MPRA
Munich Personal RePEc Archive

Products as Network: An Empirical Approximation of the Manufacturing Production Network in Indonesia

Massie, Natanael Waraney Gerald and Mangunsong, Carlos

Faculty of Economics and Business, Universitas Indonesia, DTS Indonesia

5 August 2022

Online at https://mpra.ub.uni-muenchen.de/114647/
MPRA Paper No. 114647, posted 22 Sep 2022 13:24 UTC

Products as Network: An Empirical Approximation of the Manufacturing Production Network in Indonesia

Natanael Waraney Gerald Massie
Faculty of Economics and Business, Universitas Indonesia
natanaelmassie2009@gmail.com

Carlos Mangunsong
Faculty of Economics and Business, Universitas Indonesia
alos_mangunsong@yahoo.com

August 2022

Abstract

This study aims to characterise and represent the Indonesian manufacturing sector as a production network. We specifically define any relationship between any two products in the network as a relationship that one product is used as input to produce the other, akin to the input-output models but in a much-disaggregated level of 10-digit product level. This study utilises the Indonesian annual survey of manufacturing firms, specifically the 2017 data, to construct a product-level network of industries. Using the constructed network, this study discusses which products and sectors in the Indonesian manufacturing sector are more well-connected to others, using different centrality measures commonly discussed in network theory. We find that, generally, low-to-medium technology products are the more central products in Indonesian manufacturing. We also compare our framework with other well-established product network frameworks and discuss possible further works using our framework.

1 Introduction

1.1 Background

Recently, as the "product-as-networks" and "sectors-as-networks" frameworks have been growing in the body of literature, the ability to successfully represent a country's (or region's, or the world's) economy as networks allows for a wide range of calculations to be estimated – and hence more policy questions to be addressed. Studies using network-based models have asked questions from the age-old question on economic multipliers (see W. Leontief, 1987; Richardson, 1985) to the more recent inquiries on, for instance, which product should a country transition into being more competitive at (see Hidalgo et al., 2007; Hidalgo and Hausmann, 2008), how to characterize a country's production landscape (see Hausmann and Klinger, 2008b; De La Cruz and Riker, 2012; Hausmann and Klinger, 2008a), as well as characterizing the cascading effects of idiosyncratic microeconomic shocks on the aggregate fluctuations (see Acemoglu et al., 2012).

Considering the above, this study aims to characterize the Indonesian economy's manufacturing sector as a production network. While the country's statistics bureau (BPS / Statistics Indonesia) have officially released the official input-output statistics[1], and that numerous studies have even integrated the Indonesian input-output tables to an interregional or world-level models (see Bartelme and Gorodnichenko, 2015), this study takes a step back and asks the question: *can we go more granular than the sectoral level*?

As we know, the input-output (IO) based models are aggregated at the sectoral level. The most granular view of the IO tables in Indonesia is the 185-sector version[2]. While the above model is useful and has indeed been the cornerstone to numerous economic modeling and estimations across the years, this study argues that we can do one even better. Specifically, the manufacturing sector of the Indonesian economy *can* be disaggregated to its 10-level product level of input-output linkages, but not perfectly[3].

We argue that a successful representation of product-as-networks for the Indonesian economy can lead to novel and more granular questions being answered. For the IO-based models, trivially, the disaggregation of manufacturing products allows the analysis to be conducted at a more targeted level. However, for the competitiveness-based models (for instance, the Product Space models by Hidalgo et al., 2007), the disaggregation of the manufacturing sector into a product-level network might allow the two networks to be compared in one integrated analysis.

For instance, following the PS model, suppose some product x is calculated to be one of the better products for Indonesia to transition into and attempt to be competitive. However, without the product-level network, one might struggle to thoroughly analyze whether that product x's upstream industries are well-established. Trivially, importing the upstream products is a completely feasible option; however, the ability to readily query the upstream industry on a product-level IO network is only beneficial to policymaking.

[1]See the official statistics released by BPS.

[2]See the official statistics released by BPS.

[3]This condition is caused by the potential "over-inclusion" problem that will be discussed in Section 2.

This study aims to start the discussion on such a disaggregation and provide an approximate look at how the 10-digit product-level manufacturing sector input-output linkages are constructed in Indonesia. We identify our framework to be akin to the IO-based models of Indonesia, but a limitation[4] in data shape and availability hinders this study from fully deriving the A matrix in the IO model (the intersectoral $n \times n$ matrix). As such, this study is explicitly (and carefully) claiming only to be an *approximation* of a manufacturing-only 10-digit product-level *unweighted* input-output network for Indonesia.

