Measuring the Input Rank in Global Supply Networks

Armando Rungi* Loredana Fattorini†

This version: May 2018

Abstract

In this paper, we introduce the *Input Rank* as a measure to study the organization of supply networks at the firm level. We assume that a Markov process of exploration may be started by a producer throughout her web of direct and indirect suppliers, to assess the technological relevance of each direct and indirect input, when her ability to outreach in the supply network may be limited. Therefore, each producer ends up with an input-output eigenvector centrality, which is higher when a direct or indirect input is relatively more requested to produce other direct or indirect inputs, and when that input is relatively more requested to produce other highly-requested inputs. Finally, we compute the *Input Rank* on U.S input-output tables and test its empirical validity for choices of vertical integration on a dataset made of 20,489 U.S. parent companies controlling 154,836 affiliates worldwide. Results show that a higher *Input Rank* is positively associated to a higher probability that that input is vertically integrated, relatively more when the demand faced by the parent company is more elastic. We argue that a producer reduces the risk of disruption in her supply network when a central input is vertically integrated. In this framework, the *Input Rank* is at least complementary to previous sequential metrics (e.g. *upstreamness* or *downstreamness*), because it better catches the recursive nature of real-world supply networks, whereas linear technological sequences may be just corner solutions.

Keywords: production networks, global value chains, Markov chains, input output, vertical integration, multinational enterprises

JEL codes: F23, L23, D23, C63, C67.

* Corresponding author: armando.rungi@imtlucca.it. Laboratory for the Analysis of Complex Economic Systems, piazza San Francesco 19 - 55100 Lucca - Italy.

† Mail to loredana.fattorini@imtlucca.it. Laboratory for the Analysis of Complex Economic Systems, piazza San Francesco 19 - 55100 Lucca - Italy.

1 Introduction

Modern economies are organized as webs of specialized producers. Each company can be plunged into a supply network that starts with the idea of a product in engineering, design or research labs. After that, parts and components are manufactured and assembled, then they reach producers of final products who require the services of marketing, advertising and distribution companies to get to the market.

In fact, the configuration of production processes can be much complex and recursive in nature, when the same intermediate goods and services are repeatedly needed along the supply network, at different stages of the production process. Take logistics and distribution services, which are crucial in the delivery of parts and components to other companies, as well as in the case of final goods destined to consumers. Raw materials are the basis of so many manufactured inputs and outputs. Innovation can require the services of R&D labs, engineering and design at various stages during the production process.

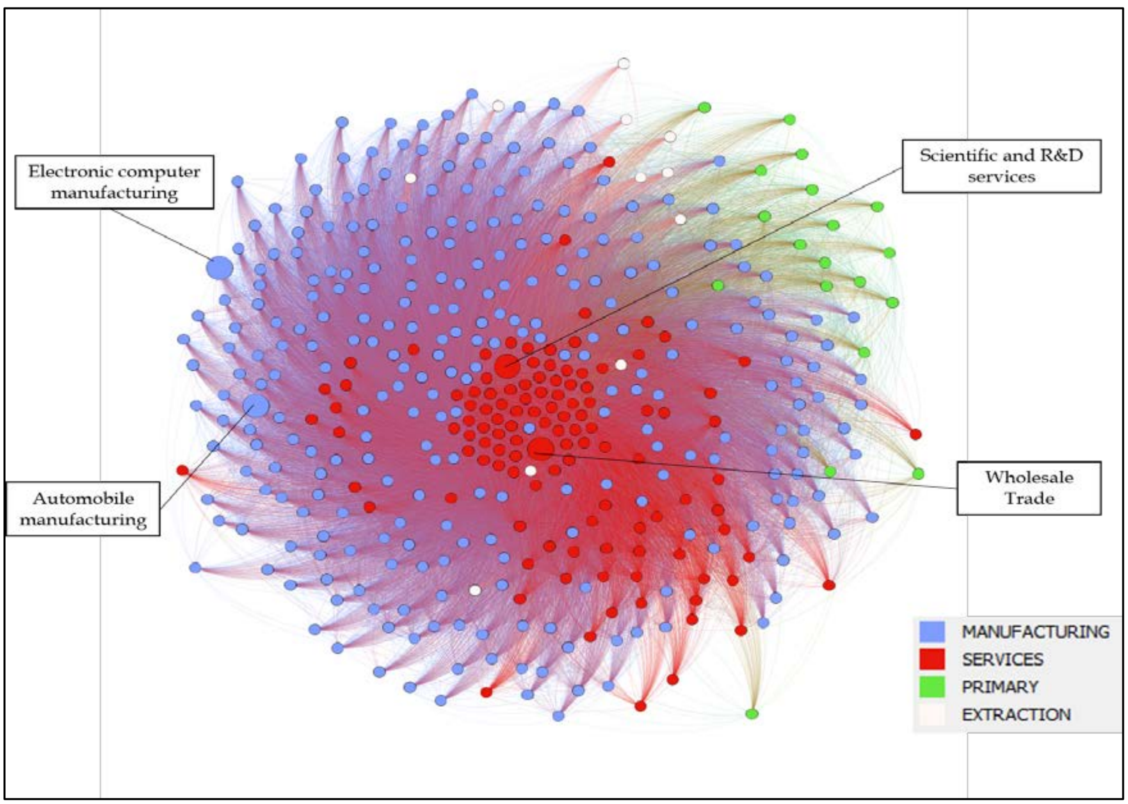
Against this background, in recent decades, supply networks have been increasingly fragmenting on a global scale since a process of unbundling started, due to the dramatic advances in transportation and communication technologies that scattered production stages across different countries (Baldwin, 2016).

Although fragmentation can originate either spider-like or snake-like configurations, depending on the technological peculiarities of the production processes (Baldwin and Venables, 2013), the international organization of production has been mainly studied after assuming a separation of production stages along ordered and linear sequences. That is, complex production networks have been proxied as snake-like processes running from the conception of the product to its delivery for final usage (Costinot, Vogel and Wang, 2013; Antràs and Chor, 2013; Fally and Hillberry, 2015; Alfaro et al., 2017; Antràs and de Gortari, 2017), therefore neglecting the spider-like nature of the organization of production for sake of simpler assumptions on theory and empirics. Specialization patterns by country along global value chains (GVC) and firm-level choices of vertical integration have been both envisaged as partitions of ideally linear segments oriented on *upstream-downstream* directions.

So far, empirical efforts have followed theory when proposing positioning metrics, e.g. the *upstreamness* or *downstreamness* of a production stage, which simulate a technological sequence constructed on input-output tables (Fally, 2012, Antràs et al., 2012, Antràs and Chor, 2013, Alfaro et al., 2017, Miller and Temurshoev, 2017, Wang et al., 2017, Antràs and Chor, 2017). Albeit an advancement for understanding the interdependence among buyers and suppliers, linear circuits certainly underestimate the relative central importance of some inputs, which can magnify or dampen a shock in presence of technological loops, kinks and corners.

Take the case of the U.S. economy, which we plot as a production network in Figure 1. According to the U.S. BEA 2002 input-output tables, the U.S. economy can be represented as a collection of 425 industries (i.e., nodes) linked by 51,768 transactions (i.e., edges). In Figure 1, we organize U.S. industries on a two-dimension space according to their reciprocal connectivity, following a Fruchterman and Reingold (1991) layout, which in our case posits more requested inputs at the center stage. Interestingly, services industries make the core of the U.S production network because they are used as direct inputs in many other manufacturing and services industries. Primary industries, like agriculture and forestry, are rather peripheral and mostly located in the north-west area of the graph. Among services, let us pick the case of R&D (code 541700) and Wholesale Trade (code 541800), which seem to be among the most connected industries. In fact, wholesalers have a prominent role in professionally distributing many intermediate inputs in different moments of the production process, whereas R&D services are pivotal in fostering innovation across most U.S. sectors. Let us consider now the case of two consumer goods industries: Electronic Computer Manufacturing (code 334111) and Automobile Manufacturing (code 336111). Although their products can be used as capital goods in some other sectors, they appear to be at the periphery of the U.S. production network, because their final usage prevails on the intermediate usage.

Figure 1: Input-Output Network from U.S. BEA 2002 I-O tables, and selected industries

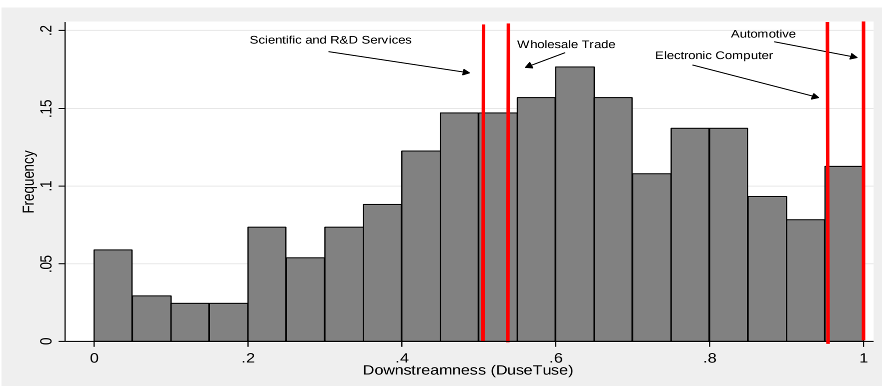


Note: Nodes represent 425 6-digit NAICS industries from the U.S. Bureau of Economic Analysis 2002 Input-Output tables. Edges represent 51,768 industry-pair transactions. The graph is visualized using a Fruchterman and Reingold (1991) layout using the GEPHI software. More connected industries (weighted out-degree) at the center of the graph. Selected industries in evidence.

In fact, at a first glance, Figure 1 shows a rather compact network with a relatively high density, i.e., the fraction of actual linkages out of all potential linkages is 0.286. The average path length connecting any two industries is just 1.7 links, pointing to a *small-world* nature of the US economy. Briefly, on average, any producer sources inputs from most of the other industries, either directly or indirectly. Indeed, the network of Figure 1 is not separable: it is self-contained in a unique connected component, and it is always possible to run seamlessly from one node to another, just following input linkages.

Once we compare the positions of selected industries in the production network with their position on the *downstreamness* segment (Antràs and Chor, 2013), in Figure 2, we curiously find that both R&D and Wholesale Trade are in the middle of the ideally linear supply chain. This is in contrast with the stylized chain we may have in mind, where a representative business line would start with R&D services and ends with distributions services. In fact, when we review computation methodologies, we find that *downstreamness* segments are essentially derived from the weighted relative usage of the tasks, intermediate *vis à vis* final, in one or more industries, therefore confounding the distance from the final demand and the central role they may have across different production processes, when the actual production network is collapsed on a segment.[1]

Figure 2: *Downstreamness* of selected industries from U.S. BEA 2002 I-O tables



Note: *Downstreamness (DuseTuse)* sourced from Antràs and Chor (2013). Frequency indicates how many industries out of total 425 from U.S. input-output tables are found in that position. Selected industries: Scientific Research and Development Services (code 541700, value 0.504); Wholesale Trade (code 541800, value 0.666); Electronic Computer Manufacturing (code 334111, value 0.959); Automobile Manufacturing (code 336111, value 0.999).

Instead, we argue that the mutually interactive and recursive nature of production processes is better understood when we consider not only how inputs enter in a different order (downstream

[1] More recently, Alfaro et al. (2017) compute a *Relative Upstreamness* to consider the heterogeneity of input positions oriented towards different outputs. However, also in this case, the position of R&D services is on average located in the middle of the output-specific technological sequences, i.e., the average *upstreamness* value is 3.044 for an indicator that originally ranges approximately from 1 to 8.9.

vs upstream), but also how central they are because they are required more than once along the same production processes (as inputs of inputs).

In the end, a bird's eye view of the U.S. production network represented in Figure 1 returns an idea of a 'global' centrality for each industry, for example rendering the crucial role of the R&D services for the entire U.S. economy. However, what we are actually interested in is a measure of the 'local' technological relevance of any (direct or indirect) input for the completion of a specific target output.

