	1.59	-0.045	-0.62***	-0.072	-1.83***	0.61***	-1.71***
	(4.04)	(2.61)	(0.15)	(0.21)	(0.18)	(0.61)	(0.18)	(0.48)
Medium parent	0.71***	0.64***	0.77***	0.75***	0.77***	0.77***	0.78***	0.73***
	(0.017)	(0.015)	(0.013)	(0.012)	(0.014)	(0.013)	(0.016)	(0.016)
Medium-large parent	1.29***	1.03***	1.56***	1.50***	1.59***	1.58***	1.66***	1.34***
	(0.039)	(0.032)	(0.030)	(0.028)	(0.035)	(0.034)	(0.034)	(0.031)
Large parent	1.83***	1.55***	2.22***	2.14***	2.25***	2.23***	2.46***	1.95***
	(0.050)	(0.037)	(0.040)	(0.038)	(0.044)	(0.044)	(0.051)	(0.038)
Very large parent	2.22***	2.06***	2.73***	2.67***	2.79***	2.77***	2.99***	2.54***
	(0.064)	(0.051)	(0.056)	(0.057)	(0.053)	(0.054)	(0.088)	(0.066)
Constant	-0.61***	-0.21**	-0.74***	-0.33***	-0.50***	-0.36***	-0.78***	-0.40***
	(0.11)	(0.10)	(0.071)	(0.064)	(0.095)	(0.087)	(0.090)	(0.094)
ln(α) (Dispersion)	0.070*	0.087**	-0.15***	-0.11***	-0.11**	-0.093**	-0.45***	-0.19***
	(0.037)	(0.034)	(0.037)	(0.038)	(0.042)	(0.043)	(0.042)	(0.036)
Observations	6,074	9140	9,446	9446	6,296	6296	6,203	8607
Pseudo R-squared	0.085	0.060	0.134	0.128	0.136	0.134	0.178	0.118
alpha variable	No	Yes	No	Yes	No	Yes	No	Yes

Errors clustered by parent in parentheses * p-value<0.1, ** p-value<0.05, *** p-value<0.01.

Besides, we explore our results using an alternative formula to compute the *Relevance Index*, where we set the tuning parameter, now α_i equals to the contractibility value of each input when available.[22] Results in Table 5 confirm that more vertical integration occurs when an input has been detected to be important in the oriented supply network. Eventually, Table 6 corroborates the increasing number of affiliates active in the relevant industry.

Table 7: Fixed effects conditional logit with α_i in the computation of the *Relevance Index*.

Dep. Var.: Yes/No Input i integrated by parent j	(1)	(2)	(3)	(4)	(6)	(7)
	CL	CL	CL	CL	CL	CL
Relevance Index$_{k_i k_j}$	13.3***	13.5***	9.77***	8.31***	9.94***	6.56***
	(0.044)	(0.047)	(0.16)	(0.15)	(0.39)	(0.15)
Downstreamness$_{k_i}$		-0.76***	-0.96***	-1.01***	0.24***	-2.20***
		(0.031)	(0.052)	(0.051)	(0.083)	(0.10)
Complements$_{k_j}$*Downstreamness$_{k_i}$			0.19**	0.25***	-0.47***	0.67***
			(0.078)	(0.079)	(0.11)	(0.15)
Complements$_{k_j}$*Relevance Index$_{k_i k_j}$				3.69***	13.7***	3.81***
				(0.52)	(2.66)	(0.39)
Observations	8,564,068	8,402,090	2,856,445	2,856,445	1,139,175	1,275,295
Pseudo R-squared (McFadden's)	0.358	0.362	0.110	0.113	0.073	0.162
Parent-level fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
BEC category of parent	All	All	All	All	Final goods	Intermediates

Errors clustered by parent in parentheses * p-value<0.1, ** p-value<0.05, *** p-value<0.01.

[22]See the Appendix A.1 for more details.

Table 8: Negative binomial regressions with α_i in the computation of the *Relevance Index*.

