Measuring the Input Rank in Global Supply Networks

Measuring the Input Rank in Global Supply Networks

Armando Rungi[*] Loredana Fattorini[†] Kenan Huremovic[§]

This version: October 2019

Abstract

In this paper, we introduce the *Input Rank* as a measure to study the organization of global supply networks at the firm level. We model the case of a firm that needs assessing the technological relevance of each direct and indirect supplier on a network-like production function with labor and intermediate inputs. In our framework, an input is technologically more relevant if a shock on that upstream market can hit harder the marginal costs of a downstream buyer, considering the topology of the supply structure. A higher labor intensity at each stage buffers the transmission of upstream shocks in the network. In addition, we provide for the possibility that producers have limited knowledge of inputs in the supply network, hence they can underestimate the relevance of more distant inputs. After applications, the *Input Rank* returns a matrix of technological centralities that order any direct or indirect input for a representative firm in any output industry. We compute the *Input Rank* on U.S. and world input-output tables. Finally, we test how it correlates with choices of vertical integration made by 20,489 U.S. parent companies controlling 154,836 affiliates worldwide. We find that a higher *Input Rank* is positively associated with higher odds that that input is vertically integrated, relatively more when final demand is elastic. A supplier's *Input Rank* remains a significant predictor of a firm's decision to integrate even after controlling for the relative positions on *upstreamness*/*downstreamness* segments.

Keywords: production networks; global value chains; Input Output tables; vertical integration; multinational enterprises; eigenvector centrality

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@alumni.imtlucca.it. Laboratory for the Analysis of Complex Economic Systems, piazza San Francesco 19 - 55100 Lucca - Italy.
[§] Mail to kenan.huremovic@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 production network that starts with the idea of a product by engineering, design or research labs, and finally reaches the consumers after a series of technological steps, including the manufacturing of parts and components, an assembly line, and the provision of post-production services by marketing, advertising and distribution industries.

In fact, the technical configuration of production processes can be much complex and recursive in nature when the same intermediate goods and services are repeatedly needed over a supply network. A global fragmentation of production processes can originate either spider-like or snake-like configurations, depending on technological peculiarities (Baldwin and Venables, 2013). Take logistics and distribution services, which are crucial in the delivery of intermediate inputs to companies, as well as in the delivery of final goods to consumers. Or else, consider the case of most innovative activities, which may require the services of R&D labs at different stages of completion before final delivery to consumers.

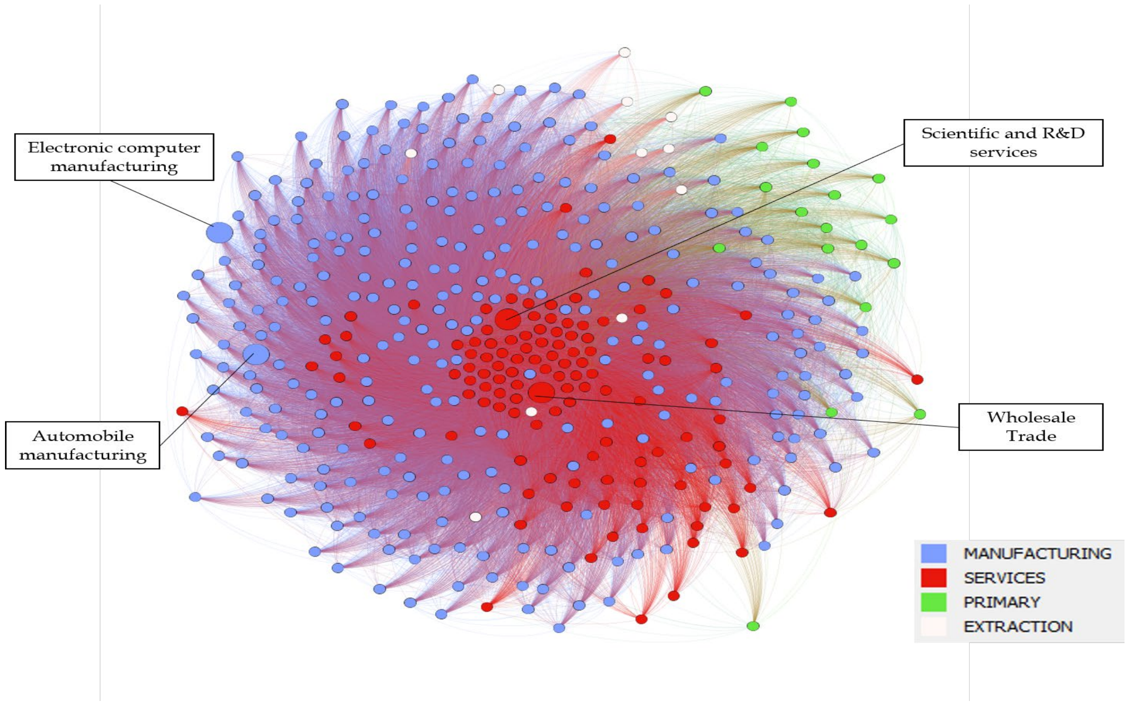
Nevertheless, Global Value Chains (GVCs) have been mainly studied assuming a separation of tasks over linear sequences, i.e. the *'chains'*, oriented on *upstream-downstream* directions, therefore neglecting the recursive nature of modern production (Costinot, Vogel and Wang, 2013; Antràs and Chor, 2013; Fally and Hillberry, 2015; Antràs and de Gortari, 2017; Alfaro et al., 2019; de Gortari, 2019). For the sake of simpler assumptions on both theory and empirics, previous works propose position metrics, e.g. the *upstreamness* or *downstreamness* of a production stage, which simulate productive sequences 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). It certainly is an advancement for understanding the mutual economic interdependence of firms organized over GVCs. However, linear approximations of complex network structures are likely to lead to an underestimation of the importance of some suppliers and an overestimation of the importance of others.

Take the case of the U.S. economy, which we plot as a production network[1] in Figure 1. According to the U.S. BEA 2002 Input-Output tables, we can represent the U.S. economy as a collection of 425 industries (i.e., nodes) linked by 51,768 transactions (i.e., edges). In Figure 1, we

[1] 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 within a production network whose structure presents a density of 0.286, i.e., the fraction of actual linkages out of all potential linkages. 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 in an output industry 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 where it is always possible to run seamlessly from one node to another just following input linkages.

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. On the other hand, 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. Now, let us consider the case of two consumer goods industries: Electronic Computer Manufacturing (code 334111) and Automobile Manufacturing (code 336111). They appear to be at the periphery of the U.S. production network because they mostly meet final consumers.

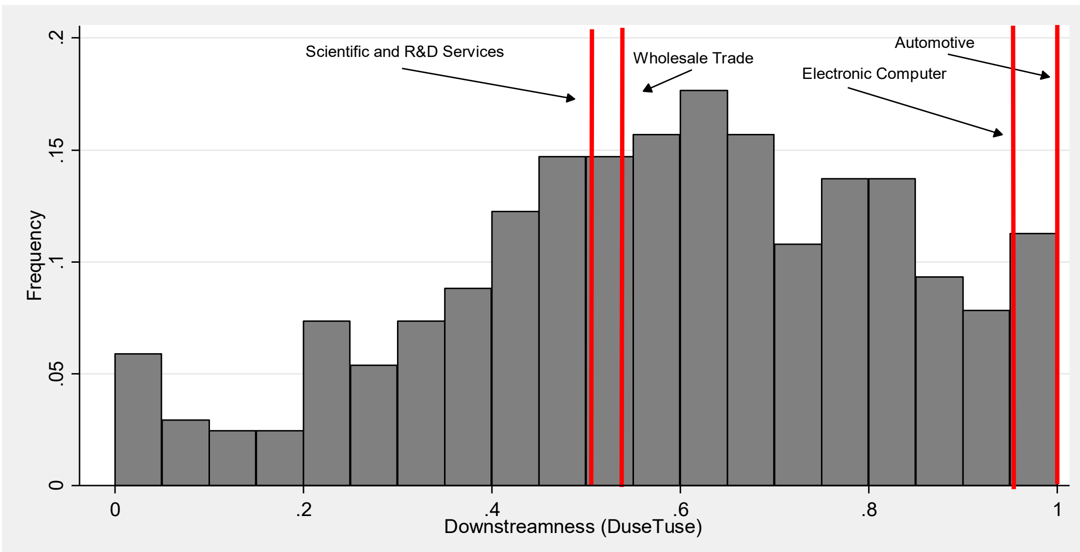
Figure 1: Input-Output Network from U.S. BEA 2002 I-O tables



Note: Nodes represent 425 6-digit NAICS industries from the U.S. BEA 2002 Input-Output tables. Edges represent 51,768 industry-pair transactions. Network density: 0.286. Average path length: 1.7 links. The graph is visualized using a Fruchterman and Reingold (1991) layout with the GEPHI software. More connected industries (weighted out-degree) at the center stage. Selected industries in evidence.

However, once we compare the network positions of selected industries in Figure 1 with their positions 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 an ideally linear supply chain. This is in contrast with the stylized chain we may have in mind, where a representative business line starts with R&D services and ends with distribution services. In fact, when we review computation methodologies, we find that *downstreamness* segments[2] are essentially obtained considering the weighted relative usage of inputs, intermediate *vis à vis* final, collapsing an otherwise complex production network on a linear sequence.

Figure 2: *Downstreamness* from the 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 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).

On the contrary, we argue that the mutually interactive and recursive nature of modern production is better understood when we consider the entire technological network, i.e. not only how inputs enter in a different order (*downstream vs upstream*), but also how central they are when they are requested as *inputs of inputs* at different stages of production.

In this respect, we introduce the *Input Rank*[3] as a bilateral measure of the technological relevance of any input-output relationship for the organization of global supply networks. We start by modeling the problem of a producer who plans the delivery of her output based on the requirements of both direct and indirect inputs. A network production function includes both labor and

[2] More recently, Alfaro et al. (2019) 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.

[3] When we compare the *Input Rank* with *Relative Upstreamness* sourced from Alfaro et al. (2019), we find that they convey different information. A Spearman rank correlation test shows that they are correlated – 0.31, with a p-value < 0.001.

intermediate inputs at each production stage. The most important upstream markets for a representative producer will be the ones that can have a higher impact on the firm-level marginal costs, when a friction transmits downstream, considering the topology of the supply structure. In our framework, a higher labor intensity at each stage potentially reduces the transmission of the shock from upstream markets. In addition, we model the possibility that a final producer has imperfect information on inputs of the supply network. In this case, the ability to outreach indirect suppliers is more limited in complex supply structures, and the producer can underestimate the role of transactions far away in the network structure.