In doing so, this study utilizes the firm-level micro-data from the Indonesian manufacturing sector, released annually by BPS. Each year, BPS releases one main dataset containing plant-level characteristics; however, our main focus in the study is exploring two closely-related datasets that are unique at the product level and not the plant level. The first dataset provides information on the raw/intermediate input used by each firm surveyed in the primary dataset, commonly referred to as *Rawin* dataset. Meanwhile, the second dataset provides information on the other side, namely, the products that are produced by each firm surveyed in the primary dataset, commonly referred to as the *Proin* dataset[5].

As one might see, combining the *Rawin* and *Proin* datasets can provide a network of which products are the upstream products (ancestor node in the graph) of some other products and, analogously, which products are the downstream product (child node in the graph) of some other products. The study borrows one of the important notions in the computer science literature, specifically the graph traversal methodologies from network theory. The literature draws back to the early seminal paper on graphs (see, for instance, Zuse, 1972; Moore, 1959) that introduces how one should approach a traversal of graphs and thus construct the whole network.

This study is structured as follows. In the current introductory section, Section 1, the study provides the relevance of the study, as well as how it compares and can be integrated into the other existing frameworks. In the following section, Section 2, the study provides the algorithm of the whole network graph generation in its first subsection. We describe the data source in the second subsection. In the third subsection, we address the potential shortcomings of the dataset and algorithm. This study argues that the handling methods of the potentially erroneous data and algorithm allow the overall framework to be used in various settings and differing degrees of data quality − a typical land-

[4]This condition is caused by the potential "over-inclusion" problem that will be discussed in Section 2.

[5]At the time of this writing, the two datasets are now unable to be officially acquired by purchasing from BPS-Statistics Indonesia, as now the institution only provides the primary data. The data used in this study was used in past studies/research projects and was acquired by authors before BPS started providing only the primary dataset.

scape in using empirical, survey-based datasets. We introduce the concept of a "strong edge" and "weak edge" to address one of the potential shortcomings of the two datasets – *Proin* and *Rawin* – which can only be joined at the firm level.

Section 3 describes the generated network. Apart from the network creation for the whole Indonesian manufacturing sector, this study also provides several critical applications that can be derived after obtaining the initial network. Finally, the last concluding section, Section 4, describes the study's limitations and potential ways forward in expanding the framework and integrating it into the existing economic frameworks.

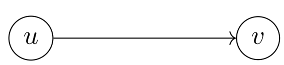
2 Model

2.1 Setup

This study defines a network by its colloquial definition; that is, a network G is defined as a pair of sets that are comprised of a set of vertices V and set of edges E, such that $G = (V, E)$ where every edge $e \in E$, any two nodes v_1, v_2 that it connects are members of the set V.

In the context of the study, any vertex v in the set V represents one product at the 10-digit product code level, either used as an input or produced as an output by some firm f in the set of firms F. The graph that the study utilizes is the directional graph (digraph), where for any edge e denoting an ancestor-children relationship between two nodes, e.g., $u, v \in V$ such that $e = (u, v)$ it is the case that in the Indonesian manufacturing survey dataset, some firm $f \in F$ uses u as an input in the *Rawin* dataset (input dataset) and produces v as an output in the *Proin* dataset (output dataset). A simple illustration is shown in Figure 1.

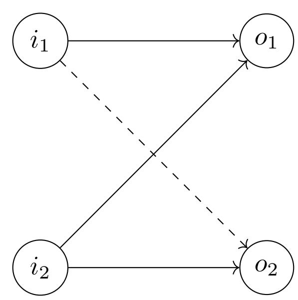
Figure 1: *Illustration of one downstream linkage of product u being used as input to produce product v*



The algorithm used to produce the complete network in this study is by using an ensemble of the standard graph traversal methodology. First, the study creates the set of edges E using the data available at the firm level. For any firm $f \in F$, we define the set of edges E_f relevant for the firm by creating a Cartesian product of any input i in the input set $I_f \subseteq V$ that the firm uses with any output o in the output set $O_f \subseteq V$ that the firm produces. By aggregating the set of edges for all firms $E_1, E_2, ..., E_n$, the study obtains the full network set of edges E.

While the above exercise might sound trivial, at the time of writing, the only join keys between the *Rawin* (input) dataset and *Proin* (output) dataset are the firm IDs[6]. This means that one can only *approximate* the network, as the information on how exactly each input is utilized in producing which output is unclear. One obvious problem that can be potentially spotted in the process is what the study refers to as the "over-inclusion" of the ancestor-child relationship. Consider a case where a firm f produces two outputs $o_1, o_2 \in V$, where it uses inputs $i_1, i_2 \in V$ both to produce o_1, but only input i_2 to produce o_2. Such a case means that the valid edges are $(i_1, o_1), (i_2, o_1), (i_2, o_2)$. However, as the method assumes a Cartesian product to create the edges, (i_1, o_2) is mistakenly included. Cases with higher cardinality of products in the input set and output set analogously follow from the simple case.