Dep. Var.: N. integrated affiliates in industry k_i	(1) NBRM	(2) NBRM	(3) NBRM	(4) NBRM	(5) NBRM	(6) NBRM	(7) NBRM	(8) NBRM
Relevance Index$_{k_i k_j}$	0.49***	0.49***	0.91***	0.91***	0.74	2.97***	6.14***	1.72***
	(0.10)	(0.097)	(0.12)	(0.12)	(0.70)	(0.49)	(1.23)	(0.37)
Downstreamness$_{k_i}$		0.27**	0.50*	0.50*	0.48*	0.41**	1.32***	-0.46
		(0.14)	(0.29)	(0.29)	(0.28)	(0.16)	(0.31)	(0.34)
Complements$_{k_j}$			0.19	0.19	0.18	-0.0048	0.14	-0.19
			(0.21)	(0.21)	(0.20)	(0.12)	(0.25)	(0.18)
Complements$_{k_j}$*Downstreamness$_{k_i}$			-0.25	-0.25	-0.24	-0.053	-0.12	0.17
			(0.36)	(0.36)	(0.35)	(0.20)	(0.36)	(0.39)
Complements$_{k_j}$*Relevance Index$_{k_i k_j}$					0.18	-1.77***	-5.34***	-0.48
					(0.71)	(0.49)	(1.23)	(0.38)
Medium parent						0.72***	0.74***	0.72***
						(0.013)	(0.022)	(0.020)
Medium-large parent						1.29***	1.28***	1.39***
						(0.027)	(0.046)	(0.041)
Large parent						1.85***	1.83***	1.99***
						(0.031)	(0.045)	(0.051)
Very large parent						2.42***	2.56***	2.40***
						(0.043)	(0.070)	(0.052)
Constant	1.42***	1.26***	1.10***	1.10***	1.11***	-0.40***	-1.08***	0.037
	(0.035)	(0.083)	(0.17)	(0.17)	(0.16)	(0.090)	(0.22)	(0.15)
ln(α) (Dispersion)	0.44***	0.44***	0.38***	0.38***	0.38***	-0.040	-0.028	-0.015
	(0.025)	(0.025)	(0.031)	(0.031)	(0.031)	(0.033)	(0.054)	(0.046)
Observations	29,621	29,535	12,512	12,512	12,512	12,512	5,490	5,413
Pseudo R-squared	0.002	0.003	0.006	0.006	0.006	0.099	0.097	0.104
BEC category of parent	All	All	All	All	All	All	Final parent	Intermediates

Errors clustered by parent in parentheses * p-value<0.1, ** p-value<0.05, *** p-value<0.01.
Small parent is set as base category.

6 Conclusion

In this contribution, we first argue that the assumption of linearity for the technological production process is not realistic. A network approach is more appropriate, to catch the complexity of sourcing strategies originated by the unbundling of production of final products from the production of intermediate inputs. First, global supply networks go beyond such linearity incorporating all kind of configurations that are possibly variable across industries. Second, they encompass all relevant sets of direct and indirect relationships capturing the multi-dimension nature of production. We show that each industry along what is assumed to be a supply chain actually can hide complex production functions, in which some inputs can play a role at different moments of the intermediate manufacturing, before reaching final consumers. As a consequence, production networks can differ across industries because sourcing strategies are implicitly heterogeneous.

In this framework, we introduce a new industry-pair measure, the *Relevance Index*, which reflects the position in the supply network of direct and indirect inputs for the production of an output. We aim at capturing the technological relevance of *inputs of inputs* within a production network, which depends on the extent to which an input is accessible and replaceable. Moreover, we consider a higher weight should be given to inputs that serve more than one supply network at the same time or enter the production process at multiple stages. Thereafter, we show how network-based measures fit better than chain-like metrics in explaining firm boundaries. We build our empirical analysis on the property-rights model of the firm boundaries by Antràs and Chor (2013). In particular, we test the validity of the *Relevance Index* on explaining the role of production stages performed by affiliates integrated into the firm boundaries. First, we find that our measure is a good predictor of vertical integration decisions which are the outcome of formally organized rules and conventions. Second, we notice more affiliates will be vertically integrated into a relevant input industry. Overall, we find evidence of the main prediction of Antràs and Chor (2013) model, according to which the ownership decisions of a firm depend on the position of an input industry in the entire production process. However, we notice that the elasticity of demand faced by the parent is not a significant determinant of integration choices if we remove the assumption of a linear sequence of stages of production where the final-output producer is located at the end.

Further research on the organization of supply networks should investigate more the dynamic nature of firms' boundaries geographically and organizationally. Until now, we are considering choices neutral to country-level factors and the evolution of production networks over time is not taken into account.