For sake of comparison with previous studies, we compute the *Input Rank* on U.S. Input-Output tables, sourced from the Bureau of Economic Analysis (US BEA, 2002), and world Input-Output tables, sourced from both WIOD and EORA. Finally, we test the correlation of the *Input Rank* with choices of vertical integration, 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, relatively more when final demand is more elastic. We argue that vertical integration allows firms reducing frictions coming from upstream markets, and this is an incentive to enlarge the boundary to inputs that are more relevant in the supply network. Yet, we also find that parent companies preferably integrate inputs that are relatively proximate on the supply network, supporting our theory that limited knowledge of complex upstream markets makes a representative producer underestimate their impact on final production. Our findings are robust to different sample compositions, to changing empirical strategies, and to the inclusion of *downstreamness/upstreamness* metrics.

The rest of the paper is organized as follows. The next section positions our contribution with respect to related literature. Section 3 introduces a compact theory for the *Input Rank*. In Section 4, we compute the *Input Rank* on both the U.S. and worldwide Input-Output tables to describe 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 6.

2 Related literature

A flourishing strand of research studies how the network dimensions in the organization of production can contribute to explaining the response of aggregate fluctuations to microeconomic shocks (Acemoglu et al., 2012; Carvalho, 2014, Acemoglu et al., 2016). According to Oberfield

(2018), buyers and suppliers establish linkages that determine both individual and aggregate productivities, as the organization of a network is the result of endogenous collective choices. From an international perspective, Chaney (2014) studies the dynamic formation of trade networks based on searching processes of partners by exploiting direct and indirect contacts in destination markets. More in general, the literature on trade and production networks is still in its infancy, and many questions remain unanswered (Bernard and Moxnes, 2018).

This is the case of the emergence of GVCs, which are mainly modelled and tested as supposedly linear technological sequence (Fally, 2012; Antràs et al., 2012; Antràs and Chor, 2013; Miller and Temurshoev, 2017; Wang et al., 2017; Antràs and Chor, 2018; Alfaro et al., 2019), even if the existence of spider-like *vs* snake-like configurations has been acknowledged as depending on engineering details (Baldwin and Venables, 2013). A first step in modeling a production network has been made by Antràs and de Gortari (2017), who assume that a linear technology interacts with central geographic locations. Richer information on the configuration of GVCs is also exploited by de Gortari (2019) to build numerical counterfactuals on the transmission of value from inputs to outputs.

In our contribution, we introduce the *Input Rank* as a network position measure of the technological relevance of any direct or indirect input for a representative producer in an output industry. The *Input Rank* considers the recursive nature of real-world webs of suppliers and buyers when trade or contractual frictions can be encountered at any production step. We model the problem of input ranking on a nested production function that encompasses upstream markets because what happens in any upstream market has consequences on the ability to delivering a final output. A producer will rank any direct or indirect input relatively higher when a sourcing friction on that input market transmits downstream with a higher impact on marginal costs. Realistically, we provide for the possibility that a producer may have only a limited knowledge of the entire supply network, discounting relatively more the risk of frictions coming from faraway upstream industries.

Our measure is to some extent inspired by the PageRank centrality first applied in social networks and search engines (Brin and Page, 1998), to assess the relevant information consumed by internet users. The PageRank tool has by now spread to many different domains[4], from biology and genetics to financial debts, bibliometrics, and engineering of road networks Gleich (2015). The main idea of the PageRank is that a web page is more important if other important web pages have hyperlinks pointing to it. In our framework, the *Input Rank* assumes that:

[4] 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. For a basic introduction to network position metrics, see Early and Kleinberg (2010).

i) a (direct or indirect) input that is relatively more requested to produce other (direct or indirect) inputs must rank relatively higher;

ii) a (direct or indirect) input that is relatively more requested to produce other highly requested inputs is relatively more relevant than a (direct or indirect) input that delivers to less-requested inputs.

Eventually, we test the association of the *Input Rank* against vertical integration choices by U.S. multinational enterprises. In this, we relate to the recent strand of research that studies the firm-level organization of GVCs[5]. 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 and participate 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 GVC assumes different shades of meaning.

We stream our work following the intuition by Antràs and Chor (2013) and Alfaro et al. (2019), who model a supply chain as a technology made of production stages where each *downstream* output depends on a set of *upstream* (direct or indirect) inputs. In that framework, all producers shall rely on a surplus from the sale of the final output, and economic dependence is established along the supply chain, for how that surplus is optimally generated by and allocated among producers. In this case, the main prediction is that final-good producers integrate stages that are relatively more *downstream* (*upstream*) when final demand is sufficiently elastic (inelastic). However, when it comes to firm-level empirics, Del Prete and Rungi (2017) find that vertical integration choices are not always in line with theoretical predictions, as parent companies and affiliates locate not so far from each other along *upstreamness/downstreamness* segments.

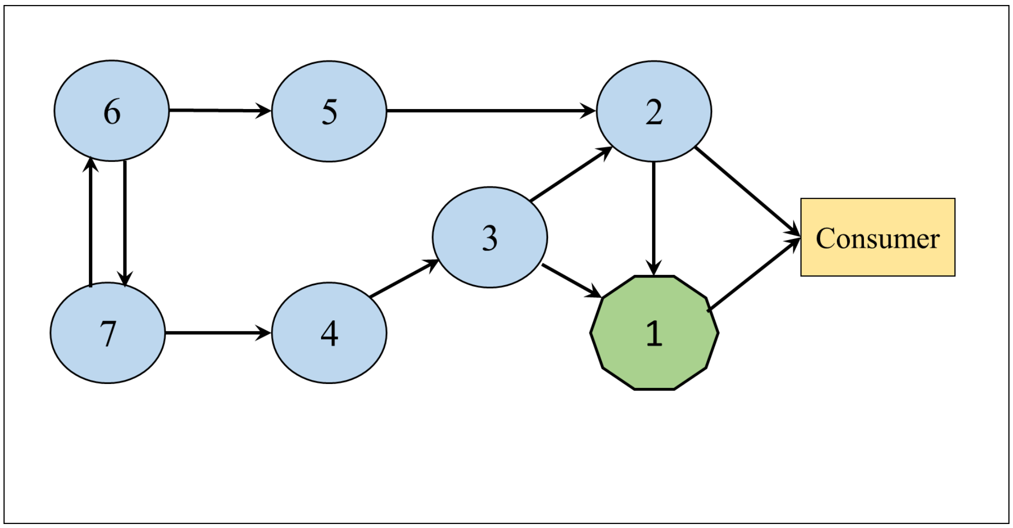
In this contribution, we build on a similar framework and find that a higher *Input Rank* is always associated with higher odds that that input is vertically integrated, even after controlling for the position on the *downstreamness/upstreamness* segments, and that proximate inputs on the network are more likely integrated than distant ones.

[5] 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.

3 A model for the Input Rank

In this section, we lay out the theoretical foundations for input ranking over supply networks. Assessing the importance of a supplier of a given firm is not a straightforward task when production processes are fragmented. To illustrate this point, we start with a stylized example depicted in Figure 3, where nodes indicate sectors or, alternatively, representative firms from those sectors, while directed links indicate the flows of goods or services. The output of each firm can be used as an intermediate input, as well as a consumption good.

Figure 3: A stylized supply network



Let us focus on the supply chain of firm 1. Failure of firm 4 to provide an appropriate input to firm 3 creates a friction that affects firm 3's production process. The friction is (partially) passed down to firm 1 and firm 2 because both use input 3 in their production. For instance, the aforementioned friction can imply an increase in the price of the intermediate input provided by firm 4, which will increase the production cost of firm 3. This will, in turn, affect both firm 1 and firm 2 that use good 3 in the production. Eventually, the production process of firm 1 is affected by both firm 2 and firm 3. More in general, we expect that a firm will be more affected by a distortion hitting an upstream supplier if: i) firms in the economy rely more on the deliveries of intermediate inputs; ii) the network is more connected, in the sense that there are more paths starting from the affected supplier and leading to that firm.

When supply networks become more complex, it is also conceivable that the manager of firm 1 does not fully observe what is happening on upstream markets, and therefore she cannot establish the relevance and the magnitude of a shock coming from those markets. This is especially true for suppliers that are relatively more distant in the supply network. For instance, the manager of firm 1 may not be able to observe the quality and the quantity of intermediate inputs that firm 6 and 7 reciprocally exchange.

In line with the above intuition, we present below a theory for the *Input Rank* that accounts for the topology of a supply network where frictions (e.g. trade barriers or contractual institutions) potentially transmit downstream, including the case that knowledge of indirect suppliers is more limited. Our theoretical framework is in many respects a standard for production networks[6], and thus we present it in a quite compact manner. Formal proofs of the claims are in Appendix A.

There are two types of agents in the economy: firms and the representative consumer. We denote the full set of firms in the economy with *N*. Firms group in *M* sectors. Each firm belongs to exactly one sector, and it produces a single differentiated variety of a sector-specific good.

3.1 Consumers

The representative consumer owns all firms in the economy and supplies one unit of labor inelastically. The preferences of the consumer over *M* goods are defined with the following Cobb-Douglas utility function:

$$U(c_1, c_2, \dots, c_M) = C = \theta \prod_{k=1}^{M} c_k^{\gamma_k} \tag{1}$$

where c_k is the consumption of good k and $\sum \gamma_k = 1$, while the parameter $\theta = \prod_{k=1}^{M} \gamma_k^{-\gamma_k}$ is a normalizing constant that simplifies computations. The composite consumption good k is defined with:

$$c_k = \left(\sum_{i=1}^{M_k} c(k,i)^{\frac{\varepsilon_k - 1}{\varepsilon_k}} \right)^{\frac{\varepsilon_k}{\varepsilon_k - 1}} \tag{2}$$

where $c(k,i)$ is the consumption of variety i of a good k, $\varepsilon_k > 1$ is the elasticity of substitution across varieties of good k, and M_k denotes the number of firms, each producing a different variety i in the kth sector.

[6] See also Baqee (2018) and Grassi (2017).

The consumer maximizes her utility subject to the following budget constraint:

$$\sum_{k=1}^{M}\sum_{i=1}^{M_k} p(k,i)c(k,i) = \sum_{k=1}^{M}\sum_{i=1}^{M_k} \pi(k,i) \tag{3}$$

where $p(k,i)$ is the price of variety i in sector k, $\pi(k,i)$ is the profit of firm i in sector k, and w is a worker's wage bill.