In this respect, we introduce the *Input Rank* as a solution to a Markov stochastic process that ranks direct and indirect inputs oriented towards a target output, once assuming that producers have limited information on indirect transactions and scarce time to outreach the intricate web of direct and indirect suppliers. For this, we get our inspiration from the *Page Rank* centrality, which is a measure originally used in social networks and search engines to assess the relevant content of information (Brin and Page, 1998). The tool has by now spread to many different domains[2], from biology and genetics, to financial debts, bibliometrics and engineering of road networks (Gleich, 2015).

In our perspective, the *Input Rank* can be seen as the result of a Markov stochastic process started by a producer that is embedded in a supply network made of direct and indirect suppliers. We can easily assume that a representative producer does not know the details of indirect transactions, further upstream, although they can be much relevant for the completion of her output. Therefore, she starts navigating her web of suppliers (e.g., making phone calls to her direct or indirect suppliers) to collect information on the quantity and quality of these transactions. In her *random walks*, she can encounter some resistance, for example, due to a reduced contractibility of some indirect inputs, which means that they are not quoted at any exchange nor are they referenced priced, possibly tailored for the specific need of their buyers. When a resistance is encountered, because information is difficult to collect, then it is more likely that the producer stops her exploration along that path, goes back to headquarters and starts following a different trail. However, at the end of each exploration, she can update her personal ranking, considering more relevant the inputs that are encountered more often and that are required more. The idea is very simple:

- A (direct or indirect) input that is also relatively more requested to produce other (direct or indirect) inputs must rank relatively higher;
- A (direct or indirect) input that is relatively more requested to produce other highly-requested (direct or indirect) inputs is relatively more relevant than a (direct or indirect) input that links with less-requested inputs.

[2] For a previous adaptation of a *Page Rank* centrality in the economics domain, see the *DebtRank* by Battiston et al. (2012), where connectivity among financial institutions and debt exposures are considered to determine the systemic importance of a node in a financial network.

In principle, our *Input Rank* can fit the analysis of complex webs of firm-to-firm transactions, as well as the study of more aggregate buyer-supplier linkages recorded in input-output tables. For sake of comparison with previous studies on the global organization of production, we compute our *Input Rank* exploiting U.S. input-output tables sourced from the Bureau of Economic Analysis (BEA). Thereafter, we test its empirical validity as a determinant of vertical integration choices in the fashion of Antràs and Chor (2013), Alfaro et al. (2017), and Del Prete and Rungi (2017), on a sample of 20,489 U.S. parent companies controlling 154,836 affiliates worldwide.

We find that a higher *Input Rank* is positively associated to higher odds that a (direct or indirect) input is vertically integrated within a firm boundary, relatively more when the demand faced by the *root* producer is more elastic. With a general reference to the contract theory of the firm, we argue that a choice of vertical integration is a way for the parent company to prevent that a central (direct or indirect) supplier reneges on her commitment, therefore endangering the functioning of the entire supply network. Our findings are robust to different sample compositions, to changes in parameters that measure the ability to outreach by a producer on the complex network structure, and to several empirical strategies. Interestingly, our findings are also robust to the inclusion of *downstreamness/upstreamness* metrics, which show some ambivalence in the case of *midstream* parents, in line with what previously found in Del Prete and Rungi (2017). Therefore, we discuss how the role of the elasticity of substitution is not clear when vertical integration is started by a producer of intermediate inputs.

The rest of the paper is organized as follows. The next section positions our contribution with respect to related literature. Section 3 introduces the *Input Rank* and its properties. In Section 4, we compute the *Input Rank* on U.S. input-output tables and describe some preliminary evidence. In Section 5, we test the role of the *Input Rank* in firm-level choices of vertical integration. Concluding remarks are offered in Section 5.

2 Related literature

A flourishing literature is emerging to study how the network dimensions in the organization of production can contribute to explain the response of aggregate fluctuations to microeconomic shocks (Acemoglu et al., 2012; Carvalho, 2014, Acemoglu et al., 2016). On the other hand, the shape of a production network is increasingly seen as the result of endogenous collective choices by buyers and suppliers, who establish reciprocal input linkages hence shaping both individual and aggregate productivities (Oberfield, 2018).

Yet, the fragmentation of global value chains (GVC) is still modelled and tested as on a linear sequence made of producers, who decide whether to *'make or buy'* an input, even though the existence of spider-like *vs* snake-like production has been acknowledged as depending on engineering details (Baldwin and Venables, 2013). To date, few works have considered the richness of buyer-supplier networks from an international perspective[3], and there is much to do for understanding the implications of network structures on the global organization of production. We make an effort to cover this gap starting with the introduction of the *Input Rank* as a measure of the technological relevance of an input, which takes into account the recursive nature of real-world webs of suppliers and buyers.

Realistically, we assume that each producer is plunged into a production network made of both direct and indirect suppliers, where inputs may be recursively used at different production stages. Therefore, we assume that a representative producer may have direct knowledge of the transactions in which she is a contracting party, but she has only limited knowledge of the transactions occurring among suppliers of suppliers. Nonetheless, what happens in the upstream technological and contractual space has consequences on her ability to deliver an output. Thus, she starts *random walks* (say, random phone calls) on her web of suppliers to acquire knowledge about indirect input transactions. We represent such a process as a Markov discrete chain in the spirit of the *Page Rank* problem, which originally ranks the consumption of information and regulates the workings of many social networks (Brin and Page, 1998). More specifically, we get inspired by the 'personalized' version of the *Page Rank* proposed by Haveliwala (2003) and White and Smyth (2003). In fact, the *Page Rank* has become a more general tool of the network theory and it is adapted to many diverse scientific domains (Gleich, 2015), from biology and bibliometrics, to neuroscience and engineering of road networks.

To show a practical empirical usage of the *Input Rank*, we test its role for the firm-level choices of vertical integration. Since the 1980s, several attempts have been made to model the '*make or buy*' decision based on the relative degree of contractibility between a buyer and a supplier[4]. Acemoglu et al. (2007) are the first to study a theoretical framework where unique headquarters commit to contracts with multiple suppliers. More recently, Harms et al. (2012) analyze the offshoring decision of firms whose production process is characterized by a sequence of steps and a non-monotonic variation of transportation costs. Costinot et al. (2013) derive a sequential multi-country model in which mistakes can occur with a given probability along a sequence. Hence, countries performing more knowledge-intensive tasks are better situated relatively more upstream, participating to a larger share in world income distribution. Interestingly, Fally and Hillberry (2015) include Coasian transaction costs to explain the length of a supply chain and the cross-country variation in gross output-to-value added ratios.

In each of the previous works, the notion of a sequence assumes different shades of meaning. More properly, we believe that Antràs and Chor (2013) and Alfaro et al. (2017) model a supply chain as a technological sequence made of production stages, from the start of a business

[3] For a first review of the first attempts made until now, see Bernard and Moxnes (2017).
[4] For a detailed review since the seminal work by Grossman and Hart (1986), see Aghion and Holden (2011). See also Antràs and Yeaple (2014) for a review on trade and firm-level organization of multinational enterprises.

line to the delivery to final demand, where each *downstream* output depends on a set of *upstream* (direct or indirect) inputs. In this framework, all producers shall rely on a surplus from the sale of the final output, and an economic and contractual dependence is established along the supply chain, for how that surplus is optimally generated by and allocated among producers. Eventually, the main prediction by the authors is that final-good producers integrate production stages that are relatively more *downstream* (*upstream*) when final demand is sufficiently elastic (inelastic).

In this respect, we believe that the latter strand of research has a potential to be extended to more complex production structures, while leaving the technological sequence, i.e., the chain, as a corner solution of real-world supply networks. In fact, more recently Antràs and de Gortari (2017) succeed in extending a supposedly linear technological sequence with the introduction of a notion of geographic centrality. In a multi-country setting, the authors predict an optimal location of a production stage to be dependent also on the geographical proximity to other stages. In that framework, *downstream* stages are preferably located in more *central* destinations, where centrality is to be interpreted in terms of geographic proximity.

When it comes to firm-level empirics, both Del Prete and Rungi (2017) and Alfaro et al. (2017) positively test the predictions by Antràs and Chor (2013), assuming as from theory that integration starts from the end of the supply chain. If final demand is sufficiently elastic (inelastic), producers of final goods integrate production stages that are more proximate to (far from) the consumers. However, in the case of *midstream* parents, when integration starts from the middle of the supply chains, Del Prete and Rungi (2017) find that the same theoretical predictions are no more statistically significant. In either case, integrated production tasks tend to be rather proximate to the parent on the *downstreamness* segment, possibly due to some local unexplored technological complementarities.

In our empirical application of the *Input Rank*, we build on the latter framework and test whether the *'make or buy'* decision can be driven by the technological centrality of an input in the specific production process. In fact, we find that a higher *Input Rank* is associated with higher odds that that input is vertically integrated, even more, when the elasticity of demand faced by the parent is more elastic. In this context, our findings seem to show that the network positioning of an input is at least as important as its distance from final consumption, the latter proxied by the *downstreamness/upstreamness* segments.

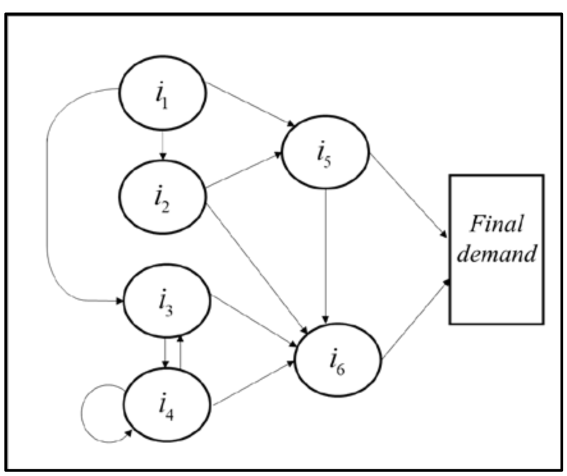
3 The Input Rank

3.1 A producer in a supply network

Our aim is to catch the central position of each input, i.e., its technological relevance for the completion of a network-like production process. We start by representing the problem of a producer who plans the delivery of her output based on the requirements of both direct and indirect inputs. In other words, a producer is aware that the completion of her production process requires the contribution of her direct transactions, for which she has immediate knowledge, and of indirect transactions among suppliers of suppliers, whose quantity and quality are not immediately known.

To clarify better the nature of the producer problem, let us consider a stylized production network sketched in Figure 3, where we represent an economy made of six producers and a final consumer. A producer i_6 transacts with four suppliers in the set $\{i_2, i_3, i_4, i_5\}$ to buy intermediate goods and services. Besides direct transactions, what happens more upstream is not directly disclosed to her. For example, there is a set of suppliers $\{i_2, i_3, i_5\}$ who directly rely on purchases from i_1. On the other hand, suppliers in $\{i_3, i_4\}$ reciprocally exchange part of their output to be used as intermediate inputs in their production process, whereas typically the supplier i_4 employs a share of her output to be reinvested in her production process as an intermediate input. Going further up in the production network, the supplier i_1 is also an indirect supplier of supplier i_4, through the production process made by i_3, who eventually is both a direct and an indirect supplier of our target producer i_6. In a nutshell, the stylized supply network of Figure 3 includes the main elements that make the production process recursive in nature.

Figure 3: A stylized production network made of six producers



Against this background, the producer problem reduces to a ranking of all direct and indirect suppliers considering their technological centrality in the supply network. What happens if one of them does not deliver? Who increases at most the risk of a disruption on the supply network?

Obviously, beyond the stylized network of Figure 3, real-life production processes can see the engagement of an incredibly high number of direct and indirect suppliers, active in many industries and in many countries. For this reason, the representative producer can find it difficult to collect information on *upstream* transactions, especially when global production is more and more fragmented. Realistically, we can imagine that the exploration process started by the producer can be proxied by a Markov discrete chain on paths of suppliers, but with heterogeneous abilities to outreach on the entire web of suppliers.