References

D. Acemoglu, P. Antràs, and E. Helpman. Contracts and technology adoption. *The American economic review*, 97(3):916–943, 2007.

D. Acemoglu, S. Johnson, and T. Mitton. Determinants of vertical integration: financial development and contracting costs. *The Journal of Finance*, 64(3):1251–1290, 2009.

D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi. The network origins of aggregate fluctuations. *Econometrica*, 80(5):1977–2016, 2012.

P. Aghion and R. Holden. Incomplete contracts and the theory of the firm: What have we learned over the past 25 years? *The Journal of Economic Perspectives*, 25(2):181–197, 2011.

L. Alfaro, P. Antràs, D. Chor, and P. Conconi. Internalizing global value chains: A firm-level analysis. *National Bureau of Economic Research, Working Paper No. 21582*, 2017.

C. Altomonte and A. Rungi. Business groups as hierarchies of firms: determinants of vertical integration and performance. *ECB Working Paper No. 1554*, 2013.

P. Antràs and D. Chor. Organizing the global value chain. *Econometrica*, 81(6):2127–2204, 2013.

P. Antràs and A. de Gortari. On the geography of global value chains. *National Bureau of Economic Research, Working Paper No. 23456*, 2017.

P. Antràs and E. Helpman. Contractual frictions and global sourcing. *National Bureau of Economic Research, Working Paper No. 12747*, 2006.

P. Antràs and S. R. Yeaple. *Multinational Firms and the Structure of International Trade*, volume 4. 2014.

P. Antràs, D. Chor, T. Fally, and R. Hillberry. Measuring the upstreamness of production and trade flows. *The American Economic Review*, 102(3):412–416, 2012.

E. Atalay, A. Hortacsu, J. Roberts, and C. Syverson. Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202, 2011.

R. Baldwin. Globalisation: the great unbundling (s). *Economic Council of Finland*, 20(3): 5–47, 2006.

R. Baldwin and J. Lopez-Gonzalez. Supply-chain trade: A portrait of global patterns and several testable hypotheses. *The World Economy*, 38(11):1682–1721, 2015.

R. Baldwin and A. J. Venables. Spiders and snakes: offshoring and agglomeration in the global economy. *Journal of International Economics*, 90(2):245–254, 2013.

S. Brin and L. Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833, 2012.

C. Broda and D. E. Weinstein. Globalization and the gains from variety. *The Quarterly* *journal of economics*, 121(2):541–585, 2006.

V. M. Carvalho. A survey paper on recent developments of input-output analysis. *Complexity Research Initiative for Systemic Instabilities*, 2013.

V. M. Carvalho. From micro to macro via production networks. *The Journal of Economic Perspectives*, 28(4):23–47, 2014.

R. H. Coase. The nature of the firm. *Economica*, 4(16):386–405, 1937.

N. M. Coe, P. Dicken, and M. Hess. Global production networks: realizing the potential. *Journal of economic geography*, 8(3):271–295, 2008.

A. Costinot, J. Vogel, and S. Wang. An elementary theory of global supply chains. *Review of Economic Studies*, 80(1):109–144, 2012.

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

D. Del Prete and A. Rungi. Organizing the global value chain: a firm-level test. *Journal of International Economics*, 109:16–30, 2017.

E. Estrada and N. Hatano. Communicability in complex networks. *Physical Review E*, 77 (3):036111, 2008.

T. Fally. Production staging: measurement and facts. *Boulder, Colorado, University of Colorado Boulder, May*, pages 155–168, 2012.

J. P. Fan and L. H. Lang. The measurement of relatedness: An application to corporate diversification. *The Journal of Business*, 73(4):629–660, 2000.

R. C. Feenstra, J. Romalis, and P. K. Schott. Us imports, exports, and tariff data, 1989-2001. *National Bureau of Economic Research, Working Paper No. 9387*, 2002.

S. Formai and F. Vergara Caffarelli. Quantifying the productivity effects of global sourcing. *Bank of Italy, Working Paper No. 1075*, 2016.

L. C. Freeman, D. Roeder, and R. R. Mulholland. Centrality in social networks: Ii. experimental results. *Social networks*, 2(2):119–141, 1979.

S. J. Grossman and O. D. Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *Journal of political economy*, 94(4):691–719, 1986.

P. Harms, O. Lorz, and D. Urban. Offshoring along the production chain. *Canadian Journal of Economics/Revue canadienne d'économique*, 45(1):93–106, 2012.