3.2 Firms

Firms in the k-th sector are symmetric and benefit from the same technology with constant returns to scale that combines labor and intermediate inputs. Each firm in a sector produces an imperfectly substitutable variety i of good k. Let us denote with $y(k,i)$ the output of firm i from sector k, with $l(k,i)$its labor input, and with $x(k,i,h,j)$ the amount of variety j of good h used in the production of variety i of good k. The profits of any firm i in sector k are simply defined with:

$$\pi(k,i) = p(k,i)y(k,i) - \sum_{h=1}^{M}\sum_{j=1}^{M_h} p(k,i)x(k,i,h,j) - wl(k,i) \tag{4}$$

The production function of firm i in sector k is defined by

$$y(k,i) = \zeta_k l(k,i)^{\beta_k}\left[\prod_{h=1}^{M}\left(\sum_{j=1}^{M_h}\big(\tau_h x(k,i,h,j)\big)^{(\varepsilon_h-1)/\varepsilon_h}\right)^{g_{hk}\varepsilon_h/(\varepsilon_h-1)}\right]^{\delta_k} \tag{5}$$

where β_k and δ_k are standard Cobb-Douglas elasticity parameters such that $\delta_k + \beta_k = 1$, while $\zeta_k = \left({\beta_k}^{-\beta_k}\prod_{h=1}^{N}{g_{hk}}^{-\delta_k g_{hk}}\right)$ is a normalizing constant that simplifies computations. Due to the Cobb-Douglas nature of the production function, the non-negative vector $(g_{hk})_{h=1}^{M}$ reflects the relative intensity with which firms in sector k use intermediate inputs $h \in \{1,2,\dots,M\}$. Hence, the sector level production structure of the economy is characterized by the (column-stochastic) adjacency matrix $\mathbf{G} = (g_{hk})_{h,k=1}^{M}$.

Crucially, we include in (5) an input-specific productivity parameter, τ_h, which catches any general distortion/friction encountered on an upstream market h. In this basic framework, market frictions are assumed exogenous and valid for all buyers of the same input[7]. The higher the friction for an input market the lower the parameter, τ_h. Distortion τ_h, in a reduced form, captures any type of distortion that results in a decrease in the productivity of intermediate inputs produced in sector *h*. For instance, when it is difficult to write a contract for sourcing goods produced in a sector *h,* it is more likely that firms will find those inputs from sector *h* less compatible with the production process, hence less productive. Alternatively, in our model, the friction τ_h will have the same effect as a price wedge. Indeed, when there is a friction τ_h, the price of 1 unit of good *h* (with productivity 1) will effectively be scaled up by factor $1/\tau_h$. One obvious case of such market frictions is tariff or non-tariff barriers, which increase the *per-unit* input price.

At this point, we define the composite intermediate input as an aggregate of varieties produced in a sector *h*, in the form:

$$x(k,i,h) = \left[\sum_{j=1}^{M_h} [x(k,i,h,j)]^{\frac{\varepsilon_h - 1}{\varepsilon_h}}\right]^{\frac{\varepsilon_h}{\varepsilon_h - 1}} \tag{6}$$

Therefore, we can rewrite (5) as

$$y(k,i) = \zeta_k l(k,i)^{\beta_k} \left\{\prod_{h=1}^{M} [\tau_h x(k,i,h)]^{g_{hk}}\right\}^{\delta_k} \tag{7}$$

3.3 Market equilibrium

We assume that firms in a sector compete in a monopolistic competition environment, and thus set their price to a constant markup over marginal costs. Following Atkeson and Burstein (2008) and Grassi (2017), we assume that firms set their prices taking as given the other sectors' prices and quantities, the wage bill, and the aggregate prices and quantities. We are now ready to introduce a notion of market equilibrium we envisage, whose existence and uniqueness follow from standard arguments. For details, see for instance Baqaee (2018).

Definition 1. *[Market Equilibrium] A market equilibrium is a collection of prices* $p(k,i)$*, wage w, input demands* $x(k,i,h,j)$*, outputs* $y(k,i)$*, consumption* $c(k,i)$*, and labor demands* $l(k,i)$ *such that:*

[7] One could further model this parameter as firm-specific, $\tau(h,j)$ or even buyer-supplier specific, $\tau(h,j,k,i)$, assuming that responses to frictions are heterogeneous. For the purpose of our analysis, we rule out heterogeneity within an industry and impose that any friction hits all varieties in an industry in the same way.

(i) Each firm maximizes its profits taking as given the sector price level and demand,

(ii) The representative consumer chooses consumption to maximize utility,

(iii) Markets for each good and labor clear.

3.4 The Input Rank

For our purpose, we consider two scenarios. First, we discuss a case when firms perfectly observe the structure of their supply network, i.e. they know the full technology made of direct and indirect inputs needed to deliver their output. More realistically, in a second scenario, we assume that firms have imperfect information on their supply networks, which may include a relatively high number of input-output relationships.

Before presenting our results, let us introduce further notation. We call a diagonal matrix **D** that matrix that contains information about sector-specific intermediate input elasticities, $\{\delta_r\}_{r=1}^{M}$. That is, each element of the matrix **D** tells us how much intermediate inputs a representative firm in a sector uses over total inputs in the equilibrium. Further, we introduce parameter $\chi_k \in (0,1]$ that captures the probability (share of) suppliers of each firm that a manager of firm $i \in k$ observes. The following definition formally introduces the concept of the *Input Rank* in production networks.

Definition 2. [Input Rank] *Denote with* e_k *the k-th unit vector. We define the Input Rank of a supplier of an input h relative to the producer of an output k as:*

$$\upsilon_{hk}(\mathbf{GD}, \chi_k) = \mathbf{e}_k'[\mathbf{I} - \chi_k \mathbf{DG}']^{-1}\mathbf{e}_h = \mathbf{e}_h'[\mathbf{I} - \chi_k \mathbf{GD}]^{-1}\mathbf{e}_k \tag{8}$$

In other words, the bilateral *Input Rank* $\upsilon_{hk}(\mathbf{GD}, \chi_k)$ is *(h,k)*-th element of the matrix $[\mathbf{I} - \chi_k \mathbf{GD}]^{-1}$. In a special case, when all sectors have the same elasticity of the output with respect to intermediate inputs, $\delta_r = \delta, \forall r = 1,2, \dots, M$, the *Input Rank* essentially captures the sum of all weighted paths from k to h in the production network, moving upstream through the network, where paths of length d are discounted by a factor $(\chi\delta)^d$. Finally, we note that the inverse in (8) exists since $\mathbf{G}$ is a column stochastic matrix, hence the spectral radius of $\mathbf{G}$ is 1. Since $\chi_k \leq 1$ and $\delta_k \leq 1,\ \forall k = 1,2, \dots, M$, the inverse in equation (8) will also exist.

From the perspective of a producer, the *Input Rank* vector, $\boldsymbol{v}_k = (\nu_{hk})_{h=1}^{M}$, encodes the information on the structure of the production technology on her supply network, and a possibly limited ability to outreach all suppliers.

Perfect information on the supply network

Let us start considering the case when a producer perfectly observes her supply network, i.e. when $\chi_k = 1$. The following proposition relates the *Input Rank*, $\upsilon_{hk}(\mathbf{GD}, 1)$, and the marginal costs of a producer, $\lambda(k, i)$, after considering her entire supply network.

Proposition 1. *Let $\breve{\lambda}(k, i)$ denote the logarithm of the marginal cost of production of firm i in sector k. Let $\breve{\tau}_h$ denote the logarithm of the friction on the upstream market h. Suppose that firms perfectly observe the production network. Then:*

$$\frac{\partial \breve{\lambda}(k, i)}{\partial \breve{\tau}_h} = -\upsilon_{hk}(\mathbf{GD}, 1) = -\mathbf{e}_h'[\mathbf{I} - \mathbf{GD}]^{-1}\mathbf{e}_k \tag{9}$$

when $\delta_r = \delta, \forall r$, then $\mathbf{GD} = \delta\mathbf{G}$.

We report proof of Proposition 1 in the Proof Appendix. In a nutshell, the higher the potential impact of frictions from an upstream market on marginal costs, the higher the ranking of that input from the perspective of a downstream buyer. In our framework, a lower distortion implies a higher productivity parameter, τ_h, hence a decrease in the marginal costs of production. Eventually, the impact of any upstream friction on the marginal costs of a final producer is a function of both the structure of the supply network, **G,** and of the relative input intensities, $(\delta_r)_{r=1}^{M}$, of each industry.

Imperfect information on the supply network

Suppose now that a producer *i* observes her supply network with some imperfections. That is, a firm *i* in a sector *k* has a probability χ_k of observing suppliers of firms of any given sector in the network.[8] We assume that this probability is output-specific, i.e., it varies across end-use sectors. In this way, we explicitly contemplate the possibility that some supply networks are too complex to explore, and a producer is able to assess the contribution of any indirect input based solely on the portion of technology she is able to observe. The following corollary directly follows from Proposition 1:

[8] More realistically we may consider the case when firms in sector *k* observe suppliers of firms in *r* with independent probability χ_{kr}, $r \in M$. Then we would replace scalar χ_k with diagonal matrix $\mathbf{H}_k$ that has diagonal elements equal to χ_{kr}. One possible interesting interpretation of χ_{kr} in that case is that it would also capture the contractibility of a sector, in the sense of Rauch (1999).

Corollary 1. *Assume that each supplier of firms in sector k is observed with probability χ_k, then:*

$$\frac{\partial \check{\lambda}(k,i)}{\partial \check{\tau}_h} = -\upsilon_{hk}(\mathbf{GD},\chi_k) = -\mathbf{e}_h'[\mathbf{I}-\chi_k\mathbf{GD}]^{-1}\mathbf{e}_k. \tag{10}$$

when $\delta_r = \delta, \forall r,$ and $\chi_k = \chi\ \forall k,$ then $\chi_k\mathbf{GD} = \chi\delta\mathbf{G}$.

Briefly, from the perspective of final producer i, any friction on an upstream market can be perceived as less important the more difficult its exploration because it is more distant in the supply network, given the damping factor, χ, and considering the relative usage of intermediate inputs, δ.