Figure 2: *Illustration of the potential "over-inclusion" in product network*



The (approximate) handling of the above problem will be more thoroughly discussed in one of the following sections that describe the study's specific handling of such a condition using what we refer to as the "strong" and "weak" edges. As one might note, the analogous problem of under-inclusion might also occur, but such a case is inevitably outside the scope of the study as it constitutes dealing with the non-existence of relationships in our data. This condition also becomes the rationale as to why this study limits its scope to creating an unweighted network. That is, the study disregards the value-added flow among products as we deem mistaken flow of value-added is more severe (especially considering cascading effects in networks) in the analysis compared to inaccurate determination of existence or non-existence of linkages between products.

As a base for creating the complete graph, the study uses the total number of firms having an ancestor-child relationship of any two products (u, v) as the weight of the edges. Formally, we define the weight w of some edge e as we as follow.

[6]This condition is a case where both the input dataset and output dataset have a product-level unit of analysis, while the joining keys between the two are at firm-level, which is more aggregated than product-level. One can only see, for instance, an arbitrary firm $f \in F$ uses inputs $i_1, i_2, ..., i_n \in V$ as inputs; and separately, the firm $f \in F$ produces outputs $o_1, o_2, ..., o_m \in V$ with no linking at product-level.

$$w_e = \sum_{f \in F} sign(E_f)$$

where:

$$sign(E_f) = \begin{cases} 0, & \text{if } (u, v) \notin E_f \\ 1, & \text{otherwise} \end{cases}$$

The study calculates the weight of each edge by taking the summation of sign functions for every firm $f \in F$ on whether the directional relationship between the vertices, for instance, (u, v) forming the edge e is present in the firm's set of edges E_f. The sign function takes the value of 1 if the edge is present and 0 otherwise.

Now that all the assumptions and definitions are laid out, we proceed to discuss the complete network creation algorithm. To create a complete network, ideally, one iteration of a pair of Breadth-First Search (BFS)[7] will be conducted from each node. The pair of BFS algorithms are simple traversals starting from every node $v \in V$, with one being conducted directionally to the upstream products and the other being conducted in the downstream direction.

Combining the two processes, the complete industry tree in the study can be constructed. We combine the two output edge lists for any starting node $v \in V$. The product of the algorithm, i.e., the complete network of industries, will be provided in Section 3, along with other analyses that can be derived from the complete graph.

2.2 Data Source

The study uses the Indonesian firm-level micro-data survey released by the statistics bureau, Statistics Indonesia (or *Badan Pusat Statistik* / BPS), on a yearly basis referred to as the *Statistik Industri* (SI) or *Industri Besar Sedang* (IBS) dataset. More specifically, the study utilizes the complimentary datasets from the surveys, namely the product-level input dataset used by each firm named the *Rawin* dataset and the product-level output dataset produced by each firm named the *Proin* dataset.

For demonstration purposes, this study uses the SI (the *Proin* and *Rawin* included) year 2017 as the primary dataset in this study. The specific dataset

[7] Breadth-First Search or BFS is one of the ubiquitous methods in conducting graph traversal in network theory. By doing BFS from some starting node $v \in V$, the user can exhaustively obtain the list of all connected (and remotely connected) nodes to v (for instance, see Cormen et al., 2009).

is used based on two main reasons. First, the dataset is the latest available. Second, the dataset is one of the first to adopt the Indonesian standardized commodity-level product code, KBKI (*Klasifikasi Baku Komoditas Indonesia*) 2015 on both the input (*Rawin*) and output (*Proin*) datasets. The KBKI codes are detailed to the tenth digit, one of the most granular data available in the Indonesian context.

2.3 Handling Over-inclusion: Weak Edge and Strong Edge

Related to the prior discussion in Section 2.1 on how the inherent shape of the datasets can potentially induce the over-inclusion problem, this section describes the suggested handling method. We use the notion of a "strong" and "weak" edge. The formal definition of a strong industrial linkage edge (SILE) is as follows.

Definition 1 *A strong industrial linkage edge is defined as some edge $e \in E$ from the industrial linkage graph such that its weight $w_e \geq w_t$ where w_t is some cutoff threshold weight that determines pruning.*

In the above definition, we define any upstream/downstream linkage whose weight is lower than some arbitrarily-chosen cutoff weight w_t is considered as a potentially irrelevant and/or mistaken connection. However, edges with a weak linkage might very well exist. We formally define such an occurrence as a weak industrial linkage edge (WILE) as follows.

Definition 2 *A weak industrial linkage edge is defined as some edge $e \in E$ from the pruned industrial linkage graph such that its weight $w_e < w_t$ where w_t is some cutoff threshold weight that determines pruning and w_e is non-zero.*

Trivially, the last possibility will be a non-existent edge (weight is zero), and this is excluded as a WILE, as if the connection has a weight of zero, then no linkage exists.