O. Hart. *Firms, contracts, and financial structure*. Clarendon Press, 1995.

T. H. Haveliwala. Topic-sensitive pagerank. In *Proceedings of the 11th international conference on World Wide Web*, pages 517–526. ACM, 2002.

K. Head, J. Ries, and D. Swenson. Agglomeration benefits and location choice: Evidence from japanese manufacturing investments in the united states. *Journal of International*

Economics, 38(3-4):223–247, 1995.

R. Hudson. Conceptualizing economies and their geographies: spaces, flows and circuits. *Progress in Human Geography*, 28(4):447–471, 2004.

D. Hummels, J. Ishii, and K.-M. Yi. The nature and growth of vertical specialization in world trade. *Journal of international Economics*, 54(1):75–96, 2001.

M. Jacomy, S. Heymann, T. Venturini, and M. Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization. *Medialab Science Po Center of Reseatch*, 2011.

G. Jeh and J. Widom. Scaling personalized web search. In *Proceedings of the 12th international conference on World Wide Web*, pages 271–279. ACM, 2003.

L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

D. McFadden. Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics. Academic Press New York*, 1974.

M. Newman. *Networks: an introduction*. Oxford university press, 2010.

S. C. North. Drawing graphs with neato. *NEATO User manual 11*, 2004.

N. Nunn. Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics*, 122(2):569–600, 2007.

N. Nunn and D. Trefler. The boundaries of the multinational firm: an empirical analysis. *The organization of firms in a global economy*, pages 55–83, 2008.

E. Oberfield. Business networks, production chains, and productivity: A theory of input-output architecture. 2012.

J. R. Pierce and P. K. Schott. A concordance between ten-digit us harmonized system codes and sic/naics product classes and industries. *Journal of Economic and Social Measurement*, 37(1, 2):61–96, 2012.

N. Ramondo, V. Rappoport, and K. J. Ruhl. Intrafirm trade and vertical fragmentation in US multinational corporations. *Journal of International Economics*, 98:51–59, 2016.

J. E. Rauch. Networks versus markets in international trade. *Journal of international Economics*, 48(1):7–35, 1999.

A. Rungi, G. Morrison, and F. Pammolli. Global ownership and corporate control networks. *IMT Lucca EIC WP Series 07/2017*, 2017.

UNCTAD. World investment report 2016. *United Nations*, 2016.

S. White and P. Smyth. Algorithms for estimating relative importance in networks. In

Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, pages 266–275. ACM, 2003.

O. E. Williamson. The vertical integration of production: market failure considerations. *The American Economic Review*, 61(2):112–123, 1971.

O. E. Williamson. Markets and hierarchies: analysis and antitrust implications: a study in the economics of internal organization. *University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship*, 1975.

O. E. Williamson. Transaction-cost economics: the governance of contractual relations. *The journal of Law and Economics*, 22(2):233–261, 1979.

A Appendix

A.1 Measures of the relative importance of a node

In our contribution, we emphasise the measures of node importance in a network relative to a set of root nodes. In particular, we argue that the relevance of any input for the production of an output influences the 'make or buy' decision of a parent company in a multinational setting. The main limitation of using 'global' measures in this context, as highlighted by White and Smyth (2003), is that using algorithms such as *PageRank*, root nodes are not given any preferential treatment in the ranking. Furthermore, using such an approach, all the nodes in a sub-network with well-defined root nodes are ranked around a root node and not with respect to it. These limitations have motivated the work presented in this paper, where measures of relative importance which consider a set of prior root nodes are investigated.

Given a directed graph G and two nodes r and i, where r, $i \in G$, $RI(i|r)$ is defined as the *Relevance Index* of a node with respect to the root node. There are two common approaches to measure the relative importance of a node with respect to a target (root) node in a network. A first approach is to consider the notion of distance according to which two nodes are related if there exists a path that links them, with the relative relevance decreasing as the path length increases. In this case, a relevant question is how to choose the set of paths between the node i and the node j. Considering the shortest paths only leads in many cases to a poor approximation, since other nodes along the path are ignored despite that they could add more importance to i relative to j. Alternatively, evaluating paths of a fixed length and without edges nor nodes in common may be a better approximation, but then the relative importance is calculated only on a small set of paths in the neighbourhood of the target node. A second approach is to treat the network as a stochastic process, using first-order Markov chains. It is based on the idea of a single 'token' visiting the nodes of a graph randomly an infinite number of times. The 'token' moves to the next node according to a stochastic function that takes into account the properties of the current node. The average fraction of time spent by the 'token' on any node, i.e. the persistence probability, can be interpreted as the importance of this node with respect to all other nodes in the network. That is the idea behind the *PageRank* algorithm.