3.2 Random walks on a production network

Let us start by considering an economy in the form of an oriented graph:

$$\mathcal{G}(N, E, V, D) \tag{1}$$

made of a set of producers, $\forall\ i, j \in N$, connected by a set of direct input-output linkages, $e_{ij} \in E$. Each producer generates an amount of output, $v_i \in V$, that is distributed through linkages whose weights correspond to (normalized) input requirements, $d_{ij} \in D$, falling in a range $[0, 1]$. Each input requirement, d_{ij}, represents the amount of the *i*th direct input necessary to produce a unit of the *j*th output.

From a producer's perspective, her supply network is a sub-graph of the entire economy that can be navigated through *upstream* paths in the form:

$$P_{ir}^{-} = (r, i_l, \ldots, i_m) \tag{2}$$

where the *r*th node represents headquarters, i_l and i_m are any direct and indirect supplier, respectively. The negative sign on P_{ir}^{-} indicates the orientation of the path[5], from the *root*

[5] For sake of generality, we could also define a *downstream* relationship running from suppliers to buyers based on input deliveries, i.e., a supply side on the production network with positively oriented paths, P_{ir}^{+}, which tell us whether a supplier can reach a *root* buyer in a positive transitive closure $\mathcal{P}_r^{+}$. For a reference, see Gilles (2010).

producer *upstream*, where at each step there is the placement of a demand order for an input. The simpler path is, of course, one that runs from the *root* producer until a direct supplier. More sophisticated paths can run through production cycles, as in the stylized network of Figure 3.

In this context, let us define a *transitive closure*, $\mathcal{P}_r^-$, as a collection of all technological paths, such that $P_{ir}^- \in \mathcal{P}_r^-$.

In other words, an oriented transitive closure, $\mathcal{P}_r^-$, represents a mapping that tells us whether and how any supplier can be reached by a *root* producer, given the entire technological network, $\mathcal{G}$, that defines an economy. Please note how we can find more than one demand path running between a *root* buyer and any supplier[6], possibly given by the recursive nature of the production processes. Usefully, the introduction of a generic *root* buyer allows us changing the reference point in the supply network for how many producers we can find in the entire economy, hence spotting local properties that are specific for each target producer and her production process.

In other words, any (supply or demand) path that is left outside a (positive or negative) transitive closure oriented to the *r*th buyer does not participate to the production process of the latter. Fixing an eye on a different *root* buyer would imply the separation of a different sub-network from the entire network economy.

Given the previous framework, we can finally solve the problem of attributing a ranking to direct and indirect suppliers from the perspective of the *root* buyer. The idea is rather simple:

- an input is more technologically relevant to the *r*th producer if it is also relatively more requested to produce other (direct or indirect) inputs;

- an input is more technologically relevant to the *r*th producer if it is also relatively more requested to produce other highly-requested (direct or indirect) inputs.

At this point, we can define a measure of the *Input Rank* as a stochastic process[7] started by the *root* producer, who needs travelling *random walks* to collect information on the characteristics of all direct and indirect inputs needed to complete her production.

[6] For example, in the simple network described by Figure 3, a (positive or negative) transitive closure oriented on i_6 includes a total of 18 paths of variable length: four paths of length one, seven paths of length two, four paths of length 3 and three paths of length 4. Among others, the indirect supplier i_1 is connected to buyer i_6 through seven paths of variable length: (i_1, i_5, i_6), (i_1, i_2, i_6), (i_1, i_3, i_6), (i_1, i_2, i_5, i_6), (i_1, i_3, i_4, i_6), $(i_1, i_3, i_4, i_3, i_6)$, $(i_1, i_3, i_4, i_4, i_6)$. Please note how the presence of reciprocal supplies (i.e., cycles) and in-house production requires the multiple accounting of some suppliers on the same path.

[7] To construct the *Input Rank*, we get our inspiration from the *Page Rank* centrality introduced by Brin and Page (1998) to organize web pages based on their connectivity with the rest of the web. Nowadays, the Page Rank centrality is considered a useful tool from network theory and is used across different domains: bibliometrics, biology, physics, etc. (Gleich, 2015). For a previous economic application, see Battiston et al. (2012), who adopt a notion of *Debt Rank*

We assume that the *root* producer travels randomly, going from one supplier to another, e.g., calling them by phone and asking on characteristics of deliveries. At any step in the web of transactions, she has a probability α to proceed in the exploration and a probability $(1-\alpha)$ to fall back to headquarters. The parameter α proxies an information wedge between the buyer and the supplier that prevents full disclosure of the attributes of transactions.

At any time-step *t*, the *root* producer collects information on the direct requirement, d_{ij}, of each transaction, and she updates her information following a Markov process as follows:

$$\boldsymbol{\pi}_{rt} = \alpha \mathbf{D} \boldsymbol{\pi}_{rt-1} + (1-\alpha)\mathbf{h}_r \tag{3}$$

where $\boldsymbol{\pi}_{rt}$ and $\boldsymbol{\pi}_{rt-1}$ are column vectors including rankings of any *i*th input at time-steps t and $t-1$, respectively. The transition matrix $\mathbf{D}$ collects all (column-normalized) direct requirement coefficients, d_{ij}, of transactions in the economy. The vector $\mathbf{h}_r$ has all its elements equal to 0 except for the *r*th element, which is 1 for the selected *root* producer[8].

In our case, the vector $\mathbf{h}_r$ uniquely identifies a (negative) transitive closure, $\mathcal{P}_r^-$, hence avoiding that the buyer falls outside her technology when back from her *random walks*. More in general, the vector $\mathbf{h}_r$ excludes that the *root* producer lingers in other areas of the production network while assuring her a safe journey back to headquarters.

The probability α falls in a range $(0, 1)$. A value proximate to 0 implies that the producer encounters a higher difficulty to travel upstream in the production network. A value proximate to 1 implies that the producer encounters almost no resistance in the exploration of her network. We shall exclude that the parameter α is equal to either 0 or 1, because in either case no exploration is needed or started at all. In the next paragraph, we will discuss an attempt to better qualify this parameter from an economic point of view, and we will discuss how sensitive results are at changing thresholds of this parameter. More in general, a constant parameter α can be seen as an information wedge between a producer and her web of suppliers, such that inputs that are more closely related to the target output are also more easily reached by the *root* producer starting from the headquarters.

What we know from the Perron-Frobenius theorem is that a stationary distribution, $\boldsymbol{\pi}_r^*$, of the Markov process in (3) can be found, it is unique, and the sum of its single elements is such that $\sum_i \pi_{ir}{}^* = 1$, because the transition square matrix $\mathbf{D}$ is positive and column-stochastic, with all

for assessing financial systemic risk.

[8] See Appendix A for a comparison between our *Input Rank* on supply networks and the original *Page Rank* on social networks and web engines.

positive single elements, $d_{ij} \geq 0$. In the stationary status, the single element $\pi_{ir}^{\ *} \in [0,\ 1]$ indicates the final rank of any *i*th supplier in the supply network of the *root* producer.

To understand the workings of the underlying Markov process, let us consider the updated distance, $\boldsymbol{\pi}_{rk} - \boldsymbol{\pi}_r^*$, which is the distance the root producer is from the true *Input Rank* solution after the *k*th exploration. It satisfies the recursion:

$$\boldsymbol{\pi}_r^* - \boldsymbol{\pi}_{rk} = \alpha \mathbf{D}(\boldsymbol{\pi}_r^* - \boldsymbol{\pi}_{rk-1}) \tag{4}$$

Since matrix $\alpha\mathbf{D}$ has a spectral radius by construction smaller than 1, $\boldsymbol{\pi}_{rk}$ tends to $\boldsymbol{\pi}_r^*$ when the number of time-steps becomes large enough, $k \to \infty$.

Let us assume that we start approximating $\boldsymbol{\pi}_r^*$ with an initial value $\boldsymbol{\pi}_{r0} = \mathbf{h}_r$. That is, let us assume that a producer's exploration starts from scratch, with no information at all on any direct or indirect transaction, when leaving headquarters represented by the unitary vector $\mathbf{h}_r$, while going through the recursion of eq. (4) up to a sufficiently large number of time-steps t.

For smaller networks, the Jacobi or the Power iterative methods for the solution of a linear system of equations are sufficient (Gentle, 1998). For bigger networks, the convergence of eq. (4) could be difficult to obtain and some adaptive methods have been suggested (e.g. Kamvar et al. 2003), according to which single nodes whose centrality has converged are not considered in following iterations, in this way introducing a degree of approximation[9].

In the simpler network represented by a relatively small input-output table, as the one we will use from Section 4, the following solution of the linear production system can be derived:

$$\boldsymbol{\pi}_r^* = (1-\alpha)(\mathbf{I} - \alpha\mathbf{D})^{-1}\mathbf{h}_r \tag{6}$$

3.3 Introducing input-specific frictions

So far, the parameter α has been considered as an arbitrary constant that represents a general information wedge that the *root* producer encounters any time she must gather information on the characteristics of any transaction. From another perspective, this parameter discounts the ability

[9] For detailed references on the mathematical properties of Page Rank tools, see Langville and Meyer (2011), and Gleich (2015).

of a producer to reach on bigger and complex webs of suppliers, when the time is scarce to navigate through all faraway transactions. The inclusion of such a parameter is certainly useful along increasingly fragmented global supply networks when producers shall navigate through many industries and countries. Conversely, the generic parameter $(1-\alpha)$ brings back the *root* producer to headquarters, possibly to start navigation on another production path contained in the transitive closure, $P_{ir}^{-} \in \mathcal{P}_{r}^{-}$.

In this section, we introduce the possibility that this parameter is a proxy for an input-specific friction in a variant of the *Input Rank*.

In other words, we introduce a full vector $\boldsymbol{\alpha}$, whose single elements, α_i, is a (normalized) indicator of input contractibility (Rauch, 1999; Nunn, 2007; Nunn and Trefler, 2013), which assesses how much that input is referenced priced and/or exchanged on thick markets. In other words, a higher input contractibility implies that the *root* producer can more easily gather the information on that transaction, for example because there is some standard contract that has been signed between the parties. If a single transaction is relatively more contractible, the *root* producer proceeds more easily *upstream* to explore the supply network beyond that transaction.

In this case, the Markov process can be expressed as:

$$\boldsymbol{\pi}_{rt} = \boldsymbol{\alpha}\mathbf{D}\boldsymbol{\pi}_{rt-1} + (\mathbf{1}-\boldsymbol{\alpha})\mathbf{h}_r \quad (7)$$

where $\mathbf{1}$ is a vector with all elements equal to one, and $\boldsymbol{\alpha}$ is the normalized contractibility vector. Main mathematical properties are kept, as from Section 3.2, including convergence in bigger networks after a reasonably big number of time steps, following iterative methods. In the following analyses, we will employ both this variant and the one in Section 3.2 for an empirical validation of the *Input Rank* on a smaller network generated by input-output tables

4 An application to U.S. input-output tables

In principle, the *Input Rank* can be computed for any producer plunged into a supply network made of firm-to-firm transactions. In this contribution, we pick more aggregate transactions sourced from the U.S. 2002 Input-Output tables, compiled by the Bureau of Economic Analysis (BEA), which we consider as a good training case for several reasons.