Given the two definitions above, any complete industrial network provided in further sections of this study will have some threshold weight. We account for the possible irrelevant or mistaken edges using such measure, and as such, any further analysis will suffer from the inherent limitations of any used threshold weight. In the following Table 1, this study summarizes the sensitivity of different cutoff weights.

Table 1: **Network shape by different threshold values**

No	Threshold	Number of products (nodes)	Number of industrial linkage (edges)
1	5	1,161	3,948
2	10	926	2,738
3	25	642	1,579
4	50	414	819
5	100	231	399

Source: Statistics Indonesia-BPS (2017)

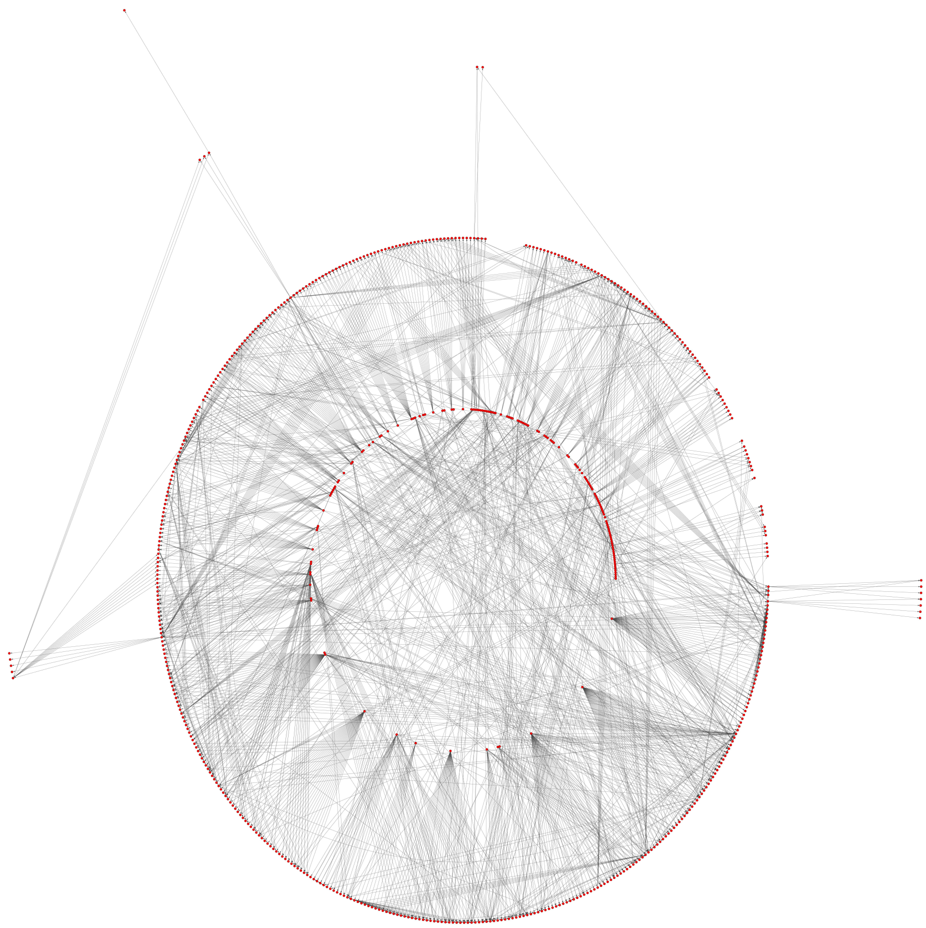
Note: The table above presents the number of products (nodes in a graph) and industrial linkages (edges in a graph) as a result of the network creation, based on the different possible thresholds in defining weak or strong edges. Note that the number of edges presented in the table omits the non-SILE edges. The selection of thresholds presented above is arbitrary.

3 Results & Discussions

3.1 The Complete Network

This first subsection provides the baseline result of the industrial network. The specification the study uses is a threshold weight w_t of 50 for an edge to be considered a strong edge. More simply put, for a product $u \in V$ to be considered as an input to the output product $v \in V$, there must be at least 50 firms reporting the use of u in the *Rawin* dataset to produce v in the *Proin* dataset, or more formally forming an edge (u, v) in the graph. A visual representation of the complete network is shown in Figure 3.

Figure 3: **Complete industrial network graph of the Indonesian manufacturing sector in 2017**



Source: Statistics Indonesia-BPS (2017); network is calculated by authors

Note: The figure above displays the complete network of the Indonesian manufacturing sector. Each red dot depicts one product, represented as one vertex in the graph. Each arrow between nodes represents a "downstream linkage" from the source node to the target node, formally defined as a relationship in which the source node (product) contributes as an input to the production of the target node (product). Only strong edges are shown, i.e., edges with less than a threshold of 50 firms reporting are excluded. Edges are drawn without any scaling by each edge's weight. High precision figure is used in the figure, allowing high-detail zoom-ins to be conducted.