To understand better the intuition introduced above, it is worth looking back once again at the fictional supply network reported in Figure 3. When exploring her supply network, any time the manager of a firm 1 tries to collect information about upstream transactions, say about transactions between firm 3 and its suppliers, she has a limited ability to know the quality and quantity of deliveries. She can call the direct supplier and ask or, alternatively, she can gather information on the market when, for example, prices and quality of upstream inputs are relatively standard. However, at any further passage upstream, e.g. from firm 4 up to firm 6, the same problem starts all over again. Eventually, the dumping rate χ discounts distant nodes relatively more than proximate nodes, hence assuming that the ability to outreach on indirect suppliers is decreasing with the distance on the supply network[9].

On the other hand, some industries are relatively less input-intensive than others, hence when δ is low, the impact of a missed delivery by a direct or indirect supplier has a smaller impact on the final producer.

3.5 Observations

On the role of centrality

Both in the case of perfect or imperfect information, the structure of the supply network, $\mathbf{G}$, is crucial to understand the impact of any frictions or shocks coming from upstream markets. From this point of view, the *Input Rank* catches the features of sophisticated technological processes, when some inputs are used more than once and with a different intensity on a firm's supply network. We insist here on the characteristics of the *Input Rank* as an eigenvector centrality, which

[9] This is easy to see once we look at the recursive computation: $[\mathbf{I}-\chi\delta\mathbf{G}]^{-1} = \mathbf{I} + \chi\delta\mathbf{G} + (\chi\delta)^2\mathbf{G}^2 + (\chi\delta)^3\mathbf{G}^3 + \ldots$

measures the technological relevance of each input from the perspective of a downstream producer, such that:

i. an input is more technologically relevant if it is requested to produce other (direct or indirect) inputs;
ii. an input is more technologically relevant in downstream industries if it is requested to produce other highly requested (direct or indirect) inputs.

For a visual intuition of the previous characteristics, let us look back at the fictional supply network of Figure 3. The first property is evident if we compare the roles of firms 2 and 3. They are both direct suppliers of firm 1, but firm 3 is relatively more central in the supply network of firm 1 because it also delivers to firm 2. If firm 3 fails to deliver, firm 2 can also have problems and the impact on firm 1 is magnified.

The second property is evident if we compare the roles of firms 4 and 5. They are both indirect suppliers of firm 1 located at the same distance on the supply network. Yet, firm 4 is relatively more important in the supply network of firm 1 because it delivers to firm 3 that is, in turn, more relevant because more 'central' among direct suppliers. In other words, firm 4 borrows some 'centrality' from firm 3. In fact, if firm 4 does not deliver, firm 3 will receive the distortion and will pass it to firm 1 through two different production paths.

On the role of frictions on input markets

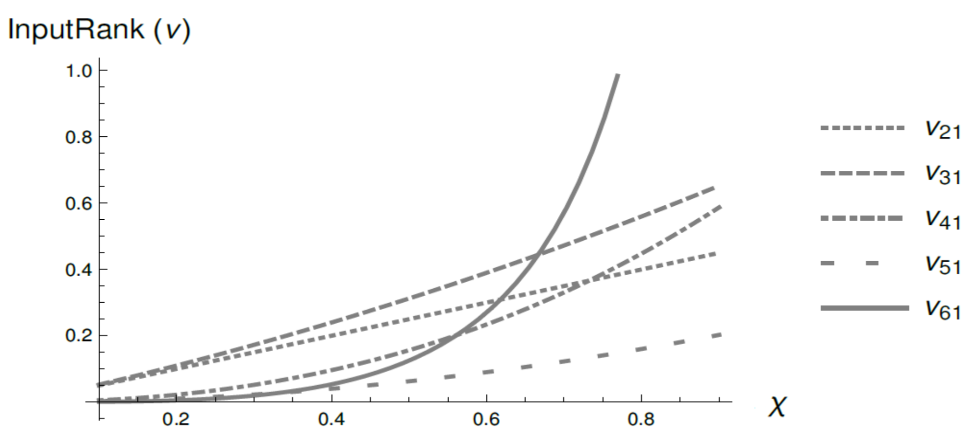
There are plentiful reasons to consider frictions over supply networks. For example, when an input is sourced from abroad, tariff or non-tariff barriers can reverberate to downstream buyers and have an indirect impact on marginal costs. Among others, a change in contracting frictions on upstream markets can have a similar impact on a downstream producer. Whether the input is delivered from abroad or not, a buyer and a supplier will never be able to sign the perfect contract that provides a detailed understanding of all the responsibilities and requirements, eliminating forever the risk of later disputes. Hence, an input delivery may not be entirely contractible, and we can assume that the degree of contractibility is specific to an input market, i.e. encompassing all the varieties offered on that market. Accordingly, our productivity parameter, $0 < \tau_h \leq 1$, catches the compatibility of that input in the production process. In our case, a higher contractibility on an input market implies a higher productivity parameter for that input, and a lower impact on marginal costs, as (9). Alternatively, one can think of the productivity parameter, τ_h, as a searching cost on an input market. In this case, a lower searching cost for any single input will reduce firm-level marginal costs. This approach is in line with the original intuition by Rauch (1999), who considered an input relatively more contractible when it is sold on organized exchanges or when its price is referenced.

However, please note how the source of any friction is considered always exogenous in our simple framework, as is the organization choice of the firm. Our basic model does not explicitly entail a case of intra-firm *vis à vis* arm's length exchange of inputs. Nonetheless, we can argue that, in an environment like the one captured by our model, it could be more beneficial for any firm, $i \in k$, to integrate a supplier of an input, $j \in h$, if the expected decrease in the per-unit cost of production, $\frac{\partial \check{\lambda}(k,i)}{\partial \check{\tau}_h}$, is larger after vertical integration. That is, if we expect that any friction becomes lower or any shock can be controlled after vertical integration, making an input more productive. Then, in our framework the productivity parameter τ_h becomes bigger and the impact on a firm's marginal costs is positive. This may be accomplished, for instance, through a better coordination of the production process or after a more efficient contract enforcement. We test this correlation in Section 5, although we can say nothing more on the optimal organization of the firm boundary, e.g. how many inputs are made *in-house* and how many are bought on the market.

On the role of the damping rate

We model imperfect information about a supply network as an output-specific dumping rate $\chi_k \in (0,1]$, which encodes the ability of a representative producer in an industry to search information on upstream markets. From a practical point of view, when χ_k is smaller, direct and indirect suppliers that are relatively closer to the final producer will have a relatively higher *Input Rank* than more distant suppliers.

Figure 4: The damping rate in a stylized supply network



Note: On the y-axis, a simulation of the *Input Rank* estimated on the fictional supply network in Figure 3. To make our point as clear as possible, we assume the economy does not include the labor input, thus $\delta = 1$.

To illustrate this property, we plot comparative statics of the *Input Rank* of main nodes detected in Figure 3 as a function of the damping rate. For simplicity, we assume that all inputs of a given firm are symmetric, meaning that for a fixed node j and any two of its suppliers r and s, we have $g_{rj} = g_{sj}$. Interestingly, although there are more paths connecting firm 1 to firm 6 than firm 1 to firm 2, firm 2 will eventually have a disproportionately higher *Input Rank* when the damping rate is smaller and smaller. Please note how, empirically, the average value of δ across sectors in the US economy is estimated 0.5 by Acemoglu et al. (2012), hence we can expect the damping factor $\chi\delta$ to be relatively small on average. In the following analyses, we first test a damping factor calibrated exclusively on 0.5, and then we discount it using input average contractibility *à la* Rauch (1999).

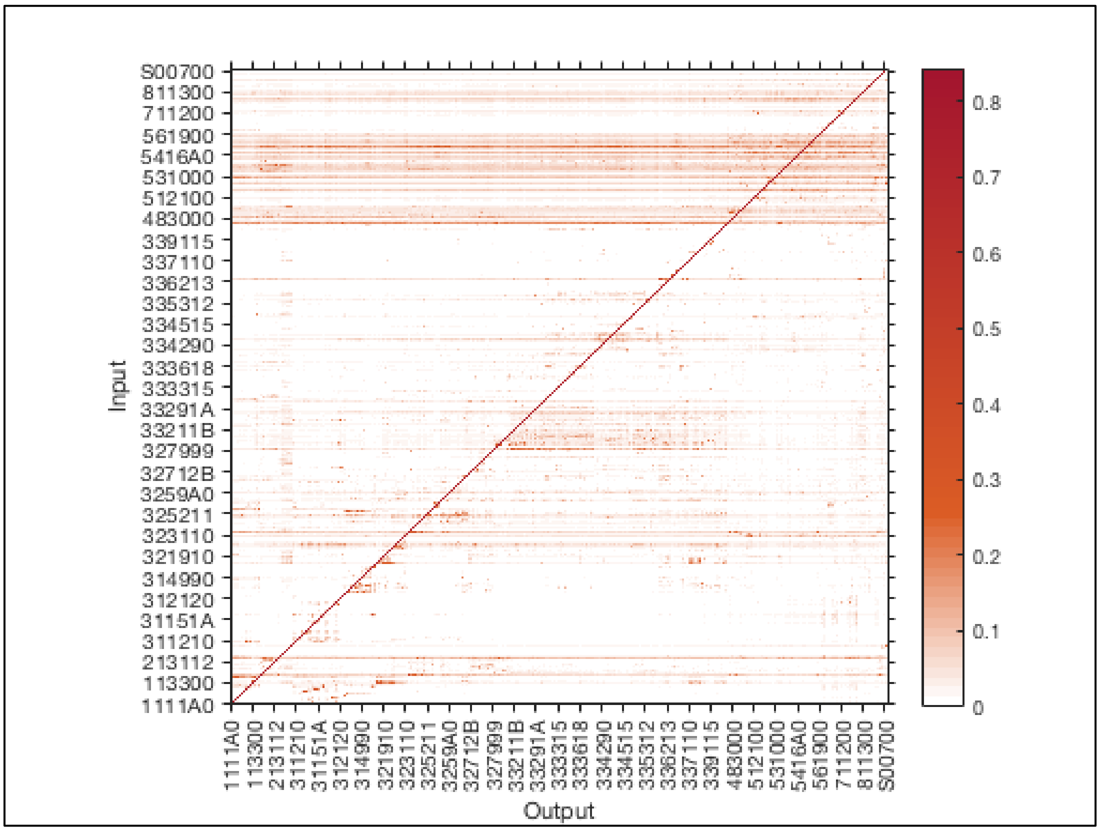
On the role of the elasticity of substitution

In the Proof Appendix, we include a demonstration that the more elastic is the demand of the final producer the higher the proportional impact of frictions on firm-level profits, $\frac{\partial \breve{\pi}(k,i)}{\partial \breve{\tau}_h \partial \varepsilon_k} > 0$, whenever $\varepsilon_k > 1$. We already gather from Proposition 1 that any downstream producer has an incentive to reduce frictions over her supply network to avoid lesser marginal costs. Here we add that such an incentive is higher when the demand faced by the final producer is relatively more elastic. See also the role of elasticity of substitution tested on vertical integration choices in Section 5.