First, U.S. BEA tables represent a reasonably detailed picture of a production networks established among 425 industries, in absence of actual firm-to-firm transactions. Second, the same

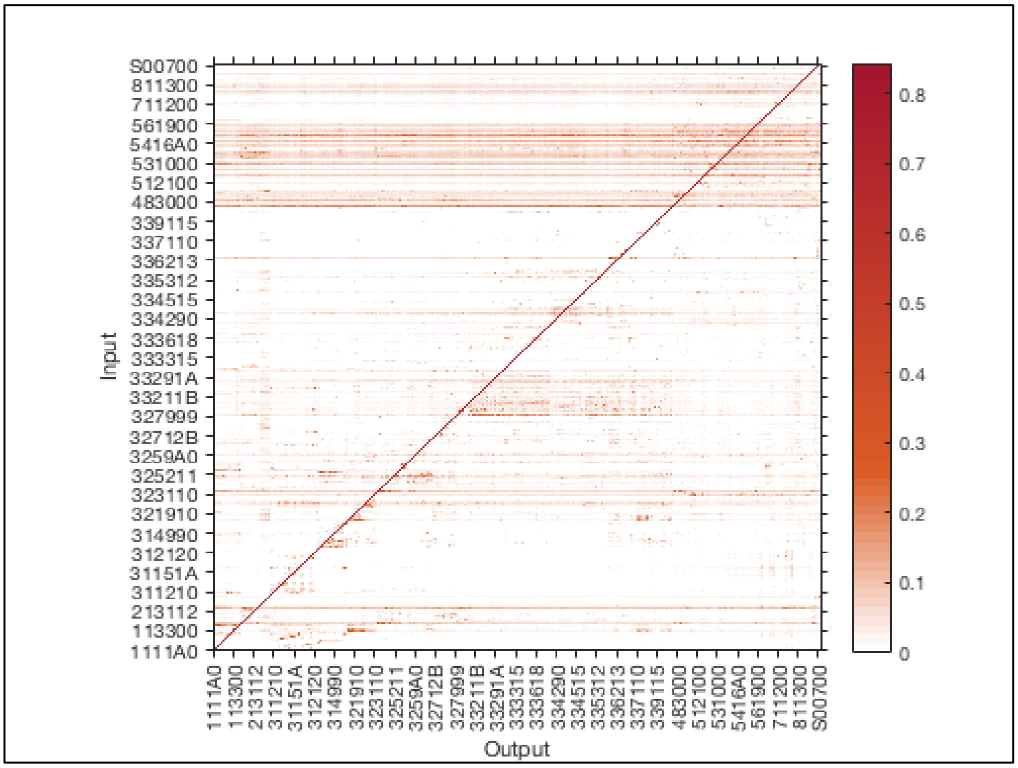
U.S. tables have been extensively used in previous works to study both production networks (Carvalho, 2014) and vertical integration choices (Acemoglu et al., 2009; Alfaro et al., 2016). Third, the same data have been recently used to compute *upstreamness* and *downstreamness* metrics[10], as a proxy for the technological distance on supposedly linear production sequences.

In fact, we already visualized in Figure 1 a solid and complex production network generated by the U.S. Input-Output tables, which collects 51,768 linkages. After a closer look at it, we can also register a strong heterogeneity of sourcing strategies developed among its 425 industries. In Appendix Figures B1 and B2, we report both the in-degree and out-degree distributions by industry, i.e., the number of inputs received, and the deliveries made by each node of the U.S. production network. As expected, the industry with the highest number of input industries (296) is the Retail Trade (code 4A0000), because retailers professionally sell physical goods to consumers. Interestingly, the industry with the highest number of purchasing industries (425) is the Wholesale Trade (code 420000), because wholesalers professionally distribute intermediate physical inputs to all industries. However, the 'global' centrality measured by out-degrees in Figure 1 is of scarce interest for our scope. More properly, the *Input Rank* introduced in Section 3 shall return a measure of 'local' eigenvector centrality, which better catches the technological relevance of an input with respect to any *root* producer.

In Figure 4, we visualize the results of the computation after following a power iteration method to derive the *Input Rank* as a vector of industry-pair values with a Matlab code that we cross-validate with a Python code, assuming a stochastic process described as in eq. (3). Each 6-digit industry is an input when on the y-axis, and it is an output when on the x-axis. A darker row implies that that industry is more technologically relevant across most industries. Interestingly, in the upper part of the figure, we find that services industries have a relatively more important role than manufacturing industries used as inputs across either manufacturing or services industries. Within manufacturing outputs, a crucial role is played by inputs coming from more aggregate Primary Metal Manufacturing (code 331) and Fabricated Metal Product Manufacturing (332). As expected, Mining industries (code 21) are technologically relevant for manufacturing producers.

[10] The 2002 U.S. Input-Output tables have been used for the computation of *absolute downstreamness* metrics in Antràs and Chor (2013), as an exercise based on previous *upstreamness* metrics proposed in Antràs et al. (2012). Alfaro et al. (2017) more recently proposed an alternative output-specific *relative upstreamness* computed on an older 1992 version of the U.S. Input-Output tables.

Figure 4: A visualization of the *Input Rank* computed on U.S. BEA rev. 2002 input-output tables



Note: *Input Rank* vectors are computed for each *root* output among 425 industries classified at the 6-digit in the U.S. BEA 2002 tables after using the power iteration method. Inputs on the y-axes and outputs on the x-axes by alphabetical order. A darker cell implies that input is more technologically relevant for that output.

In Tables 1 and 2, we report some moments of *Input Rank* distribution: first for all the top 20 inputs, then for the top 20 manufacturing inputs, excluding services. These are useful to look at some details of the input usage. Here, as well, services industries are on average ranked higher than manufacturing industries. The first highly ranked input is the Management of Companies and Enterprises (code 550000)[11], which unquestionably points to a general professional nature of the management of U.S companies. Some post-production services also rank relatively high, as expected, as in the case of Wholesale Trade (code 420000) and Advertising (code 541800). Immediately after, we spot Electric Power Generation (code 221100) and bank credit, as included in Monetary Authorities and Depository Credit Intermediation (code 52A000). In Appendix Table B1, we look at the rank of R&D input services (code 541700) and discover that the latter are more

[11] As from the original definition (BLS, 2018): This sector comprises: i) companies that hold financial activities (securities or other equity interests) in other companies for the purpose of a corporate control to influence management decisions; ii) companies that professionally administer, oversee, and manage other companies through strategic or organizational planning and decision making.

relevant to the General Federal Defense Government Services (code S00500) than to other life sciences industries (In-vitro Diagnostic Substance Manufacturing, code 325413; Biological Product Manufacturing, code 325413; Pharmaceutical Preparation Manufacturing, code 325412; Medicinal and Botanical Manufacturing, code 325411).

Table 1: Top 20 inputs (all industries) by *Input Rank* (alpha = 0.5), from U.S. BEA 2002 I-O tables

IO code	Input name	mean	p50	sd	min	max
550000	Management of companies and enterprises	0.0323	0.0306	0.0143	0.0068	0.0936
420000	Wholesale trade	0.0277	0.0279	0.0124	0.0030	0.0949
531000	Real estate	0.0235	0.0170	0.0174	0.0066	0.1215
541800	Advertising and related services	0.0145	0.0125	0.0078	0.0042	0.0606
221100	Electric power generation, transmission, and distribution	0.0116	0.0093	0.0079	0.0023	0.0749
52A000	Monetary authorities and depository credit intermediation	0.0115	0.0092	0.0072	0.0041	0.0589
517000	Telecommunications	0.0093	0.0073	0.0062	0.0032	0.0666
484000	Truck transportation	0.0090	0.0079	0.0065	0.0011	0.0785
331110	Iron and steel mills and ferroalloy manufacturing	0.0088	0.0022	0.0156	0.0003	0.1192
523000	Securities, commodity contracts, investments, and related activities	0.0084	0.0064	0.0142	0.0026	0.2471
324110	Petroleum refineries	0.0083	0.0045	0.0141	0.0017	0.1307
561300	Employment services	0.0078	0.0053	0.0062	0.0028	0.0382
211000	Oil and gas extraction	0.0072	0.0040	0.0144	0.0012	0.1975
541100	Legal services	0.0071	0.0064	0.0029	0.0030	0.0246
533000	Lessors of nonfinancial intangible assets	0.0070	0.0059	0.0052	0.0017	0.0770
541610	Management, scientific, and technical consulting services	0.0065	0.0049	0.0044	0.0018	0.0451
722000	Food services and drinking places	0.0061	0.0048	0.0040	0.0018	0.0250
230301	Nonresidential maintenance and repair	0.0054	0.0043	0.0056	0.0018	0.0790
522A00	Nondepository credit intermediation and related activities	0.0054	0.0041	0.0065	0.0022	0.1042

Table 2: Top 20 inputs (manufacturing only) by *Input Rank* (alpha = 0.5), from U.S. BEA 2002 I-O tables

IO code	Input name	mean	p50	sd	min	max
331110	Iron and steel mills and ferroalloy manufacturing	0.0088	0.0022	0.0156	0.0003	0.1192
324110	Petroleum refineries	0.0083	0.0045	0.0141	0.0017	0.1307
336300	Motor vehicle parts manufacturing	0.0052	0.0024	0.0143	0.0010	0.1686
325211	Plastics material and resin manufacturing	0.0052	0.0015	0.0139	0.0002	0.1584
325190	Other basic organic chemical manufacturing	0.0051	0.0019	0.0108	0.0003	0.0934
334413	Semiconductor and related device manufacturing	0.0041	0.0030	0.0061	0.0004	0.0792
322210	Paperboard container manufacturing	0.0039	0.0022	0.0051	0.0003	0.0418
32619A	Other plastics product manufacturing	0.0039	0.0020	0.0044	0.0005	0.0299
334418	Printed circuit assembly (electronic assembly) manufacturing	0.0035	0.0024	0.0047	0.0003	0.0400
321100	Sawmills and wood preservation	0.0030	0.0006	0.0109	0.0002	0.1318
323110	Printing	0.0030	0.0016	0.0057	0.0007	0.0704
322120	Paper mills	0.0028	0.0010	0.0086	0.0002	0.0863
326110	Plastics packaging materials and unlaminated film and sheet manufacturing	0.0027	0.0010	0.0045	0.0001	0.0380
332710	Machine shops	0.0026	0.0019	0.0025	0.0002	0.0143
3259A0	All other chemical product and preparation manufacturing	0.0023	0.0015	0.0026	0.0003	0.0207
322130	Paperboard mills	0.0021	0.0012	0.0051	0.0002	0.0627
33131A	Alumina refining and primary aluminum production	0.0020	0.0003	0.0084	0.0001	0.1146
332800	Coating, engraving, heat treating and allied activities	0.0019	0.0018	0.0015	0.0001	0.0081
325220	Artificial and synthetic fibers and filaments manufacturing	0.0019	0.0001	0.0105	0.0000	0.1271

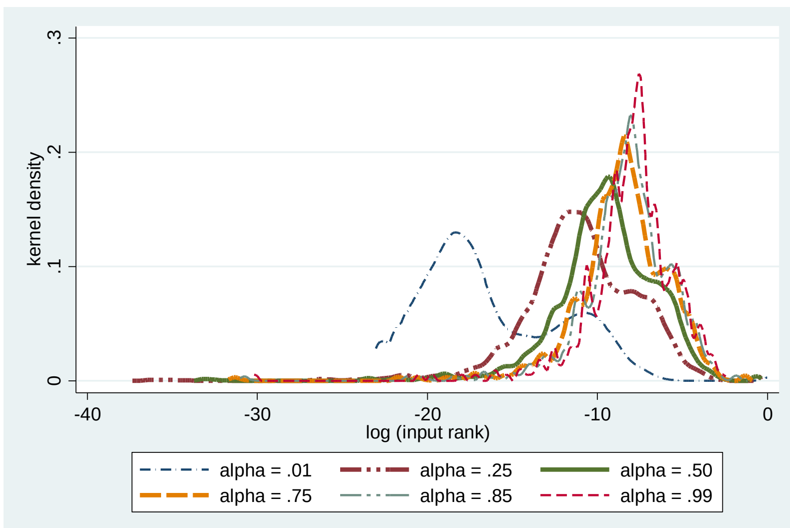
The first manufacturing input encountered among the top 20 is the Iron and Steel Mills and Ferroalloy Manufacturing (code 331110), which comes only after Truck Transportation (code 484000). In general, we observe a high variation of the *Input Rank* across *root* industries, as shown by relatively high values of the standard deviations.[12] In this case, it is more useful to look from the perspective of selected *root* industries (Electronic Computer Manufacturing, code 334111; Automobile Manufacturing, code 336111), in Appendix Tables B2 and B3 we find that the *Input Rank* is indeed specific to the technological nature of the production processes.