Visually, one can observe that there are roughly four layers of product linkages from the outermost layers to the innermost, where the linkages are in high cardinality. The observation is close, but, allowing potential linkages between the same layers, the study finds that there are products involved in a production chain of 6 products (by the highest eccentricity in Table 2), although such an occurrence is virtually idiosyncratic.

3.2 Summary statistics of the complete network

Besides such observation, Figure 3 is far too high in dimensionality to be interpreted. Taking a more general approach, several key summary statistics of the network's characteristics can be derived by iterating the nodes and edges (products and industrial linkages) in the graph. Table 2 provides some summary statistics of some of the ubiquitous characteristics of a graph, e.g., centrality measures, eccentricity, among others.

Table 2: **Product-level (node-level) summary statistics of the complete network**

Variable	**N**	**Mean**	**Std. deviation**	**Min.**	**Median**	**Max.**
Network characteristics						
Node indegrees *(Explains how many other products are used as an input for a particular product)*	353	2	2	0	2	9
Node outdegrees *(Explains how many other products uses a product as an input)*	353	1	4	0	0	35
Betweenness centrality *(Explains how "central" a product is by showing how many shortest paths among every other products passes through the node)*	353	6	39	0	0	490
Node PageRank *(Explains how "central" a product is by scoring each product based on their connections and the connections' connections)*	353	0.0016	0.0007	0.0012	0.0014	0.0083
Node eccentricity *(Explains the length of the longest shortest path starting from a node/product)*	353	0	1	0	0	6
Producing-firms characteristics						
Number of workers (persons)	353	231	618	20	156	11,428

(continued on the next page)

Table 2: **Product-level (node-level) summary statistics of the complete network**

Variable	**N**	**Mean**	**Std. deviation**	**Min.**	**Median**	**Max.**
Network characteristics						
Number of female workers (persons)	353	112	478	0	50	8,787
Number of male workers (persons)	353	119	158	10	97	2,641
Average value-added per worker (million IDR/worker)	353	0.6	1.0	0.01	0.3	13.9
Foreign ownership (%)	353	0.6	1.0	0.01	0.3	13.9
Wage of production workers (million IDR)	353	8.5	23.6	0.2	5.3	433.5
Wage of other workers (million IDR)	353	2.6	4.8	0	1.7	54.0

Source: Statistics Indonesia-BPS (2017)

Note: The table above displays the summary statistics of the products contained in the complete industrial network graph with threshold weight of 50 and product code depth at 10 digits of KBKI product codes. The summary statistics included are the count of non-missing observations, the arithmetic mean, standard deviation, minimum value, median value, and maximum value of each variable. Only products that are successfully joined between the firm-level datasets (that produces the "producing firms characteristics" calculations) and the product-level datasets (that produces the "network characteristics" calculations) are included.

Several general observations can be noted from Table 2. First, note that the number of observations found as a result of an inner-join between the product-level dataset (which shows the number of products/linkages on Table 1) and the firm-level dataset is now showing less number of products (414 products in Table 1 to 353 in Table 2). This indicates only 353 of the 414 products are reported to be domestically produced by Indonesian firms.

Further, this study finds that, on average, a product serves as an input for one other product, as explained by the "node outdegrees" characteristics. Meanwhile, on average, a product also uses two other products as input, illustrated by the "node indegrees" characteristics. Taking the higher extremes, some product(s) takes an input of as many as six other products, while some product(s) is used by 35 other products as input. Other centrality measures (e.g., betweenness, PageRank) and characteristics are also provided in the table. They will be discussed in greater detail further in the manuscript as the relevant discussions are presented.

3.3 Product-level snapshots

3.3.1 Simple applications on several policy questions

Next, we delve deeper at the product level. Specifically, this section aims to shed light on some ubiquitous policy questions related to an economy's manufacturing sector. Whether one is assessing policies in the context of aiming to improve the manufacturing sector's total value added, or in the context of considering which sector to support more prominently than others due to some idiosyncratic shock[8], one of the common question to arise is questions related to *which sector should we prioritize?*

The logic is simple: a sub-optimal decision in choosing the targeted industry might be costly in either of the two mentioned cases. We argue that much like the sectoral-level input-output where the total economic impact might be estimated by considering which sectors yield the highest economic multipliers (see, for instance, Poot, 1991; Zuhdi, 2015; Tui and Adachi, 2021), the product-level version of such multipliers can be used to pick the products (or industries) with the highest degree of connection to the whole economy, and thus avoid making the said sub-optimal decision.

While technically constructible (and trivially preferable), multiplier calculations for this model require the *weighted* network. This study demonstrates the usefulness of successfully representing an economy's manufacturing sector as networks – even its unweighted version – by arguing that one of the possible angles to approach such a question is by calculating the *"centrality"* aspect of the sectors/products, with respect to the whole network of products in the said economy. In network theory, centrality refers to metrics that indicate *"how well-connected"* a node is with respect to the whole network. The centrality calculations are applicable for both weighted and unweighted networks, thus fitting with one of the data limitations in this study as we only characterize the *unweighted* network of products.