4 Applications of the Input Rank

We compute the *Input Rank* on the U.S. and world Input-Output tables. U.S. I-O 2002 tables, compiled by the Bureau of Economic Analysis (BEA), sketch a reasonably fine-grained supply network established among 6-digit industries. The same tables have been extensively used to study production networks (Carvalho, 2014), vertical integration choices (Acemoglu et al., 2009; Alfaro et al., 2016), and to compute *Upstreamness*/*Downstreamness* metrics (Antràs and Chor, 2013: Alfaro et al., 2019). In Figure 1, we already showed how a solid and complex production network emerges from these tables, made of 51,768 linkages established among 425 industries. After a closer look, we register a strong heterogeneity in the sourcing strategies at the industry level. For example, in Appendix Figures C1 and C2, 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. On average, the in-degree of an industry is higher than its out-degree. 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. On the other hand, 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. Yet, 'global' centralities measured by in- or out-degrees are of scarce interest to understand the 'local' role of an upstream industry with respect to each specific downstream output. More properly, the *Input Rank* shall return the technological relevance[10] of that input market considering the peculiar topology of a supply network for any representative producer in an output industry.

Figure 5: *Input Rank* computed on U.S. 2002 Input-Output tables (damping factor = 0.5)



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. We assume $\chi = 1$ and $\delta_k = \delta = 0.5$. Inputs on the y-axes and outputs on the x-axes by alphanumeric order. A darker cell implies that an input is more technologically relevant for an output.

In Figure 5, we visualize the results from the computation of the *Input Rank* as a matrix of industry-pair values. For the moment, we assume that producers do not have limits in exploring

[10] As in similar works that use I-O tables, we implicitly assume that the latter represent a technology made of input-output relationships fixed in the medium term. We assume that the bundles of inputs, as well as the order in which they come, are fixed in the medium term. Inputs and processes can change with innovation only in a longer term. Expenditure shares are more endogenous to changing demand-supply equilibria. Further assumptions include the absence of economies of scale and the existence of representative firms.

the supply network, thus $\chi = 1$, and that the share of intermediate inputs across industries is constant, thus $\delta_k = \delta = 0.5$. The latter figure finds support on estimates by Acemoglu et al. (2012) made for the ensemble of the U.S. economy. A darker cell in Figure 5 implies that that input industry is more technologically relevant for that specific output. Interestingly, in the upper part of the figure, we find that services industries are much important across many manufacturing and services industries. Among manufacturing outputs, a crucial role is played by Primary Metal Manufacturing (code 331), Fabricated Metal Product Manufacturing (332), and Mining industries (code 21).

In Tables 1 and 2, we report some moments of the *Input Rank* distributions, first for all the top 20 inputs, then for the top 20 manufacturing inputs excluding services. 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), which includes headquarters services by holding firms[11]. Some post-production services also rank relatively high, as expected, e.g. Wholesale Trade (code 420000) and Advertising (code 541800). Further, we spot on top of rankings Electric Power Generation (code 221100) and bank credit (Monetary Authorities and Depository Credit Intermediation, code 52A000). From Appendix Table C1, R&D input services (code 541700) are much relevant for General Federal Defense Government Services (code S00500) and 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). 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). When we look from the perspective of selected *root* industries (Electronic Computer Manufacturing, code 334111; Automobile Manufacturing, code 336111), in Appendix Tables C2 and C3, we find that the *Input Rank* is indeed much heterogeneous across production processes, with relatively high standard deviations across end-use industries.

In an Online Appendix, we also report computations of the *Input Rank* on world Input-Output Tables, namely WIOD (World Input-Output Tables, see Timmer et al., 2015) and Eora Global MRIO tables. Both databases have been extensively used in settings where the geographical dimension of GVCs is important. However, we prefer keeping our baseline analyses using U.S. tables for two main reasons. First, U.S. tables have a fine-grained disaggregation of industries that

[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."

reduces the possibility of mixing intermediate inputs and final goods[12]. Second, our theoretical setup does not provide any foundation for considering the country of origin of an input. In fact, we assumed that an industry-level technology is fixed in the medium-long term, but we cannot extend this assumption to the origin countries of a sourcing strategy.

Table 1: Top 20 inputs (all industries) by *Input Rank* (damping factor = 0.5), as 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

[12] Both WIOD data and Eora Global MRIO adopt 2-digit industry classifications. Please note how wider industrial aggregations may bias the *Input Rank*, as each aggregation potentially include a bigger set of intermediate inputs. This problem had been already acknowledged by Alfaro and Charlton (2009), when they found that 2-digit classifications led to an underestimation of vertical FDI and an overestimation of horizontal FDI.

Table 2: Top 20 inputs (manufacturing only) by *Input Rank* (damping factor = 0.5), as 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

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 both 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) or, alternatively, 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*.

5.1 A sample of U.S. parent companies

We source firm-level data from the Orbis database, compiled by the Bureau van Dijk. For our scope, we collect information on 20,489 U.S. parent companies controlling 154,836 subsidiaries around the world at the end of the year 2015[13]. 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 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, NAFTA members, i.e. 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 values for the year 2014.

For the scope of our analysis, we map industry affiliations of both parent companies and 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, including the *Input Rank* we computed in Section 4, the *Relative Upstreamness* segments sourced from Alfaro et al. (2019), and a measure of *Network Distance* between any industry pair[14] calculated on the same U.S tables. In the absence of actual data on firm-to-firm transactions, such a mapping[15] allows us proxying buyer-supplier relationships. Finally, we complement our data

[13] We follow international standards for the identification of corporate control 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 that is controlled after a concentration of voting rights (> 50%). See also Rungi et al. (2017). Similar data structures have been used in Alviarez et al. (2016), Cravino and Levchenko (2017), Del Prete and Rungi (2017).

[14] The *Network Distance* between any input and any output in the U.S. I-O tables is the minimum number of downstream linkages that connect them through

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

with industry-level estimates of demand elasticity sourced from Broda and Weinstein (2006), and with a measure of input contractibility retrieved from Antràs and Chor (2013).

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.

5.2 Baseline results

We test a conditional logit model with parent-level fixed effects, as it is a natural empirical strategy for the multinomial case with a set of *ex-ante* alternatives[16]. That is, we test the determinants of vertical integration choices controlling for the characteristics of both the production stages that are vertically integrated and not integrated by the parent company.

Let $h = 1,2,\dots,N$ denote the set of inputs, as from the input-output tables, and let $r = 1,2,\dots,R$ denote the set of parent companies, each active in an output industry, $k = 1,2,\dots,K$. The dependent variable, $y_{hr(k)}$, takes on a value 1 when at least one subsidiary in the *h*-th input market has been integrated by a parent *r* in industry *k*, and 0 otherwise. Therefore, for each parent company, we have a vector $\mathbf{y}_{hr(k)} = \left(y_{1r(k)}, \dots, y_{Nr(k)}\right)$ made of 0s and 1s when a h-th input has been integrated or not, respectively. At this point, we can consider the probability that a generic parent chooses among a set of alternatives such that:

$$\Pr\left(\mathbf{y}_{r(k)} \middle| \sum_{h=1}^{N} y_{hr(k)}\right) = \frac{\exp\left(\sum_{h=1}^{N} y_{hr(k)}\, \mathbf{x}_{hr(k)} \boldsymbol{\beta}\right)}{\sum_{s_h \epsilon \mathbf{S}_h} \exp\left(y_{hr(k)} \mathbf{x}_{hr(k)} \boldsymbol{\beta}\right)} \tag{11}$$

where $\mathbf{S}_h$ is a set of *ex ante* alternative binary choices and each of its elements, $s_h \,\epsilon\, \mathbf{S}_h$, is equal to 1 when the *h*-th input is integrated, and 0 otherwise. Therefore, we identify a vector of covariates for each input-parent pair, $\mathbf{x}_{hr}$,which includes: the *Input Rank* of the *h*-th input with respect to the *k*-th industry estimated with a damping factor equal to 0.5; the minimum distance in a supply network of any *h*-th input from a *k*-th output; a binary variable *Complements* relative to the *h*-th input market; the input-output *upstreamness* sourced from Alfaro et al. (2019); the input-specific *Contractibility* derived from Rauch (1999); the bilateral normalized *Direct requirement* coefficient from I-O tables. As in Antràs and Chor (2013) and Alfaro et al. (2019), the variable *Complements* is equal to 1 when the elasticity of substitution of the output market is below the median $(\rho_k > \rho_{med})$, and 0 otherwise $(\rho_k < \rho_{med})$. Errors are clustered by the parent company. Fixed effects are at the parent level. Results from nested specifications are reported in Tables 5.

The coefficient of immediate interest to us is the one on the *Input Rank*, which indicates whether the odds of vertical integration are higher for a more relevant input in the supply network. We do find that the coefficient of the *Input Rank* is positive and significant throughout all our estimates. Exponentiating the coefficients, we obtain a range of higher odds for vertical integration

[16] See McFadden (1974) and Chamberlain (1980) for more details. Present notation is in line with 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.

in a range between 1.21 and 1.56. In the first columns, we consider all parent companies, whether they are active in a manufacturing or a service industry. Please note that in further columns, when we introduce industry controls, the sample reduces to manufacturing parents only. This is mainly due to the inclusion of the elasticity of substitution by Broda and Weinstein (2006), as it is originally estimated only on manufacturing imports.