[12] Please note how in absence of actual firm-to-firm transactions, different aggregations of the input-output industries may alter the *Input Rank*. We expect that a higher aggregation of an input industry entails an overestimation of its *Input Rank* for any other *root* industry. In this case, we shall rely on official statistics offices that separate industries based on their technological relevance in modern economies, as is the scope of periodic updates of input-output tables.

4.1 The sensitivity to the parameter α

In general, the parameter α can be interpreted as the probability to proceed further in the exploration of a supply network, therefore its complement to one is the probability to stop exploration and fall back to headquarters at each time-step. In Section 3.3, we discussed how we can make use of this parameter to introduce an input-specific friction: the contractibility of an input. The latter would catch the thickness of the input market in the fashion of Rauch (1999), Nunn (2007), and Nunn and Trefler (2013). In this case, we could assume that a *root* producer can more easily gather information on the implicit characteristics of the transaction thanks, for example, to the existence of a reference price or the signature of a standard supply contract for that input.

Figure 5: Sensitivity of *Input Rank* to changing values of the parameter alpha



The alternative is to make reference to other notable constants previously used in the use of some *Page Rank* centralities. For example, Brin and Page (1998) originally suggest a damping factor $\alpha = 0.85$. As neutral as possible, $\alpha = 0.5$ implies an equal probability of proceeding through exploration or stopping at each time-step. Indeed, the latter is the reference value we use for main descriptive statistics and for baseline regressions in this text. However, we will make sure in the next paragraphs that our econometric results are robust to changing thresholds of the parameter α.

In Figure 5, we plot the changing shapes of the (log) distributions of the *Input Rank*, as computed on industry pairs from the U.S. 2002 Input-Output Tables, at different constant values of the parameter α. In Table 3, we report pairwise correlations among these distributions, including also the case when α_i is the (normalized) input specific contractibility. Finally, in Appendix Figure B3, we compare the latter case with a parameter $\alpha = 0.5$. In general, we find that shapes can be very similar, although shifting to the left at lower parameter values, because they discount relatively less faraway input industries in the supply network. In the end, the rankings of inputs are highly correlated at varying values of the parameter, at least before approaching the unit value.

Table 3: Pairwise correlations: *Input Rank* computed at several values of the alpha

Values of alpha	= .01	= .25	= .50	= .75	= .85	= .99	= input contractibility
= .01	1.00						
= .25	0.99	1.00					
= .50	0.98	0.99	1.00				
= .75	0.92	0.94	0.97	1.00			
= .85	0.80	0.83	0.88	0.97	1.00		
= .99	0.08	0.11	0.18	0.39	0.61	1.00	
= input contractibility	0.78	0.79	0.80	0.78	0.72	0.18	1.00

From Table 3, it is evident that when $\alpha = 0.99$ the distribution becomes very different. In fact, the stochastic process described in eq. (4) degenerates when $\alpha \rightarrow 1$. In Appendix Figure B3, we find that the distribution after the introduction of an input-specific friction, as in the case of input contractibility, resembles the case of $\alpha = 0.5$. When we perform our econometric exercises in the next sections, we will consider both $\alpha = 0.5$ and input-specific α_i as baseline metrics, while checking for the robustness of our findings at different threshold values of the parameter α. As we will see, the magnitudes can change considerably but statistical significance will not. Therefore, one can choose which value to assign to the parameter α as dependent on the nature of the frictions to proxy in the supply network.

5 The role of the *Input Rank* in choices of vertical integration

The decision to *make or buy* an input is an example of a situation when a producer needs gathering information on the technological relevance of direct and indirect inputs. In this Section, we test whether the *Input Rank* can play a role as a determinant for the decision to integrate a production stage within the firm boundary (i.e., vertical integration), as an alternative to signing supply contracts with independent firms (i.e., outsourcing). For our purpose, we will make use of a dataset of U.S. parent companies that have integrated at least one production stage over time. Our empirical strategy explicitly takes on the theoretical framework by Antràs and Chor (2013), while augmenting the estimates by Del Prete and Rungi (2017) with the inclusion of the *Input Rank*, therefore assuming that a *root* producer has an albeit reduced ability to collect information on her supply network.

5.1 A sample of U.S. parent companies

Our firm-level data are sourced from the Orbis database, compiled by the Bureau van Dijk, which gathers financial and ownership information for companies on a global scale. For our scope, we collect information on 20,489 U.S. parent companies controlling 154,836 subsidiaries in 210 countries at the end of the year 2015[13]. The selection of U.S. parent companies is coherent with the subsequent use of regressors that are based on U.S. input-output tables and trade data. In Table 4, we provide some descriptive statistics of the geographic coverage of the subsidiaries.

Both subsidiaries and parent companies can be active in any industry: manufacturing (28.86%), services (69%), primary (0.29%), and extractive (1.85%). About 81% of subsidiaries integrated by U.S. parents are domestic. Not surprisingly, U.S. parent companies are involved mainly in global supply networks that spread across other OECD economies, where 96% of their subsidiaries are located. The member States of the European Union host the largest number of foreign subsidiaries. Among them, Germany, the United Kingdom, and the Netherlands attract a significant share of U.S. foreign affiliates active in services industries. Not surprisingly, member States of NAFTA, Canada and Mexico, mainly host manufacturing of final and intermediate goods. However, a non-negligible share of subsidiaries is present in Asia, Africa and the Middle East.

To validate our sample, we compare with official 'Data on Activities of Multinational Enterprises' (BEA, 2018) and OECD Statistics on Measuring Globalization (OECD, 2018). In 2015, BEA (2018) reports 6,880 billion dollars of total sales by foreign affiliates and 12,628 billion dollars of total sales by parent companies. The U.S. multinational enterprises present in our sample account for 94% and 92% of the BEA (2018) values, respectively. The number of foreign affiliates in our sample corresponds to 88.6% on the total of U.S. foreign subsidiaries reported in OECD (2018), although the latter source only reports the latest value valid for the year 2014.

[13] To build our sample of parents and subsidiaries, we follow international standards for complex ownership structures (OECD, 2005; UNCTAD, 2009; UNCTAD, 2016), according to which the unit of observation is the control link between a parent company and each of its subsidiary after a concentration of voting rights is detected (> 50%). See Rungi et al. (2017). Similar data structures were used in Alviarez et al. (2016), Cravino and Levchenko (2017), and Del Prete and Rungi (2017).

Table 4: Sample geographic coverage by country of subsidiaries

Country of subsidiaries	**Final goods**		**Intermediates**		**Services**		**All industries**	
	N.	*%*	**N.**	*%*	**N.**	*%*	**N.**	*%*
United States	20,571	*16.3*	24,590	*19.5*	80,729	*64.1*	125,890	*100.0*
European Union	1,934	*11.5*	2,084	*12.3*	12,872	*76.2*	16,890	*100.0*
of which:								
Germany	273	*13.2*	306	*14.8*	1,494	*72.1*	2,073	*100.0*
France	171	*11.0*	213	*13.7*	1,167	*75.2*	1,551	*100.0*
United Kingdom	563	*11.4*	624	*12.7*	3,734	*75.9*	4,921	*100.0*
Italy	136	*19.4*	139	*19.8*	427	*60.8*	702	*100.0*
Netherlands	158	*6.8*	171	*7.3*	2,005	*85.9*	2,334	*100.0*
Canada	980	*30.4*	923	*28.6*	1,325	*41.1*	3,228	*100.0*
Russia	18	*11.7*	30	*19.5*	106	*68.8*	154	*100.0*
Asia	251	*15.0*	312	*18.7*	1,109	*66.3*	1,672	*100.0*
of which:								
Japan	87	*11.5*	76	*10.1*	592	*78.4*	755	*100.0*
China	92	*12.1*	66	*8.7*	605	*79.3*	763	*100.0*
India	122	*15.7*	149	*19.1*	508	*65.2*	779	*100.0*
Africa	67	*14.2*	93	*19.7*	313	*66.2*	473	*100.0*
Middle East	82	*18.2*	80	*17.8*	288	*64.0*	450	*100.0*
Latin America	221	*12.1*	395	*21.6*	1,210	*66.3*	1,826	*100.0*
of which:								
Argentina	24	*8.1*	70	*23.6*	203	*68.4*	297	*100.0*
Brazil	137	*14.6*	219	*23.3*	583	*62.1*	939	*100.0*
Mexico	98	*23.3*	154	*36.6*	169	*40.1*	421	*100.0*
Australia	123	*14.2*	157	*18.1*	586	*67.7*	866	*100.0*
Rest of the world	489	*16.5*	585	*19.7*	1,892	*63.8*	2,966	*100.0*
Total	**24,834**	***16.0***	**29,403**	***19.0***	**100,599**	***65.0***	**154,836**	***100.0***

Note: intermediate and final manufacturing categories based on industry affiliates and following the BEC rev. 4 classification provided by the UN Statistics Division.

For the scope of our analysis, we map industry affiliations of both parent companies and their subsidiaries, from the NAICS rev. 2012 classification into the 2002 U.S. BEA I-O Input-Output Tables. The match by industry affiliations allows us combining firm-level data with sector-level metrics, like the *Input Rank* we computed in Section 4 and the *upstreamness* segments sourced from Alfaro et al. (2017). In absence of actual data on firm-to-firm transactions, such a mapping onto input-output tables[14] allows us proxying the technological relevance of a (direct or indirect) input with reference to a *root* output, in the case of the *Input Rank*, and the relative technological distance of an input from the target output, in the case of the *upstreamness* segment. Finally, we complement our data with industry-level estimates of demand elasticity from Broda and Weinstein (2006), and with a measure of input contractibility retrieved from Antràs and Chor (2013) based on the methodology by Nunn (2007).

5.2 Baseline results

We test a conditional logit model with parent-level fixed effects[15]. The fixed effects conditional logit is a natural empirical strategy for the multinomial case of *ex-ante* alternatives. That is, we can test the determinants of vertical integration choices controlling for the characteristics of the production stages that were both vertically integrated and not integrated by the parent company.

Let $i = 1, 2, ..., N$ denote all the inputs, as from the input-output tables, and let $r = 1, 2, ..., R$ denote the *root* parent companies. The dependent variable, y_{ir}, takes on a value of 1 when at least one subsidiary has been integrated that produces the *i*th input, and 0 otherwise. Therefore, for each rth parent company, we have a vector $\mathbf{y}_r = (y_{1r}, ..., y_{Nr})$ made of 0s and 1s when each input has been integrated or not, respectively.

We want to consider the probability that a generic parent chooses a value of $\mathbf{y}_r$ conditional on $\sum_{i=1}^{N} y_{ir}$:

$$\Pr\left(\mathbf{y}_r \middle| \sum_{i=1}^{N} y_{ir}\right) = \frac{\exp\left(\sum_{i=1}^{N} y_{ir}\mathbf{x}_{ir}\boldsymbol{\beta}\right)}{\sum_{s_i \in S_i} \exp\left(\sum_{i=1}^{N} y_{ir}\mathbf{x}_{ir}\boldsymbol{\beta}\right)} \tag{8}$$

where the element s_{ir} of the vector $\mathbf{s}_i$ is equal to 1 when the *i*th input is integrated, and 0

[14] For similar mappings of firm-level data into input-output tables by industry affiliations, see Alfaro and Charlton (2009), Acemoglu et al. (2010), Alfaro et al. (2016), Rungi and Del Prete (2018).

[15] See McFadden (1974) and Chamberlain (1980) for more details. Present notation is borrowed from Hamerle and Ronning (1995), and Hosmer et al. (2013). See also Head et al. (1995) and Del Prete and Rungi (2017) for previous applications in international economics, the latter with reference to firm-level vertical integration choices.

otherwise.