Specifically, we will describe the products in the Indonesian manufacturing sector by delving into the centrality measures (see illustrations in Figure 4) that are exactly related to policy questions such as:

- Which products utilize the most number of other products as inputs?
- Which products are utilized the most by other products as inputs?
- Which products are the most connected with other products, with respect to the whole production network?

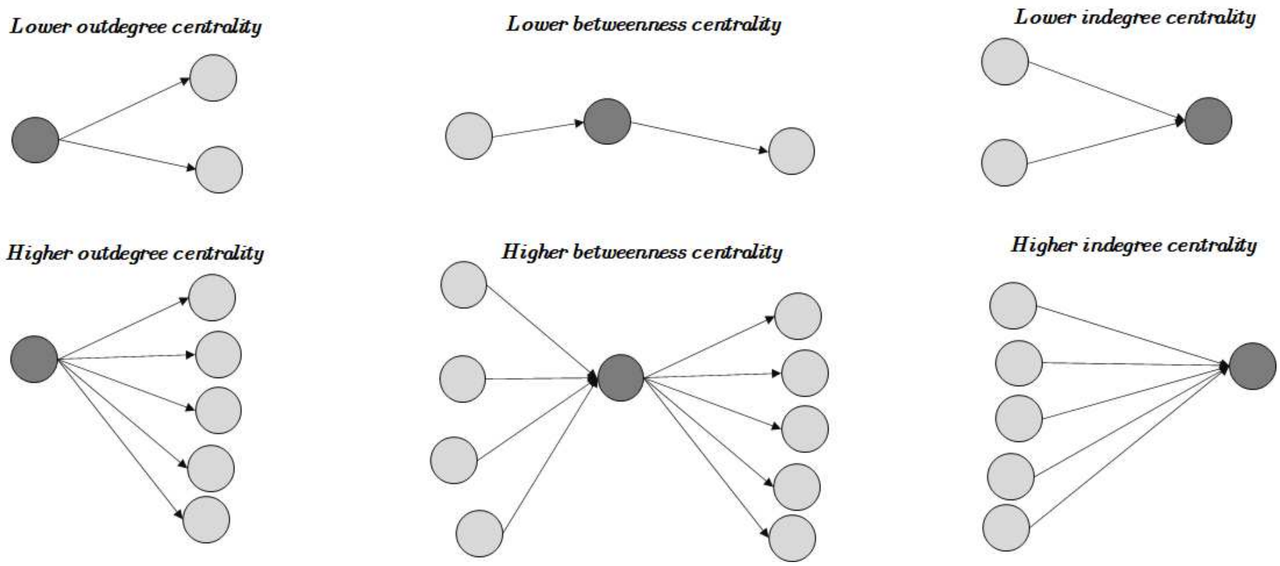
The first two questions will be addressed by describing the product-level *indegree* and *outdegree* centralities, as the two centrality measures address the

[8]Take, for instance, the recent example of COVID-19 pandemic which forced a considerable portion of the manufacturing sector in many economies to shut down.

questions quite nicely. The third question is answered quite differently: we use two different measures to answer the same question, each giving a different angle.