A personalised version of the *PageRank* has been proposed by Haveliwala (2002), Jeh and Widom (2003), White and Smyth (2003), where personalised ranks of the root nodes are introduced to take into account the prior knowledge. Three different type of probabilities are defined a priori:

- $\mathbf{p}_R = (p_1, ..., p_{|N|})$ is the vector of prior probabilities attached to every node, where R is the set of root nodes and $p_i = \frac{1}{|R|}$ for $i \in R$, and $p_i = 0$ otherwise, so that $p_i > 0$ and $\sum_{i=1}^{|N|} p_i = 1$;
- α is the back probability, i.e. the probability to jump back to the root set, and it has to be chosen in the $[0, 1]$ range;
- $p(i|j) = \frac{1}{D_{out}(j)}$ is the probability of transitioning from node j to node i in one step.

From the original formulation of the *PageRank*:

$$C_{PR}(i) = (1 - f) + f \sum_{j \in S_i} \frac{1}{D_{out}(j)} C_{PR}(j) \tag{7}$$

where S_i is the set of the source nodes that link to the node i, $D_{out}(j)$ is the out-degree of the node j, and f is the residual probability to account for nodes with zero out-degree, usually set equal to 0.85. Given $RI(i|r) = \pi(i)$, we can modify the previous formula by adding the probabilities defined a priori, as follows:

$$\pi(i)_t = (1 - \alpha) \left(\sum_{j \in S_i} p(i|j) \pi(j)_{t-1} \right) + \alpha p_i \tag{8}$$

where the choice of α is subjective and different values reflect the bias in the ranking process.

Proposed algorithm adaptation

The probabilities defined a priori are:

- $\mathbf{e}_r = (e_1, ..., e_{|N|})$ is the column vector of prior probabilities attached to every node, where $e_i = 1$ if $i = r$, and $e_i = 0$ otherwise, so that $e_i \geq 0$ and $\sum_{i=1}^{|N|} e_i = 1$;
- the back probability α chosen in the in the $[0, 1]$ range is interpreted as a distance factor which penalises nodes far away from the root node;
- $\mathbf{P}$ is the Markov matrix where each element is the transition probability from node i to node j. In particular, each entry d_{ij} is the direct requirement coefficient retrieved from the U.S. BEA input-output tables, i.e. the amount of the commodity required to produce one unit of the industry's output.

In our framework, a node i is considered as an input in the technological production process of a target output, represented by the root node r. The proposed *PageRank* adjustment, named the *Relevance Index*, can be computed for all pairs of nodes $i - r$, as follows:

$$\boldsymbol{\pi}_t \leftarrow (1-\alpha)\mathbf{P}\boldsymbol{\pi}_{t-1} + \alpha\mathbf{e}_r \tag{9}$$

Since we are interested in the steady-state condition $\boldsymbol{\pi}_t = \boldsymbol{\pi}_{t-1} \triangleq \boldsymbol{\pi}_r$, the above relation becomes

$$\boldsymbol{\pi}_r = (1-\alpha)\mathbf{P}\boldsymbol{\pi}_r + \alpha\mathbf{e}_r, \tag{10}$$

for which the solution is

$$\boldsymbol{\pi}_r = \alpha[\mathbf{I} - (1-\alpha)\mathbf{P}]^{-1}\mathbf{e}_r \tag{11}$$

The final result is a matrix composed of persistence probability vectors, i.e. $\boldsymbol{\pi}_r$, where each column contains the ranking of all the inputs used in the production of a particular output according to their importance.

The parameter α specifies the degree to which the computation is biased toward prior probabilities. The value of α is empirically chosen, and it spreads uniformly over the rank. For $\alpha = 1$, a meaningless distribution is concentrated in the root nodes, while other nodes have a null value. Conversely, as α tends to 0, root nodes become irrelevant to the final ranking.

We choose to use $\alpha = 0.5$ heuristically in our computations, after inspecting the values of the *Relevance Index* for several values of α. In general, for well-connected networks the value of α becomes irrelevant. The input-output networks, examined in this paper, have a rather high degree of connectivity, as confirmed by the density measure. Therefore the choice of α should not bias the analysis.