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

Dependent variable: Input is integrated ==1	**(1)**	**(2)**	**(3)**	**(4)**
Input Rank	0.445***	0.259***	0.185***	0.285***
	(0.011)	(0.024)	(0.016)	(0.006)
Input Rank * Complements		0.100***	0.098***	0.209***
		(0.025)	(0.018)	(0.018)
Input upstreamness		-0.812*	-0.843***	-0.566
		(0.415)	(0.335)	(0.618)
Input upstreamness * Complements		0.033	-0.177***	0.015
		(0.039)	(0.018)	(0.078)
Contractibility		-0.390***	-0.645***	-0.249***
		(0.017)	(0.032)	(0.025)
Input Network Distance	-0.115***	-0.134*	-0.256	-0.105**
	(0.028)	(0.069)	(0.309)	(0.055)
Direct requirement	0.093***	0.015*	0.010*	0.026
	(0.014)	(0.008)	(0.006)	(0.034)
Observations	8,564,068	1,151,908	595,218	542,872
N. parent companies	20,294	4,084	2,110	1,925
Pseudo R-squared	0.515	0.752	0.698	0.842
Clustered errors by parent	Yes	Yes	Yes	Yes
Activity of parent companies	All	Manu-facturing	Final goods	Interme-diate goods

Note: Input Rank estimated with a damping factor = 0.5. 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 proxies the relative technological distance between an input and a 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. (2019) 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 column in Table 5), although results are not significant anymore in the case of *midstream* parents (last column), i.e., the prediction by Antràs and Chor (2013) and Alfaro et al. (2019) is not verified in the case of integration choices started by producers of intermediate inputs. To extend the role of the elasticity of substitution to the case of supply networks, we include a similar interaction term between the variable *Complements* and 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. More in general, the latter result is in line with our basic framework (see Section 3.5 and Proposition 2 in Appendix A), according to which a more elastic final demand makes downstream buyers more vulnerable to upstream frictions. In this case, we can argue, vertical integration could be a way to reduce the impact of upstream frictions.

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.

5.3 Robustness checks

Our baseline findings are robust to several checks of robustness. In Table 6, we check whether sample compositions and changes in parameters can have an impact on the sign and significance of our coefficients of interest while keeping the *Input Rank* with a damping factor equal to 0.5. In the first column, we exclude inputs coming from the same 2-digit industry of the parent companies. In the second column, we exclude services inputs to check whether these are exclusively driving the correlation with the Input Rank, as we expect them to be more relevant on average, as from descriptive statistics. In the third column, we modify our indicator of *Complements*, explicitly considering the difference between the elasticities of the output and the one of each input, $(\rho_k - \rho_h)$, i.e. introducing a reference point internal to the supply network. In the fourth column, we reduce our sample to the Top 100 (direct or indirect) inputs with the highest *Input Rank*. 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 C4 and C5, we further control for: i) sample compositions when we consider only *midstream* manufacturing parents, i.e. parent that produce intermediate inputs; ii) empirical specifications different from the fixed-effects conditional logit. All main findings are similar in sign and significance with baseline estimates.

Finally, in Appendix Table C6, we modify the Input Rank by plugging in the contractibility index retrieved by Rauch (1999), in order to proxy a second component of the damping factor, χ_k, which allows us catching the knowledge of a supply network. All our main tenets are stable, although magnitudes of coefficients on the *Input Rank* generally increase.

Table 6: Robustness on sample composition, parent-level fixed effects conditional logit

Dependent variable: Input is integrated ==1	No horizontal	Only manufacturing inputs	Output minus input elasticity	Top 100 inputs
Input Rank	0.679***	0.251***	0.255***	0.334***
	(0.152)	(0.044)	(0.044)	(0.065)
Input Rank * Complements	0.106***	0.095***	0.078***	0.128***
	(0.026)	(0.023)	(0.015)	(0.026)
Input upstreamness	-0.892**	-0.768	-0.649*	-0.611**
	(0.468)	(0.630)	(0.334)	(0.309)
Input upstreamness * Complements	0.254	0.313	0.347	0.248*
	(0.156)	(0.240)	(0.651)	(0.132)
Contractibility	-0.511***	-0.289***	-0.413***	-0.456***
	(0.027)	(0.018)	(0.017)	(0.023)
Input Network Distance	-0.052	-0.095*	-0.105	0.089
	(0.079)	(0.049)	(0.085)	(0.126)
Direct requirement	0.028	-0.008	0.028*	0.040
	(0.035)	(0.005)	(0.017)	(0.055)
Observations	741,066	905,640	1,151,908	156,705
N. parent companies	2,637	3,903	4,084	2,847
Pseudo R-squared	0.118	0.285	0.290	0.309
Clustered errors by parent	Yes	Yes	Yes	Yes
Activity of parent companies	Manufacturing	Manufacturing	Manufacturing	Manufacturing

Note: Input Rank estimated with a damping factor = 0.5. 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 a measure to catch the technological relevance of direct and indirect inputs in the supply network. We frame the input ranking problem as the solution of a representative producer that needs minimizing the impact of frictions coming from upstream markets when they can hit her marginal costs in a network-like production function. The main intuition is that we must consider the peculiar topology of any supply network to derive the impact of any input on downstream marginal costs. Eventually, a (direct or indirect) input will rank relatively higher when considering the entire supply network: i) its direct requirements are higher; ii) it has a central position, i.e. it delivers to many other inputs; iii) it delivers to other inputs that have a central position; iv) the usage of intermediate inputs is higher than the usage of labor. For sake of comparison with previous GVC positioning metrics (e.g., *downstreamness* and *upstreamness* segments), we compute it on U.S. 2002 BEA Input-Output tables, and then we test how it correlates with firm-level choices of vertical integration made by U.S. parent companies worldwide. We do find that an input with a higher ranking more likely will be integrated within the boundary of the firm, even more so when the demand of the final product is more elastic. We argue that vertical integration allows downstream buyers reducing the possibility that frictions on upstream markets could hit their production processes, even more so when the margins from final sales are smaller. Our findings are robust to several checks on sample compositions, parameter choices, and empirical specifications. More in general, we argue that the *Input Rank* catches the recursive and complex nature of real-world supply networks, which have been too often represented as supposedly linear chains in studies on the international organization of production. Certainly, both empirics and theory need better considering the technological loops, kinks, and corners that can magnify or dampen a shock in a supply network, finally shaping the organizational response of the companies.

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 (2019). “Internalizing global value chains: A firm-level analysis”, Journal of Political Economy 127 (2): 509-559.

Alfaro, L. and Charlton A. (2009). "Intra-industry Foreign Direct Investment." American Economic Review, 99 (5): 2096-2119.

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 (2018). "On the Measurement of Upstreamness and Downstreamness in Global Value Chains". World Trade Evolution: Growth, Productivity and Employment, 126-194. Taylor & Francis Group.

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.

Atkeson, A. and Burstein, A. (2008). Pricing-to-market, trade costs, and international relative prices. American Economic Review, 98(5):1998–2031.

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.

Baqaee, D. R. (2018). Cascading failures in production networks. Econometrica.

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 (2018). Networks and Trade, Annual Review of Economics, 10:65-85.

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.

Chaney T. (2016). "Networks in International Trade", in Oxford Handbook of the Economics of Networks, (eds.) Bramoullé Y., Galleoti A. and Rogers B. Oxford University Press.

Chaney T. (2014). “The Network Structure of International Trade”, American Economic Review, 104 (11): 3600-34.

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.

de Gortari A. (2019). "Disentangling Global Value Chains", mimeo.

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

Easley D. and Kleinberg J. (2010). Networks, Crowds, and Markets: Reasoning about a Highly Connected World. Cambridge University Press.

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.

Grassi, B. (2017). IO in I-O: Size, Industrial Organization, and the Input-Output Network Make a Firm Structurally Important, mimeo.

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.

Newman, M. E. J. (2008) Mathematics of Networks. In: Palgrave Macmillan (eds) The New Palgrave Dictionary of Economics. Palgrave Macmillan, London.

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. and Del Prete D. (2018). “The smile curve at the firm level: Where value is added along supply chains”, Economics Letters, 164 (C): 38-42.

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 – Proof Appendix

Lemma 1. *Let N_k^+ denote the set of inputs of firms in sector k. The cost function of firm i is given by*

$$c[y(k,i);\ w,\boldsymbol{p}] = \lambda(k,i)y(k,i) \tag{A1}$$

where $\lambda(k,i) = w^{\beta_k}\prod_{h\in N_k^+} p_h^{\delta_k g_{hk}}\tau_h^{-\delta_k g_{hk}}$.

Proof of Lemma 1. The Lagrangian of the cost minimization problem of firm i from sector k is:

$$\mathcal{L} = wl(k,i) + \sum_{h=1}^{M} p_h x(k,i,h) - \lambda(k,i)\left[\zeta_k l(k,i)^{\beta_k}\left(\prod_{h=1}^{M}\left(\tau_h x(k,i,h)\right)^{g_{hk}}\right)^{\delta_k} - y(k,i)\right] \tag{A2}$$

From the first-order necessary conditions (also sufficient, given convexity), we deliver the following conditional demand functions:

$$x(k,i,h) = \lambda(k,i)\delta_k g_{hk}\frac{y(k,i)}{w} \tag{A3}$$

$$l(k,i) = \lambda(k,i)\beta_k\frac{y(k,i)}{w} \tag{A4}$$

Plugging (A3) and (A4) into the cost of firm *i*, it directly follows that $c[y(k,i);\ w,\boldsymbol{p}] = \lambda(k,i)y(k,i)$. Hence, $\lambda(k,i)$ is the marginal cost of production of firm *i*. Substituting (A3) and (A4) into the production function, we obtain:

$$\begin{aligned} y(k,i) &= \zeta_k\left[\frac{\lambda(k,i)\beta_k y(k,i)}{w}\right]^{\beta_k}\prod_{h\in N_k^+}\left[\tau_h\frac{\lambda(k,i)\delta_k g_{hk}y(k,i)}{p_h}\right]^{g_{hk}\delta_k} = \\ &= \zeta_k\,\lambda(k,i)y(k,i)\left(\frac{\beta_k}{w}\right)^{\beta_k}\prod_{h\in N_k^+}\left[\tau_h\frac{\delta_k g_{hk}}{p_h}\right]^{g_{hk}\delta_k} = \\ &= \lambda(k,i)w^{-\beta_k}\prod_{h\in N_k^+}\left[\frac{p_h}{\tau_h}\right]^{-g_{hk}\delta_k} y(k,i) \end{aligned} \tag{A5}$$

where for the last equality we used that $\zeta_k = \beta_k^{-\beta_k} \prod_{h \in N_k^+} (\delta_k g_{hk})^{-\delta_k g_{hk}}$. Solving for $\lambda(k,i)$, we get:

$$\lambda(k,i) = w^{\beta_k} \prod_{h \in N_k^+} p_h{}^{\delta_k g_{hk}} \tau_h{}^{-\delta_k g_{hk}} \tag{A6}$$