For each input-parent pair, we can identify a vector of covariates, $\mathbf{x}_{ir}$, which includes: the *Input Rank* relative to the *i*th input with respect to the output of the *r*th parent company; the interaction term of the *Input Rank* with the binary variable *Complements* relative to the *i*th input; the *Input upstreamness* sourced from Alfaro et al. (2017) measuring the technological distance of the *i*th input to the *r*th target output; the interaction term of the *Input upstreamness* with *Complements*; the input-specific *Contractibility* derived as in Nunn (2007); the *Direct requirement* coefficient available from the U.S. input-output tables that ranges in $[0, 1]$. In this context, the variable *Complements* is equal to 1 when the elasticity of substitution of the parent industry is below the median $(\rho_r > \rho_{med})$, and 0 otherwise $(\rho_r < \rho_{med})$. Errors are clustered by parent companies and variables are standardized. Results from nested specifications are reported in Tables 5 and 6. In Table 5, we report findings when we assume there is an equal probability that a producer proceeds exploring at each time-step the supply network or she falls back to the headquarters. In Table 6, we consider that the probability to proceed with the exploration is dependent on the input-specific contractibility. In the latter case, we do not include the input contractibility as a separate variable in the specifications.

The coefficient of immediate interest to us is the one on the *Input Rank*, which indicates whether the odds of vertical integration change for a more central input in the supply network. It is positive and significant throughout our estimations.

In the first columns, we consider all parent companies whether active in a manufacturing or a service industry. We find that one additional standard deviation of the *Input Rank* is correlated with 1.35 higher odds of vertical integration. Please note that in further columns, when we introduce subsequent controls, the sample reduces to manufacturing parents only, because the elasticity of substitution by Broda and Weinstein (2006) is estimated on U.S. imports of manufacturing only. Our specification is complete in the fourth column of Table 5 and in the third column of Table 6, where a standard deviation increase of the *Input Rank* correlates with 1.16 and 1.08 higher odds of vertical integration, respectively.

Table 5: Baseline regressions I: parent-level fixed effects conditional logit

Dependent variable: Input is integrated ==1	(1)	(2)	(3)	(4)	(5)	(6)
Input Rank (alpha = 0.5)	0.318***	0.196***	0.197***	0.149***	0.189***	0.085***
	(0.001)	(0.003)	(0.003)	(0.004)	(0.005)	(0.006)
Input Rank * Complements				0.090***	0.014*	0.209***
				(0.005)	(0.008)	(0.008)
Input upstreamness		-0.595***	-0.583***	-0.782***	-0.443***	-1.166***
		(0.023)	(0.025)	(0.031)	(0.035)	(0.068)
Input upstreamness * Complements		-0.044*	-0.037	0.336***	-0.177***	0.955***
		(0.025)	(0.028)	(0.039)	(0.048)	(0.078)
Contractibility			-0.385***	-0.390***	-0.645***	-0.249***
			(0.018)	(0.017)	(0.032)	(0.025)
Direct requirement	0.063***	0.049***	0.025***	0.015***	0.010*	0.026***
	(0.004)	(0.003)	(0.003)	(0.005)	(0.006)	(0.004)
Observations	8,564,068	1,437,785	1,151,908	1,151,908	595,218	542,872
N. parent companies	20,294	4,203	4,084	4,084	2,110	1,925
Pseudo R-squared	0.409	0.215	0.250	0.257	0.203	0.342
Log pseudolikelihood	-96,831.2	-29,841.3	-22,779.5	-22,560.8	-12,739.4	-9,281.4
Clustered errors by parent	Yes	Yes	Yes	Yes	Yes	Yes
Activity of parent companies	All	Manu-facturing	Manu-facturing	Manu-facturing	Final goods	Intermediate goods

Errors clustered by parent in parentheses. Variables are standardizes. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Our findings are robust after the inclusion of the *Input Upstreamness*, which should proxy the relative technological distance from an input and its target output. In this case, more distant inputs are less likely integrated by the parent company. The central tenet of the theoretical framework by Antràs and Chor (2013) and Alfaro et al. (2017) is tested by the sign of the interaction term between the *Input Upstreamness* and *Complements*. According to these authors, when final demand is sufficiently elastic (inelastic), parents integrate production stages that are more proximate to (far from) final demand. This seems to be the case for producers of final goods (penultimate columns in Tables 5 and 6), although a sign reversal is observed in the case of *midstream* parents (last columns), i.e., when considering integration choices by producers of intermediate goods, in line with what tested by Del Prete and Rungi (2017).

In an effort to extend the role of the elasticity of substitution to the case of supply networks, we include a similar interaction term of the variable *Complements* with the *Input Rank*. In this case, when final demand is sufficiently elastic, we find that the odds are proportionally higher that a central input is integrated within the boundary of the firm.

Please note how, as expected, the *Direct Requirement* and the input-specific *Contractibility* have a positive and negative coefficient, respectively. In the first case, a higher value of the transaction (if any) is trivially correlated with higher odds of vertical integration. In the second case, a more contractible input is less likely integrated because the agreement between a producer and an independent supplier can be more easily enforced by law, and the incentives for vertical integration are lesser.

Table 6: Baseline regressions II, parent-level fixed effects conditional logit

Dependent variable: Input is integrated ==1	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Input Rank (alpha = contractibility)	0.307***	0.140***	0.074***	0.129***	0.016**
	(0.001)	(0.003)	(0.004)	(0.006)	(0.007)
Input Rank * Complements			0.118***	0.034***	0.181***
			(0.008)	(0.010)	(0.011)
Input upstreamness		-0.769***	-0.860***	-0.685***	-1.078***
		(0.026)	(0.028)	(0.029)	(0.063)
Input upstreamness * Complements		-0.074**	0.080**	-0.144***	0.399***
		(0.033)	(0.033)	(0.034)	(0.040)
Direct requirement	0.122***	0.076***	0.073***	0.084***	0.065***
	(0.004)	(0.002)	(0.003)	(0.005)	(0.003)
Observations	8,564,068	1,437,785	1,437,785	745,554	675,473
N. parent companies	20,294	4,203	4,203	2,179	1,975
Pseudo R-squared	0.369	0.166	0.172	0.133	0.219
Log pseudolikelihood	-103,307.9	-31,720.9	-31,485.0	-17,830.4	-13,396.9
Clustered errors by parent	Yes	Yes	Yes	Yes	Yes
Activity of parent companies	All	Manu-facturing	Manu-facturing	Final goods	Intermediate goods

Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

5.3 Robustness checks

Our main findings are robust to several checks of robustness. First, in Table 7, we check whether sample compositions can have an impact on the sign and significance of coefficients.

In the first column, we exclude cases of inputs coming from the same 2-digit industry of the parent companies. In the second column, we exclude services inputs because some of them could uniquely lead to previous results, as they are more central than manufacturing (see Figure 1) in most production processes. In the third column, we modify our indicator of *Complements*, explicitly considering the difference between the elasticities of the output and of the input $(\rho_r - \rho_i)$, which more specifically provides a reference point to understand how much elastic the demand of the *root* producer is. In the fourth column, we reduce our sample to the top 100 (direct) inputs of the parent output, as from I-O tables, to check whether the role of the *Input Rank* is exclusively driven by direct vs indirect inputs. In all these cases, when an input is more technologically relevant in the supply network, the odds are higher that the parent companies will *make* rather than *buy* the input from an independent supplier.

In Appendix Tables B4, B5, B6 and B7, we further control for: i) sample compositions when the *Input Rank* is built by considering the input-specific contractibility at each time-step of the network exploration (see Section 3.3); ii) sample compositions when we consider only *midstream* manufacturing parents; iii) changing values of the constant parameter α; iv) empirical specifications different from the fixed-effects conditional logit. All main findings are similar in sign and significance with baseline estimates, with the exception of a lack of statistical significance of the coefficient of the *Input Rank* in Table B4, when we exclude possibly horizontal strategies in the 2-digit industries of the parent company.

Table 7: Robustness on sample composition, parent-level fixed effects conditional logit

Dependent variable: Input is integrated ==1	No horizontal	Only manuf inputs	Input vs output elast	Top 100 inputs
Input Rank (alpha = 0.5)	0.749***	0.152***	0.171***	0.134***
	(0.151)	(0.004)	(0.004)	(0.005)
Input Rank * Complements	0.306***	0.091***	0.064***	0.118***
	(0.076)	(0.006)	(0.005)	(0.006)
Input upstreamness	-0.892***	-0.768***	-0.649***	-0.611***
	(0.035)	(0.030)	(0.026)	(0.051)
Input upstreamness * Complements	0.286***	0.313***	0.137***	0.748***
	(0.046)	(0.040)	(0.038)	(0.062)
Contractibility	-0.505***	-0.289	-0.413***	-0.453***
	(0.025)	(0.016)	(0.017)	(0.023)
Direct requirement	-0.028*	-0.008	0.028***	0.040***
	(0.015)	(0.005)	(0.003)	(0.003)
Observations	741,066	905,640	1,151,908	156,705
N. parent companies	2,637	3,903	4,084	2,847
Pseudo R-squared	0.080	0.281	0.254	0.398
Log pseudolikelihood	-19,399.7	-18,622.2	-22,653.3	-7,949.5
Clustered errors by parent	Yes	Yes	Yes	Yes
Activity of parent companies	Manu-facturing	Manu-facturing	Manu-facturing	Manu-facturing

Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

6 Conclusions

In this contribution, we introduced the *Input Rank* as an eigenvector centrality measure applicable to recursive production networks made of transactions among many buyers and suppliers. It measures the technological relevance of suppliers in the entire supply network of a *root* producer. Adapted from previous metrics of information consumption in social networks and web engines, the *Input Rank* proxies a stochastic process that is started by a *root* producer, who needs gathering information on the technology of her entire supply network made of both direct and indirect suppliers. After random walks throughout her supply network (e.g., random phone calls), she comes at each time-step with a numerical value and an updated ranking of how important is that (direct or indirect) input for the completion of her production process. Given increasingly complex webs of suppliers and an increasing fragmentation of the production processes, it is possible that a generic *root* producer is scarcely able to navigate faraway areas of her supply network, therefore a dumping parameter is introduced at each time-step to discount the difficulty to gather information on a specific transaction, e.g., due to its contractibility, and therefore the probability that the *root* producer falls back home and starts a new journey of exploration.

The *Input Rank* can be computed either on firm-to-firm transactions or on input-output tables, keeping its simple mathematical properties. For sake of comparison with previous positioning metrics (e.g., *downstreamness* and *upstreamness* segments) of GVCs, we compute it on U.S. 2002 BEA Input-Output tables, and thereafter we test firm-level choices of vertical integration by U.S. parent companies. We find that a higher *Input Rank* correlates with higher odds that that input is vertically integrated, even more so when the demand faced by the parent company is more elastic. We argue that vertical integration allows reducing the possibility that otherwise independent suppliers renege on commitments and disrupt the supply network, generating more damage for the completion of the production processes. Even more so when the margins on which the *root* producer can rely are smaller. Our findings are robust to several checks on sample compositions, parameter choices and empirical models.

More in general, we argue that the *Input Rank* better catches the recursive and complex nature of real-world supply networks, which have been so far represented as supposedly linear technological sequences in studies for the international organization of production. Certainly, both empirics and theory need better considering the technological loops, kinks and corners, which can magnify or dampen a shock in a supply network, finally shaping the organizational response of the company.

References

Acemoglu D., P. Antràs, and E. Helpman (2007). "Contracts and Technology Adoption", The American

Economic Review, 97(3): 916–943.

Acemoglu D., S. Johnson, and T. Mitton (2009). "Determinants of Vertical Integration: Financial Development and Contracting Costs", The Journal of Finance, 64(3):1251–1290.

Acemoglu D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). "The Network Origins of Aggregate Fluctuations", Econometrica, 80(5):1977–2016, 2012.

Acemoglu D., A. Malekain and A. Ozdaglar (2016). "Network Security and Contagion", Journal of Economic Theory, pp. 536–585.

Aghion P. and R. Holden (2011). "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.

Alfaro L., P. Antràs, D. Chor, and P. Conconi (2017). "Internalizing global value chains: A firm-level analysis". NBER Working Paper No. 21582.