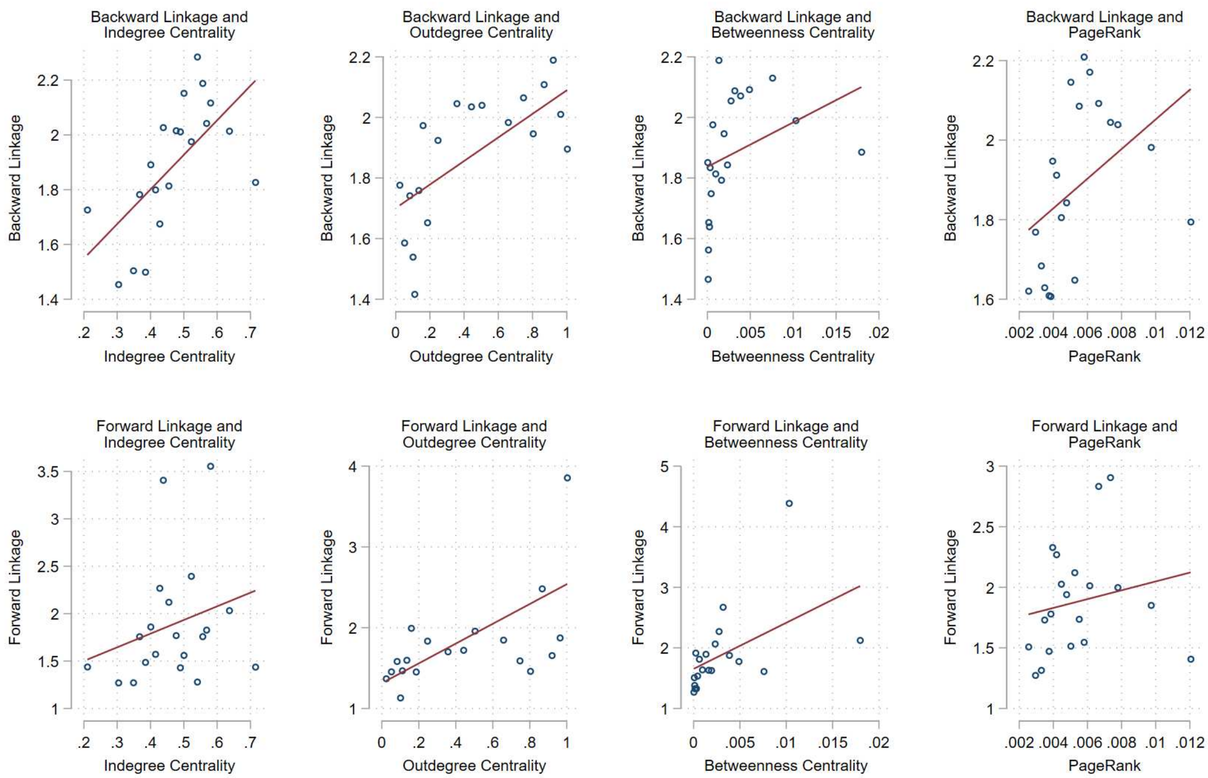
First, we try asking, "which product is the most *connecting* for the other products?" Such a question is addressed by discussing the *betweenness centrality* of the products. For the second approach, we try asking a unique question: "How complex and overreaching is the network of products *leading up to* a product?" This question is addressed by discussing the *pagerank* scores for the products.

Figure 4: **Illustration of different centrality measures**



While the detailed definition for each centrality measure is discussed in the respective sections, we conclude this argument that centrality measures can be used as proxies for the multiplier calculations in the case of our unweighted network by providing a visualization of relationships between the two on the data where both sets of calculations (e.g., centrality and multipliers) are available: the 185-sectors input-output table for Indonesia in 2016. Figure 5 plots the relationship between the two.

Figure 5: **Relationship between unweighted centrality measures and input-output multipliers, based on Indonesia's 2016 Input-Output Table**



Source: Statistics Indonesia-BPS (2016); calculated by authors
Note: The figure above presents the binscatter plot between two input output multipliers (e.g., backward and forward linkage) and four centrality measures (e.g., indegree, outdegree, betweenness, and pagerank) calculated from the 2016 Indonesian Input-Output table. Centrality measures are calculated by converting the input-output table to its unweighted representation to simulate this study's condition. Binscatter plot construction follows Stepner, 2013.

3.3.2 Products' indegrees: which products utilize the most inputs?

This study starts from a motivating question that states the following trivial question: which products utilize the most inputs in their production? Such a question might serve as an essential policy (and academic) question, especially in determining crucial policy options within the product space.

Addressing such a query, the first indicator discussed is the *indegree centrality* of each product: a node (product) characteristic that calculates how many other nodes (products) are the ancestor (used as an input) by the particular node (product). By definition, the indegree centrality $IC(u)$ of a node in a graph (product $u \in V$ in our context) can be defined (see, for instance, Hansen et al., 2011) as follow.

$$IC(u) = \sum_{v \neq u \in V} sign(v \longrightarrow u)$$

where:

$$sign(v \longrightarrow u) = \begin{cases} 1, & \text{if } (v,u) \in E \\ 0, & \text{otherwise} \end{cases}$$