Q. E. D. □

Lemma 2. *The following relations are valid between firm-level marginal cost $\lambda(k,i)$ and sector-level marginal cost λ_k, and between firm-level markup $\mu(k,i)$ and sector-level markup μ_k hold:*

$$\lambda_k = M_k^{\frac{1}{1-c_k}} \lambda(k,i) \tag{A7}$$

$$\mu_k = \mu(k,i) = \frac{\varepsilon_k}{\varepsilon_k - 1} \tag{A8}$$

Proof of Lemma 2. Using results from the theory of monopolistic competition, the sector-level price of good k, p_k, and the sector-level output, y_k, are given with:

$$p_k = \left[\sum_{i=1}^{M_k} p(k,i)^{1-\varepsilon_k} \right]^{\frac{1}{1-\varepsilon_k}} \tag{A9}$$

$$y_k = \left[\sum_{i=1}^{M_k} y(k,i)^{\frac{\varepsilon_k - 1}{\varepsilon_k}} \right]^{\frac{\varepsilon_k}{\varepsilon_k - 1}} \tag{A10}$$

In the symmetric equilibrium, $p(k,i) = p(k,j)$, and $y(k,i) = y(k,j)$. Hence:

$$p_k = M_k^{\frac{1}{1-\varepsilon_k}} p(k,i) \tag{A11}$$

$$y_k = M_k^{\frac{\varepsilon_k}{\varepsilon_k - 1}} y(k,i) \tag{A12}$$

Thanks to the assumption of constant returns to scale and using the expression for y_k, we can write the sector level marginal cost of production as:

$$\lambda_k = \sum_{i=1}^{M_k} \lambda(k,i)\frac{y(k,i)}{y_k} = \frac{M_k\lambda(k,i)y(k,i)}{M_k^{\frac{\varepsilon_k}{\varepsilon_k-1}}y(k,i)} = M_k^{\frac{1}{1-\varepsilon_k}}\lambda(k,i) \tag{A13}$$

Finally, from the firm's pricing rule we have $p(k,i) = \mu(k,i)\lambda(k,i)$. Plugging the pricing rule in the expression for price from A (11) and A (12), we get:

$$p_k = \left\{\sum_{i=1}^{M_k}[\mu(k,i)\lambda(k,i)]^{1-\varepsilon_k}\right\}^{\frac{1}{1-\varepsilon_k}} = [M_k\mu(k,i)^{1-\varepsilon_k}\lambda(k,i)^{1-\varepsilon_k}]^{\frac{1}{1-\varepsilon_k}} =$$

$$= M_k^{\frac{1}{1-\varepsilon_k}}\mu(k,i)\lambda(k,i) = \mu(k,i)\lambda_k \tag{A14}$$

Q. E. D. □

Proof of Proposition 1. Taking logarithms[17] of (A6), using the pricing rule ($p_h = \mu_h\lambda_h$), and normalizing $w = 1$, we get:

$$\breve{\lambda}(k,i) = \delta_k \sum_{h\in N_k^+} g_{hk}(\breve{\mu}_k - \breve{\tau}_h) + \delta_k \sum_{h\in N_k^+} g_{hk}\breve{\lambda}_h \tag{A15}$$

From Lemma 2, we can write the above equation in terms of sector level marginal costs as:

$$\breve{\lambda}_k = \frac{1}{\varepsilon_k - 1}\breve{M}_k + \delta_k \sum_{h\in N_k^+} g_{hk}(\breve{\mu}_k - \breve{\tau}_h) + \delta_k \sum_{h\in N_k^+} g_{hk}\breve{\lambda}_h \tag{A16}$$

Let $\mathbf{E} = diag\left[\left(\frac{1}{\varepsilon_k-1}\breve{M}_k\right)_{k=1}^{M}\right]$. Writing the above equation for all k's in vector notation, we get:

$$\breve{\boldsymbol{\lambda}} = \mathbf{E} + \mathbf{DG}'\breve{\boldsymbol{\mu}} - \mathbf{DG}'\breve{\boldsymbol{\tau}} + \mathbf{DG}'\breve{\boldsymbol{\lambda}} =$$

[17] To ease notation, we indicate logarithms with an accent on variables.

$$= (\mathbf{I} - \mathbf{DG}')^{-1}[\mathbf{E} + \mathbf{DG}'(\breve{\boldsymbol{\mu}} - \breve{\boldsymbol{\tau}})] \quad \text{(A17)}$$

Finally, by differentiating we get:

$$\frac{\partial \breve{\lambda}(k,i)}{\partial \breve{\tau}_h} = \frac{\partial \breve{\lambda}_k}{\partial \breve{\tau}_h} = -\mathbf{e}_k'(\mathbf{I} - \mathbf{DG}')^{-1}\mathbf{DG}'\mathbf{e}_h = -\mathbf{e}_k'[(\mathbf{I} - \mathbf{DG}')^{-1} - \mathbf{I}]\mathbf{e}_h =$$
$$= -\mathbf{e}_h'[(\mathbf{I} - \mathbf{GD})^{-1} - \mathbf{I}]\mathbf{e}_k \quad \text{(A18)}$$

Whenever $k \neq h$, the above equation has a form:

$$\frac{\partial \breve{\lambda}(k,i)}{\partial \breve{\tau}_h} = -\mathbf{e}_h'[\mathbf{I} - \mathbf{GD}]^{-1}\mathbf{e}_k \quad \text{(A19)}$$

Q. E. D. ☐

Lemma 3. *Given the demand for good k, then partial derivative* $\frac{\partial \breve{y}(k,\ i)}{\partial \breve{\tau}_h}$ *is positive and increasing in* ε_k.

Proof of Lemma 3. Let $x(h,\ k,\ i)$ denote the aggregate demand for variety $i \in k$ by firms in a sector h. From the market clearing condition for variety *i*, we have

$$y(k,i) = c(k,i) + \sum_{h=1}^{M} x(h,k.i) =$$
$$= \gamma_k \left[\frac{p(k,i)}{p_k}\right]^{-\varepsilon_k} \frac{p_c C}{p_k} + \sum_{h=1}^{M} \delta_h g_{kh} \left[\frac{p(k,i)}{p_k}\right]^{-\varepsilon_k} \frac{p_h y_h}{p_k} =$$
$$= \gamma_k p(k,i)^{-\varepsilon_k} p_k^{\varepsilon_k - 1} p_c C + \sum_{h=1}^{M} \delta_h g_{kh}\, p(k,i)^{-\varepsilon_k} p_k^{\varepsilon_k - 1} p_h y_h \quad \text{(A20)}$$

Taking logs and simplifying, we get:

$$\breve{y}(k,i) = -\varepsilon_k \breve{p}(k,i) + (\varepsilon_k - 1)\breve{p}_k + \log(p_k y_k) \quad \text{(A21)}$$

As demand for good k is fixed, $(p_k y_k)$ remains constant, and we note that:

$$\frac{\partial \breve{p}_k}{\partial \breve{\tau}_h} = \frac{1}{p_k}\frac{\partial p_k}{\partial \breve{\tau}_h} = \frac{1}{M_k^{\frac{1}{1-\varepsilon_k}} p(k,i)} [M_k p(k,i)^{1-\varepsilon_k}]^{\frac{\varepsilon_k}{1-\varepsilon_k}} p(k,i)^{-\varepsilon_k} \frac{\partial p(k,i)}{\partial \breve{\tau}_h} =$$

$$= \frac{1}{M_k^{\frac{1}{1-\varepsilon_k}} p(k,i)} M_k^{\frac{\varepsilon_k}{1-\varepsilon_k}} \frac{\partial p(k,i)}{\partial \breve{\tau}_h} = M_k^{-1} \frac{\partial p(k,i)}{\partial \breve{\tau}_h} \frac{1}{p(k,i)}, \tag{A22}$$

where we used the symmetry property of the equilibrium $p(k,i) = p(k,j)\ \ \forall i,j \in k$, and the fact that $\frac{\partial p(k,j)}{\partial \breve{\tau}_h} = 0\ , \quad \forall j \neq i$. We now write:

$$\begin{aligned} \frac{\partial \breve{y}(k,j)}{\partial \breve{\tau}_h} &= -\varepsilon_k \frac{\partial \breve{p}(k,i)}{\partial \breve{\tau}_h} - (1-\varepsilon_k) M_k^{-1} \frac{\partial p(k,i)}{\partial \breve{\tau}_h} \frac{1}{p(k,i)} \\ &= -\left[\frac{\varepsilon_k (M_k - 1) + 1}{M_k}\right] \frac{\partial \breve{p}(k,i)}{\partial \breve{\tau}_h} \end{aligned} \tag{A23}$$

Finally, we note that $\frac{\partial \breve{p}(k,i)}{\partial \breve{\tau}_h} = \frac{\partial \breve{\lambda}(k,i)}{\partial \breve{\tau}_h}$. From Proposition 1, we know that $\frac{\partial \breve{\lambda}(k,i)}{\partial \breve{\tau}_h} < 0$ and that $\frac{\partial \breve{\lambda}(k,i)}{\partial \breve{\tau}_h}$ is independent of ε_k. From this and the fact that $M_k \geq 1$ and $\varepsilon_k > 1$, the claim follows directly.

Q. E. D. □

Proposition 2. *Given the demand for good k, then* $\frac{\partial \breve{\pi}(k,i)}{\partial \breve{\tau}_h \partial \varepsilon_k} > 0$ *whenever* $\varepsilon_k > 1$.

Proof of Proposition 2. We can write the profit of firm *i* as:

$$\pi(k,i) = \mu_k \lambda(k,i) y(k,i) - \lambda(k,i) y(k,i) =$$

$$= (\mu_k - 1) \lambda(k,i) y(k,i) = \frac{1}{\varepsilon_k - 1} \lambda(k,i) y(k,i) \tag{A24}$$

Then, taking logs and differentiating:

$$\frac{\partial \breve{\pi}(k,i)}{\partial \breve{\tau}_h} = \frac{\partial \breve{\lambda}(k,i)}{\partial \breve{\tau}_h} + \frac{\partial \breve{y}(k,i)}{\partial \breve{\tau}_h} \tag{A28}$$

From Proposition 1, we know that $\frac{\partial \breve{\lambda}(k,i)}{\partial \breve{\tau}_h}$ is proportional to the Input Rank and independent of ε_k. From Lemma 3, we have that $\frac{\partial \breve{y}(k,i)}{\partial \breve{\tau}_h}$ increases with ε_k. This means that $\frac{\partial \breve{\pi}(k,i)}{\partial \breve{\tau}_h}$ is also increasing in ε_k .