Alfaro L., P. Conconi, H. Fadinger, and A. F. Newman (2016). "Do Prices Determine Vertical Integration?" Review of Economic Studies 83, no. 3: 855–888.

Alviarez V., J. Cravino, and A. Levchenko (2017). "The Growth of Multinational Firms in the Great Recession," Journal of Monetary Economics, 85: 50-64.

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

Antràs P. and D. Chor (2017). "On the Measurement of Upstreamness and Downstreamness in Global Value Chains". NBER Working Paper No. 24185.

Antràs P. and A. de Gortari (2017). "On the Geography of Global Value Chains. NBER Working Paper No. 23456.

Antràs P and S. R. Yeaple (2014). "Multinational Firms and the Structure of International Trade", Handbook of International Economics, Volume 4, Pages 55–130.

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

Atalay E., A. Hortacsu, J. Roberts, and C. Syverson (2014) "A.Network Structure of Production," Proceedings of the National Academy of Sciences, 108 (13): 5199-5202.

Baldwin R. (2016). The Great Convergence: Information Technology and the New Globalization. Cambridge: Belknap Press of Harvard University Press.

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

Battiston S., M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli (2012). "DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk", Scientific Reports 2, 541.

Bernard A. B. and A. Moxnes (2017). “Networks and Trade,” forthcoming on Annual Review of Economics.

Bernard A. B., A. Moxnes, and Y. Saito (2017). “Production Networks, Geography and Firm Performance,” forthcoming on Journal of Political Economy.

Brin S. and L. Page (1998). Reprint of: “The anatomy of a large-scale hypertextual web search engine”, Computer Networks and ISDN Systems, 30:107-117.

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

BLS (2018). Industries at a Glance. Bureau of Labor Statistics.

Carvalho V. M. (2014). “From Micro to Macro via Production Networks”, The Journal of Economic Perspectives, 28(4): 23–47.

Chamberlain, G. (1980). “Analysis of covariance with qualitative data”, Review of Economic Studies 47: 25–238.

Costinot A., J. Vogel, and S. Wang (2012). “An Elementary Theory of Global Supply Chains”, Review of Economic Studies, 80(1): 109–144.

Cravino J. and A. Levchenko (2017). "Multinational Firms and International Business Cycle Transmission," The Quarterly Journal of Economics, 132(2): 921-962.

Del Prete D. and A. Rungi (2017). “Organizing the Global Value Chain: a firm-level test”, Journal of International Economics, 109:16–30.

Fally T. (2012). “Production staging: measurement and facts”, mimeo.

Fally T. and R. Hillberry (2015). “A Coasian Model of International Production Chains”, NBER Working Paper No. 21520.

Fruchterman, T. M. J. and E. M. Reingold (1991), "Graph Drawing by Force-Directed Placement", Software – Practice & Experience, 21 (11): 1129–1164.

Gentle J. E. (1998). Numerical Linear Algebra for Applications in Statistics. Springer Publishing.

Gilles, R. P. (2010). The Cooperative Game Theory of Networks and Hierarchies. Springer.

Gleich D. F. (2015). “PageRank Beyond the Web”, Society for Industrial and Applied Mathematics – SIAM Rev., 57(3), 321–363.

Grossman J. S. and O. D. Hart (1986). “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration”, Journal of Political Economy, 94(4):691–719, 1986.

Hamerle, A., and G. Ronning (1995). “Panel Analysis for Qualitative Variables”, Handbook of Statistical Modeling for the Social and Behavioral Sciences, ed. G. Arminger, C. C. Clogg, and M. E. Sobel, 401–451. New York: Plenum.

Harms P., O. Lorz, and D. Urban (2012). “Offshoring along the Production Chain”, Canadian Journal of Economics/Revue Canadienne d’économique, 45(1): 93–106, 2012.

Haveliwala T. H. (2003). “Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search”, IEEE Transactions on Knowledge and Data Engineering archive, 15(4): 784-796.

Head H., J. Ries, and D. Swenson (1995). “Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States”, Journal of International Economics, 38(3-4): 223–247.

Hosmer, D. W., Jr., S. A. Lemeshow, and R. X. Sturdivant (2013). Applied Logistic Regression. 3rd ed. Hoboken, NJ: Wiley.

Kamvar S., Haveliwala T., Manning C. and Golub G. (2003). Exploiting the block structure of the web for computing PageRank. Technical Report 2003-17, Stanford InfoLab

Katz L. (1953). “A new Status Index derived from Sociometric Analysis,” Psychometrika, 18(1):39–43.

Langville A. N. and Meyer C. D. (2011). Google's PageRank and Beyond: The Science of Search Engine Rankings. Princeton University Press.

McFadden D. (1974). “Conditional Logit Analysis of Qualitative Choice Behavior”. Frontiers in econometrics. Academic Press New York.

Miller, R. E., and U. Temurshoev, (2017), “Output Upstreamness and Input Downstreamness of Industries/Countries in World Production,” International Regional Science Review 40(5): 443-475.

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

Nunn N. and D. Trefler (2013). “Incomplete Contracts and the Boundaries of the Multinational Firm”, Journal of Economic Behavior & Organization 94, 330–344.

OECD (2005). Guidelines for Multinational Enterprises. OECD, Paris.

Oberfield E. (2018). “A Theory of Input-Output Architecture”, Econometrica, 86 (2): 559–589.

Rauch J. E. (1999). “Networks versus Markets in International Trade”, Journal of International Economics, 48(1): 7–35.

Rungi, A., G. Morrison, and F. Pammolli (2017). "Corporate boundaries of Multinational Enterprises: a network approach". EIC working paper series 07/2017. IMT Institute for Advanced Studies Lucca.

UNCTAD (2009). "Training Manual on Statistics for FDI and the Operations of TNCs Vol. II.". United Nations, Geneva.

UNCTAD (2016). World investment report 2016. United Nations, Geneva.

Wang, Z., S.-J. Wei, X. Yu, and Zhu K. (2017), “Measures of Participation in Global Value Chains and Global Business Cycles,” NBER Working Paper N. 23222

White S. and P. Smyth (2003). “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.

A – Appendix: from the *Page Rank* to the *Input Rank*

The intuition of the *Input Rank* is adapted from the 'personalized' version of the *Page Rank* centrality, first used in social networks and search engines (Brin and Page, 1998) to present to users the most pertinent content. Some variants of the *Page Rank* have been used in many domains (bibliometrics, biology, physics, engineering of infrastructures, financial exposure, etc.) as an alternative to the Katz (1953) centrality (Gleich, 2015). The underlying assumption is that more important nodes (in our case, *inputs*) are likely to receive more links from other nodes (in our case, *inputs of inputs*), and that proximity to central nodes implies, in turn, a relatively higher centrality.

For our scope, the main limitation of the original formulation of the *Page Rank* is its 'global' outreach on a supposedly unique network, whereas we are interested in a 'local' outreach of a specific *root* buyer in her oriented supply network. Therefore, we needed a 'personalization' of the *Page Rank*, in the spirit of Haveliwala (2003) and White and Smyth (2003), where different rankings are possible for different *root* nodes, given an initial prior knowledge of the stochastic process.

Starting from the original formulation of the *PageRank*, adopting the notation proposed by Gleich (2015), the eigenvalue problem can be represented by the following identity:

$$\left[(1-\alpha)\mathbf{P}+\alpha\mathbf{v}\mathbf{e}^{T}\right]\mathbf{x}=\mathbf{x} \tag{A1}$$

For our scope, we substitute each term with a corresponding in our *Input Rank* from eq. (5), to take into account the peculiar economic process at stake:

- In the *Page Rank*, a transition matrix $\mathbf{P}$ contains the probabilities that an internet user clicks on one page following a web link present on the one she is visiting, column-normalized by the total number of received links, i.e. its in-degree. In the *Input Rank*, we substitute the matrix $\mathbf{P}$ with an input-output matrix $\mathbf{D}$, whose single elements are column-normalized buyer-supplier transactions, d_{ij}. Even though in the present work we propose estimates of the *Input Rank* using input-output tables (US BEA, 2002), our framework is open to applications to firm-to-firm transactions. In the latter case, an element of the matrix $\mathbf{D}$ would be a firm-to-firm actual shipment, always normalized by column (i.e., expressed by receiving firm as a percentage of all input shipments).

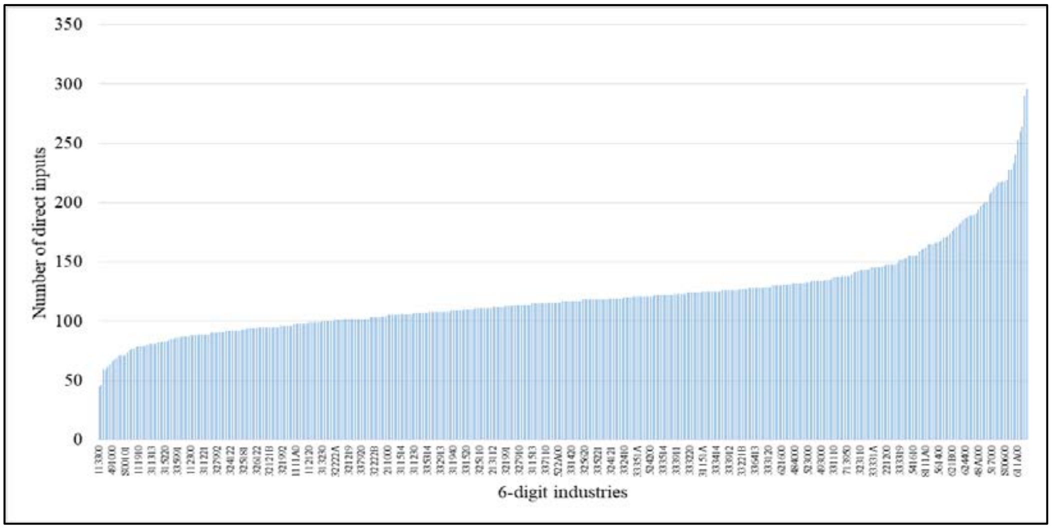
- A vector $\mathbf{v}$ is a critical tool that allows for the 'personalization' of the *Page Rank*. In absence of 'personalization', this vector contains just a uniform distribution of probability across all web pages. Therefore, personalized versions of the *Page Rank* make it non-uniform, so that a particular region of the internet is highlighted with a higher probability. At the same time, the vector $\mathbf{e}$ is a vector of 1s that algebraically extends the same (uniform or non-uniform) distribution in $\mathbf{v}$ to all web users. In our *Input Rank*, we substitute $\mathbf{v}$ with a *root*-specific unitary vector, $\mathbf{h}_r$ made of all 0s except for the *r*th element that is equal to 1. Together with the term α, the unitary vector $\mathbf{h}_r$ avoids that the *r*th buyer navigates outside her network of suppliers while pointing at headquarters.

- The term $\alpha \in (0,\ 1)$ is a *teleportation* parameter in the *Page Rank*, otherwise called a damping factor. It indicates the probability that a 'web surfer' interrupts a random navigation following page-to-page links and falls elsewhere, on any other web page not directly linked to the one she is visiting. By converse, $(1-\alpha)$ is the probability that the user goes on randomly following her web path made of cross-link citations. In our *Input Rank*, α must be read in connection to the peculiarity of $\mathbf{h}_r$. In our case, α is the probability that the *root* producer stops travelling in her production network and goes back to the headquarters (i.e., the 1 in the unitary vector $\mathbf{h}_r$). Conversely, $(1-\alpha)$ is the probability that the producer goes on exploring her web of suppliers. Later on, in Section 3, we further personalize introducing input-specific α_i, which considers the contractual friction that prevents a producer to collect, for example, the input degree of contractibility (Rauch, 1999; Nunn, 2007; Nunn and Trefler, 2013).

- Finally, $\mathbf{x}$ is the solution to the eigenvalue problem in (A1), which indicates the relevance of the web content in the case of the *Page Rank*. In our *Input Rank*, the *root*-specific solution $\boldsymbol{\pi}^*_r$ in eq. (5) represents the technological relevance resultant from the exploration of the input-output linkages by the *root* producer.