As a further robustness check, we substitute α with a vector α_i in the above relation (9), allowing the parameter to change across the inputs. The alternative *Relevance Index* formula is:

$$\boldsymbol{\pi}_t = (1-\alpha_i)\mathbf{P}\boldsymbol{\pi}_{t-1} + \alpha_i\mathbf{e}_r \tag{12}$$

In the empirical analysis we set α_i equals to the contractibility values of each input when available, and we choose to use $\alpha_i = 0.5$ heuristically otherwise in our computation.

B Data Appendix

B.1 Variables Description

Downstreamness: This variable is calculated by Antràs and Chor (2013) based on data from the 2002 U.S. I-O tables. They propose two alternative measures the 'downstreamness' of an industry in a production process. The *DUse_TUse* is the ratio of aggregate direct use to aggregate total use of a particular input industry i, where the direct use is the value of goods from industry i to industry j to produce goods for final use, while the total use is the value of goods from industry i used either directly or indirectly in producing industry's j output for final use. The higher is the *DUse_TUse* for a given industry i, the most of the contribution occurs relatively far downstream. Alternatively, the *DownMeasure* corresponds to a weighted index of the average position in the value chain at which an industry's output is used, with the weights being given by the ratio of the use of that industry's output in that position relative to the total output of that industry.

Demand elasticity: It was computed by Antràs and Chor (2013) based on the widely used U.S. import demand elasticities for Harmonized System ten digit (HS10) products computed by Broda and Weinstein (2006). These were merged with a comprehensive list of HS10 codes from Pierce and Schott (2012). For each HS10 code missing an elasticity value, they assigned a value equal to the trade-weighted average elasticity of the available HS10 codes with which it shared the same first nine digits. This was done successively up to codes that shared the same first two digits, to fill in as many HS10 elasticities as possible. Using the IO-HS concordance provided by the BEA with the 2002 U.S. I-O tables, they then took the trade-weighted average of the HS10 elasticities within each IO2002 category. At each stage, the weights used were the total value of U.S. imports by HS10 code from 1989– 2006, calculated from Feenstra et al. (2002). This yielded import elasticities for the industry that sells the input in question. For the average buyer elasticity, they took a weighted average of the elasticities of industries that purchase the input in question, with weights equal to these input purchase values as reported in the 2002 U.S. I-O tables.

Complements (Substitutes): We follow the procedure set by Antràs and Chor (2013) in splitting the parents' industries into complements and substitutes. Using demand elasticity from Broda and Weinstein (2006) and elaborated by Antràs and Chor (2013), we check when the elasticity is respectively higher or lower than the sample median. In a robustness check, the subset of parents' industries with above-median (below-median) affiliates' demand elasticity are labelled as complements (substitutes).

Contractibility: Nunn (2007) proposes a measure of contract intensity based on the proportion of an industry's intermediate inputs that are customized, i.e. neither reference priced nor traded on an organized exchange. He builds on the classification of goods of Rauch (1999) which identifies goods sold on organized exchanges, reference priced in trade publications, and all residuals assumed to be customized. Then, using U.S. I-O Use tables he infers the extent to which each industry's intermediate inputs are relation-specific. In our analysis, we use the measure of *Contractibility* computed by Antràs and Chor (2013) using the 2002 U.S. I-O tables.

Size: It is computed at the parent level from our sample, sourced from the Orbis database. The size categories (small, medium, medium-large, large, very large) are based on a combination of criteria: revenues, or total assets, or number of employees, or capitalization, or listed on the stock exchange.

B.2 Descriptive Statistics

Table 9: Control variables.

	Obs.	mean	median	sd	min	max
Relevance Index ($\alpha = 0.5$)	146,934	0.25	0.04	0.26	0	0.64
Relevance Index (α_i)	146,934	0.24	0.03	0.28	0	0.93
Parents						
Elasticity of substitution	6,900	7.56	4.24	8.32	1.30	84.19
Affiliates						
Downstreamness	152,892	0.56	0.50	0.21	0.22	1
Elasticity of substitution	59,666	9.13	4.75	12.14	1.30	108.50
Contractibility	59,666	0.29	0	0.41	0	1

Piazza San ponziano 6, 55100 Lucca, Italy.
www.imtlucca.it