Q. E. D. □

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

The empirical intuition of the *Input Rank* comes from the 'personalized' version of the *PageRank* 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 *PageRank* 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, we are interested in the 'local' outreach of a specific *root* buyer in her oriented supply network. Therefore, we need a 'personalization' of the ranking problem in the spirit of Haveliwala (2003) and White and Smyth (2003), because different rankings are possible for different *root* nodes. Starting from the original formulation of the *PageRank*, adopting the notation proposed by Gleich (2015), the eigenvalue/eigenvector problem can be represented by the following identity:

$$[(1-\alpha)\mathbf{P} + \alpha\mathbf{v}\mathbf{e}']\mathbf{x} = \mathbf{x} \tag{B1}$$

For our scope, we discuss the correspondence between elements in (B1) and their counterparts in the *Input Rank* (see Equation 11), in light of the peculiar economic process at stake:

- In the *PageRank*, a transition matrix **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 use the matrix of an Input-Output table, $\mathbf{G}$, whose single elements are column-normalized buyer-supplier transactions, $g_{hk} \in [0,\ 1]$.

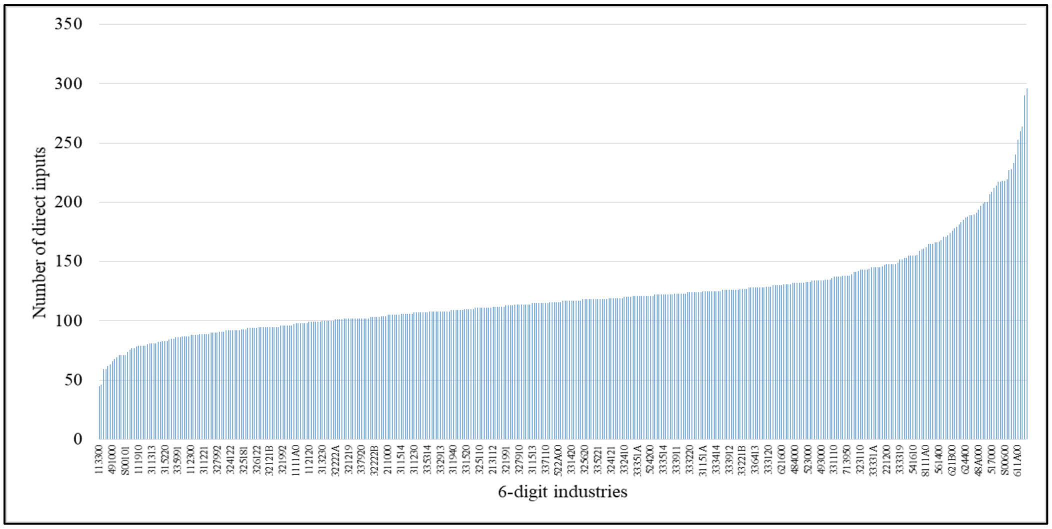
- A vector $\mathbf{v}$ is a critical tool that allows for the 'personalization' of the *PageRank*. In the absence of 'personalization', this vector contains just a uniform distribution of probability across all web pages, which is valid for all users. A 'personalization' entails a non-uniform distribution of probabilities, such that a region of the web will be more likely visited. At the same time, the vector $\mathbf{e}$ is a unitary vector that algebraically extends the same (uniform or non-uniform) distribution in $\mathbf{v}$ across web users. In our *Input Rank*, we can think of the combination of $\mathbf{v}$ and $\mathbf{e}$ as a specific portion of the production network, where the supply network of a root producer can be found. In a nutshell, in the case of the Input Rank, the representative producer in an output market will be able to explore only the technology of its supply network.

- The term $\alpha \in (0, 1)$ is a *teleportation* parameter in the *PageRank*, otherwise called a damping factor. It indicates the probability that a 'web surfer' interrupts 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*, we substitute α with the product of two parameters, δ_k and χ_k. The first indicates the relative usage of intermediate inputs, as from the Cobb-Douglas production function of Equation (5). The second indicates the ability of the producer to navigate her supply network, as from Definition 2.

- Finally, $\mathbf{x}$ is the solution to the eigenvalue problem in (B1). In the case of the *PageRank*, it indicates the relevance of the web content for each user, as for example after a query from a search engine. In the *Input Rank*, a vector $\boldsymbol{v_k}$ represents an ordering of all direct and indirect inputs based on the impact that any upstream friction may have on the marginal costs of a representative firm in the *k*-th industry.

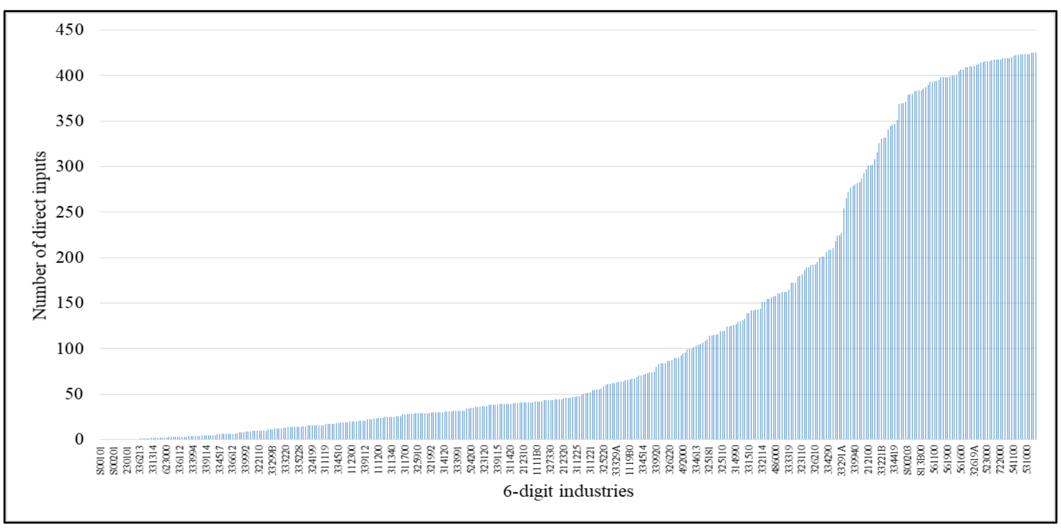
C – Appendix Tables and Graphs

Figure C1: 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 C2: 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 C1: 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 C2: 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 C3: 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

Table C4: Robustness to sample composition considering *midstream* parents only, parent-level fixed effects conditional logit

Dependent variable: **Input is integrated ==1**	No horizon-tal	Only indirect inputs	Output minus input elast	Top 100 in-puts
Input Rank	0.092***	0.180***	0.173***	0.149***
	(0.033)	(0.036)	(0.035)	(0.037)
Input Rank * Complements	0.131***	0.092***	0.190***	0.234***
	(0.050)	(0.048)	(0.048)	(0.048)
Input upstreamness	-1.015**	-1.197	-0.794	-1.618***
	(0.515)	(0.765)	(0.547)	(0.116)
Input upstreamness * Complements	0.792	1.112	0.385***	1.904***
	(0.582)	(0.977)	(0.062)	(0.132)
Contractibility	-0.191***	-0.163***	-0.271***	-0.276***
	(0.030)	(0.024)	(0.025)	(0.029)
Input Network distance	0.116	0.114	0.115	0.125
	(0.128)	(0.124)	(0.124)	(0.222)
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	0.201	0.289	0.314	0.356
Clustered errors by parent	Yes	Yes	Yes	Yes
Activity of parent companies	Intermediate goods	Intermediate goods	Intermediate goods	Intermediate goods

Note: The Input Rank is estimated with a damping factor equal to 0.5. Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table C5: Robustness to changing empirical strategy

Dependent variable: **Input is integrated ==1**	Linear probability model	Logit	Probit
Input Rank	0.012***	0.151***	0.074***
	(0.001)	(0.005)	(0.002)
Input Rank * Complements	0.084***	0.222***	0.119***
	(0.011)	(0.014)	(0.012)
Input upstreamness	-0.003**	-0.766***	-0.243***
	(0.002)	(0.354)	(0.029)
Input upstreamness * Complements	-0.001	-0.215	-0.210**
	(0.001)	(0.119)	(0.099)
Contractibility	-0.001***	-0.354***	-0.141***
	(0.001)	(0.017)	(0.007)
Input Network distance	0.051	-0.098*	-0.145
	(0.145)	(0.051)	(0.178)
Direct requirement	0.015	0.018***	0.011***
	(0.004)	(0.004)	(0.001)
Constant	0.005***	-4.824***	-.789***
	(0.001)	(0.035)	(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
Clustered errors by parent	Yes	Yes	Yes
Activity of parent companies	Manufacturing	Manufacturing	Manufacturing

Note: The Input Rank is estimated with a damping factor equal to 0.5. Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.

Table C6: Input Rank with sector-specific damping factors

Dependent variable: Input is integrated ==1	(1)	(2)	(3)	(4)
Input Rank	0.506***	0.274***	0.199***	0.091***
	(0.013)	(0.025)	(0.015)	(0.008)
Input Rank * Complements		0.101***	0.106***	0.158***
		(0.036)	(0.028)	(0.058)
Input upstreamness		-0.309	-0.303***	-0.044
		(0.215)	(0.040)	(0.418)
Input upstreamness * Complements		0.036	-0.155	0.255***
		(0.030)	(0.130)	(0.068)
Input Network Distance	-0.088***	-0.132	-0.121	-0.105
	(0.030)	(0.068)	(0.309)	(0.095)
Direct requirement	0.094***	0.012***	0.010***	0.026
	(0.013)	(0.003)	(0.002)	(0.034)
Observations	8,564,068	1,151,908	595,218	542,872
N. parent companies	20,294	4,084	2,110	1,925
Pseudo R-squared	0.548	0.759	0.701	0.815
Clustered errors by parent	Yes	Yes	Yes	Yes
Activity of parent companies	All	Manufacturing	Final goods	Intermediate goods

Note: The Input Rank is estimated with a damping factor equal to input contractibility à la Rauch (1999), specific for each input industry. Errors clustered by parent in parentheses. ***, **, * stand for p-value < 0.01, p-value < 0.05 and p-value < 0.10, respectively.