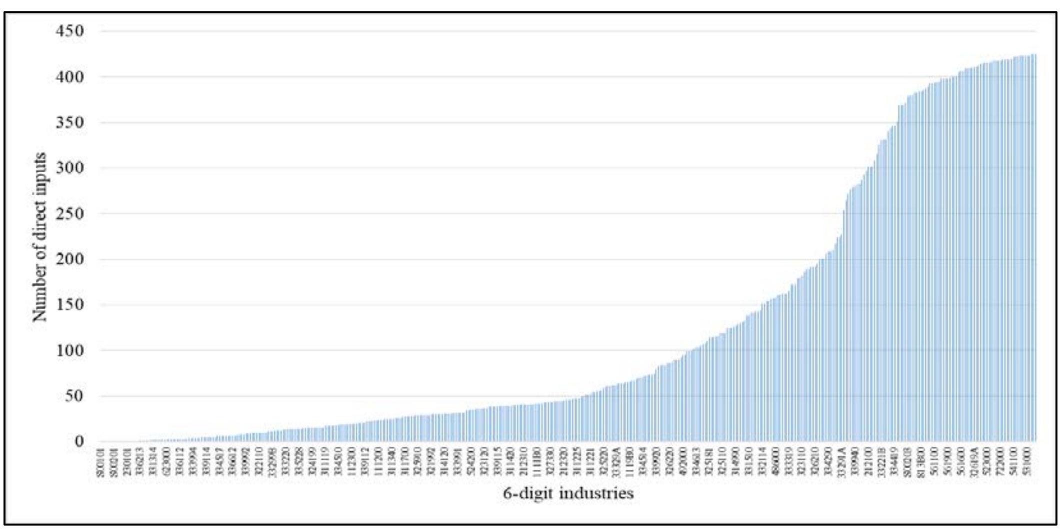
B – Appendix Tables and Graphs

Figure B1: In-degree distribution of Input-Output Network from U.S. BEA 2002 I-O tables



Note: Number of input industries by output ordered on the x-axis. Average: 122. Minimum at the Logging industry (code 113300) is 45. Maximum at the Retail Trade (code 4A0000) is 296.

Figure B2: Out-degree distribution of Input-Output Network from U.S. BEA 2002 I-O tables



Note: Number of buying industries by output ordered on the x-axis. Average: 122. Minimum at the Museums, Historical Sites, Zoos, and Parks (code 712000) is 0. Maximum at the Wholesale Trade (code 420000) is 425.

Table B1: Top 10 highest *Input Rank* values of the R&D services (code 541700) by output

IO code	Output name	R&D Input rank (alpha =0.5)
S00500	General Federal defense government services	0.0384
325413	In-vitro diagnostic substance manufacturing	0.0317
325414	Biological product (except diagnostic) manufacturing	0.0293
325412	Pharmaceutical preparation manufacturing	0.0247
325411	Medicinal and botanical manufacturing	0.0226
325320	Pesticide and other agricultural chemical manufacturing	0.0211
3259A0	All other chemical product and preparation manufacturing	0.0211
325620	Toilet preparation manufacturing	0.0193
325910	Printing ink manufacturing	0.0192
325610	Soap and cleaning compound manufacturing	0.0190

Table B2: Top 10 direct or indirect inputs by Input Rank for the Automotive Manufacturing (code 336111)

IO code	Input name	Input rank (alpha = 0.5)
336300	Motor vehicle parts manufacturing	0.1686
420000	Wholesale trade	0.0353
550000	Management of companies and enterprises	0.0302
331110	Iron and steel mills and ferroalloy manufacturing	0.0101
531000	Real estate	0.0087
541800	Advertising and related services	0.0078
334413	Semiconductor and related device manufacturing	0.0072
484000	Truck transportation	0.0071
32619A	Other plastics product manufacturing	0.0057
221100	Electric power generation, transmission, and distribution	0.0054

Table B3: Top 10 direct or indirect inputs by Input Rank for the Electronic Computer Manufacturing (code 334111)

IO code	Industry name	Input rank (alpha = 0.5)
334112	Computer storage device manufacturing	0.0568
420000	Wholesale trade	0.0553
550000	Management of companies and enterprises	0.0467
334418	Printed circuit assembly (electronic assembly) manufacturing	0.0400
334413	Semiconductor and related device manufacturing	0.0374
511200	Software publishers	0.0305
33411A	Computer terminals and other computer peripheral equipment manufacturing	0.0190
541800	Advertising and related services	0.0132
531000	Real estate	0.0121
541700	Scientific research and development services	0.0112

Figure B3: Distributions of the (logs of) *Input Rank* when alpha is 0.5 and alpha is input-specific contractibility

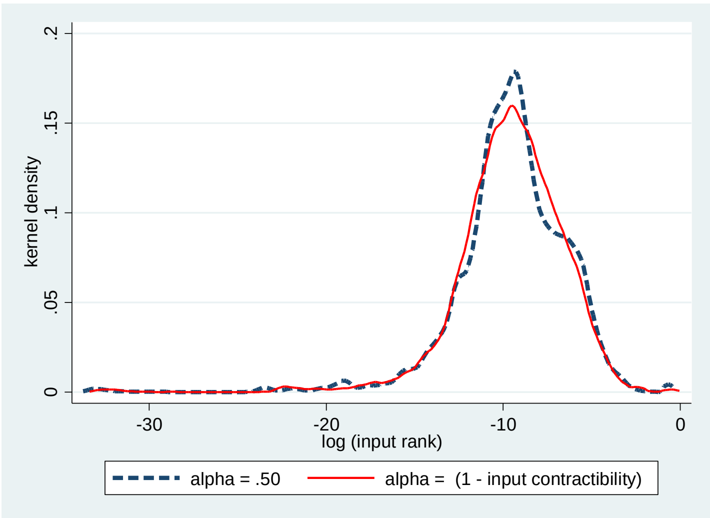


Table B4: Robustness to sample composition when alpha is input contractibility, parent-level fixed effects conditional logit

Dependent variable: **Input is integrated ==1**	No horizontal	Only manuf inputs	Input vs output elast	Top 100 inputs
Input Rank (alpha = contractibility)	0.614***	0.067***	0.114***	0.080***
	(0.050)	(0.003)	(0.004)	(0.005)
Input Rank * Complements	0.808***	0.126***	0.109***	0.136***
	(0.060)	(0.009)	(0.007)	(0.014)
Input upstreamness	-0.842***	-0.889***	-0.882***	-0.695***
	(0.032)	(0.028)	(0.026)	(0.046)
Input upstreamness * Complements	0.220***	0.025	0.091**	0.352***
	(0.039)	(0.037)	(0.039)	(0.054)
Direct requirement	0.017***	0.039***	0.064***	0.080***
	(0.005)	(0.003)	(0.003)	(0.003)
Observations	953,458	925,155	1,151,908	252,897
N. parent companies	2,796	3,903	4,084	3,020
Pseudo R-squared	0.094	0.202	0.210	0.201
Log pseudolikelihood	-25,884.9	-20,793.8	-23,986.9	-14,925.6
Clustered errors by parent	Yes	Yes	Yes	Yes
Activity of parent companies	Manu-facturing	Manu-facturing	Manu-facturing	Manu-facturing

Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table B5: Robustness to sample composition considering *midstream* parents only, parent-level fixed effects conditional logit

Dependent variable: Input is integrated ==1	No horizontal	Only manuf inputs	Input vs output elast	Top 100 inputs
Input Rank (alpha = 0.5)	0.097	0.080***	0.173***	0.049***
	(0.177)	(0.006)	(0.005)	(0.007)
Input Rank * Complements	0.931***	0.223***	0.090***	0.234***
	(0.080	(0.008)	(0.008)	(0.009)
Input upstreamness	-1.254***	-1.197***	-0.794***	-1.618***
	(0.073)	(0.065)	(0.047)	(0.116)
Input upstreamness * Complements	0.793***	1.112***	0.385***	1.904***
	(0.084)	(0.078)	(0.062)	(0.132)
Contractibility	-0.191***	-0.163***	-0.271***	-0.276***
	(0.030)	(0.024)	(0.025)	(0.029)
Direct requirement	-0.016	0.012***	0.033***	0.011**
	(0.017)	(0.004)	(0.003)	(0.005)
Observations	316,429	437,805	542,872	87,847
N. parent companies	1,126	1,887	1,925	1,591
Pseudo R-squared	0.096	0.363	0.313	0.460
Log pseudolikelihood	-8,141.9	-8,063.6	-9,688.9	-3,848.8
Clustered errors by parent	Yes	Yes	Yes	Yes
Activity of parent companies	Intermediate goods	Intermediate goods	Intermediate goods	Intermediate goods

Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table B6: Robustness to changing values of the parameter alpha, parent-level fixed effects conditional logit

Dependent variable: Input is integrated ==1	alpha = 0.01	alpha = 0.25	alpha = 0.50	alpha = 0.75	alpha = 0.85	alpha = 0.99
Input Rank	0.139***	0.142***	0.149***	0.167***	0.198***	0.209***
	(0.004)	(0.004)	(0.004)	(0.005)	(.006)	(0.013)
Input Rank * Complements	0.089***	0.089***	0.090***	0.097***	0.115***	0.128***
	(0.005)	(0.005)	(0.005)	(0.006)	(.007)	(0.015)
Input upstreamness	-0.779***	-0.781***	-0.782***	-0.780***	-0.777***	-0.898***
	(0.030)	(0.031)	(0.031)	(0.031)	(.032)	(0.030)
Input upstreamness * Complements	0.332***	0.334***	0.336***	0.342***	0.350***	-0.073*
	(0.038)	(0.039)	(0.039)	(0.039)	(.017)	(0.039)
Contractibility	-0.388***	-0.388***	-0.390***	-0.393***	-0.395***	-0.410***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)
Direct requirement	0.036***	0.026***	0.015***	0.001	-0.007	0.070***
	(0.004)	(0.004)	(0.005)	(0.005)	(.005)	(0.002)
Observations	1,151,908	1,151,908	1,151,908	1,151,908	1,151,908	1,151,908
N. parent companies	4,084	4,084	4,084	4,084	4,084	4,084
Pseudo R-squared	0.257	0.257	0.257	0.257	0.257	0.147
Log pseudolikelihood	-22,569.0	-22,564.7	-22,560.8	-22,553.1	-22,548.5	-25,905.4
Clustered errors by parent	Yes	Yes	Yes	Yes	Yes	Yes
Activity of parent companies	Manu-facturing	Manu-facturing	Manu-facturing	Manu-facturing	Manu-facturing	Manu-facturing

Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table B7: Robustness to changing empirical strategy

Dependent variable: Input is integrated ==1	Linear probability model	Logit	Probit
Input Rank (alpha = 0.5)	0.011***	0.151***	0.074***
	(0.001)	(0.005)	(0.002)
Input Rank * Complements	0.024***	0.242***	0.117***
	(0.001)	(0.004)	(0.002)
Input upstreamness	-0.003***	-0.766***	-0.273***
	(0.001)	(0.025)	(0.009)
Input upstreamness * Complements	-0.001***	-0.415***	-0.160***
	(0.001)	(0.025)	(0.009)
Contractibility	-0.001***	-0.354***	-0.141***
	(0.001)	(0.017)	(0.007)
Direct requirement	-0.001***	0.018***	0.011***
	(0.001)	(0.004)	(0.001)
Constant	0.005***	-5.894***	-2.789***
	(0.001)	(0.033)	(0.012)
Observations	1,257,668	1,257,668	1,257,668
N. parent companies	4,717	4,717	4,717
R squared / Pseudo	0.124	0.209	0.215
Log pseudolikelihood	-	-29,320.3	-29,098.1
Clustered errors by parent	Yes	Yes	Yes
Activity of parent companies	Manufacturing	Manufacturing	Manufacturing

Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.