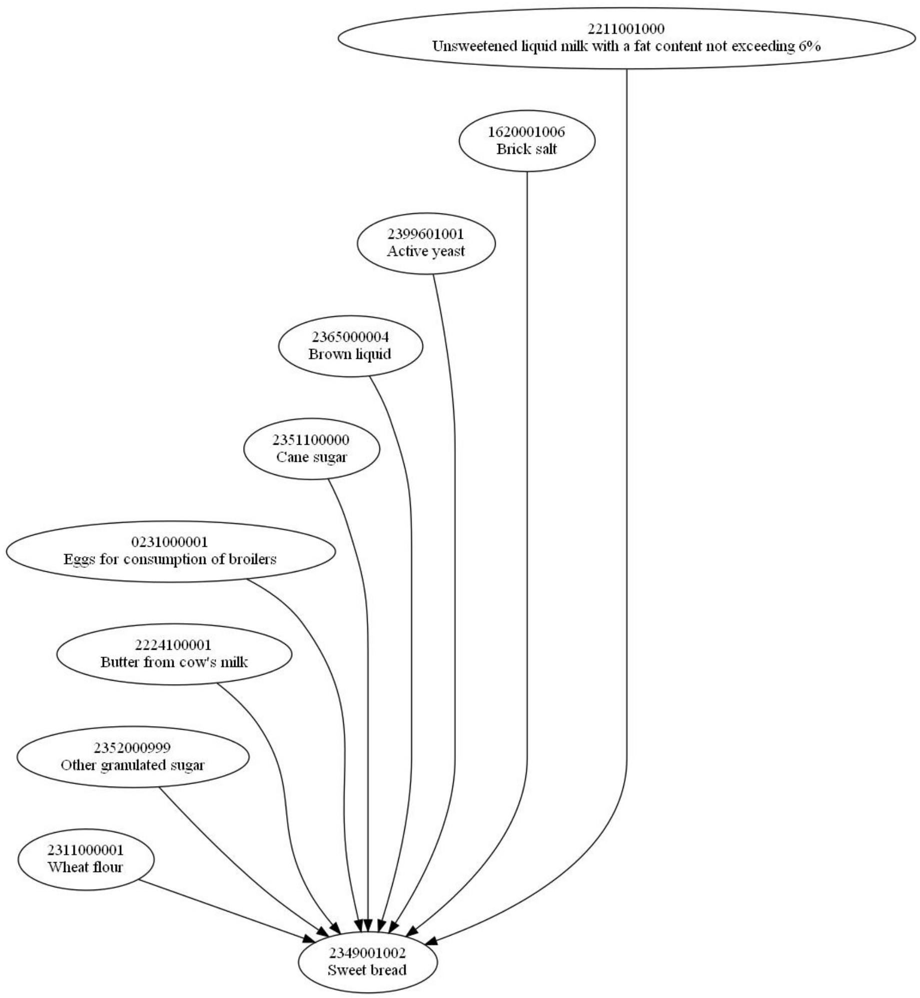
Essentially, the indegree centrality of a product (or node) $u \in V$ is simply the count of edges that starts from the other products $v \in V$ for any product $v \neq u$ and ends at the product u. This denotes the *usage of other products* by product u, thus enabling one to address the question of whether product u utilizes many other products or not.

The products at the 10-digits level that use the most number of inputs are displayed in Table 3. Considering the product-level results in Table 3, a more abstract view of the subject can also be studied. Table 4 presents the frequency of how many times each sector's (2-digit level) product (at the 10-digit level) is present in the top 50 products with the highest indegree.

A general observation would be that the top indegree sector is dominated by the basic-to-intermediate level of manufactured products, such as paddy and rice products, meat/fish products, tobacco products, and base metal. Such an observation might further support the notion that the Indonesian manufacturing sector is still skewed to the upstream and basic industries, as a sophisticated downstream industry should arguably lead to more complex, high-technology products (such as electronics) being located at the top-50 inputs used.

Let us look deeper into one of the 10-digit products mentioned in the Table. Specifically, let us take a look at how *"Sweet bread"* (product code *"2349001002"*) is created by seeing its downstream and upstream nodes both traversing by 1 stage of linkages in either direction. Figure 6 shows the illustration of the network.

Figure 6: **An illustration of *Sweet bread*'s immediate network**



Source: Statistics Indonesia-BPS (2017); network is calculated by authors
Note: The figure above presents the products (represented as nodes in a graph) with upstream and/or downstream linkages with *Sweet bread* (product code *2349001002*). The figure uses a network specification that seeks the central node (sweet bread)'s downstream and upstream nodes both traversing by 1 stage of linkages into either directions, with pruning threshold of 50. Note that the edges presented in the table omits the non-SILE edges.

Figure 6 shows how sweet bread is connected to a high number of other products as input, making it one of the products with the highest number of indegree centrality scores. Looking at the product descriptions of such related nodes, the inputs that our algorithm produces for sweet bread also generally pass a sense check: the algorithm outputs sugar, milk, flour, eggs, and yeast, among others, as inputs in creating a sweet bread.

Note that in this case, despite stating that this example aims to traverse by one stage toward either direction (both upstream or downstream) starting from sweet bread in the graph, we only have its upstream products and no downstream product. Such an illustration shows that, indeed, sweet bread is a final product. Thus, we find no other products using it as input. Analogously, the algorithm can be readily applied to other products as its centre node.

3.3.3 Products' outdegrees: which products are utilized by the most number of downstream products?

An analogous analysis regarding the outdegree characteristics can also be derived. A motivating question that states the following: *which products are utilized by the most number of downstream products?* That is, which products are used by other products as input the most, a relationship represented in a graph such that the product forms a directed edge toward the said other (more downstream) products. A similar − only reversed − calculation compared to indegree, the outdegree centrality $OC(u)$ of a node in a graph (product $u \in V$ in our context) can be defined (see, for instance, Hansen et al., 2011) as follows.

$$OC(u) = \sum_{v \neq u \in V} sign(u \longrightarrow v)$$

where:

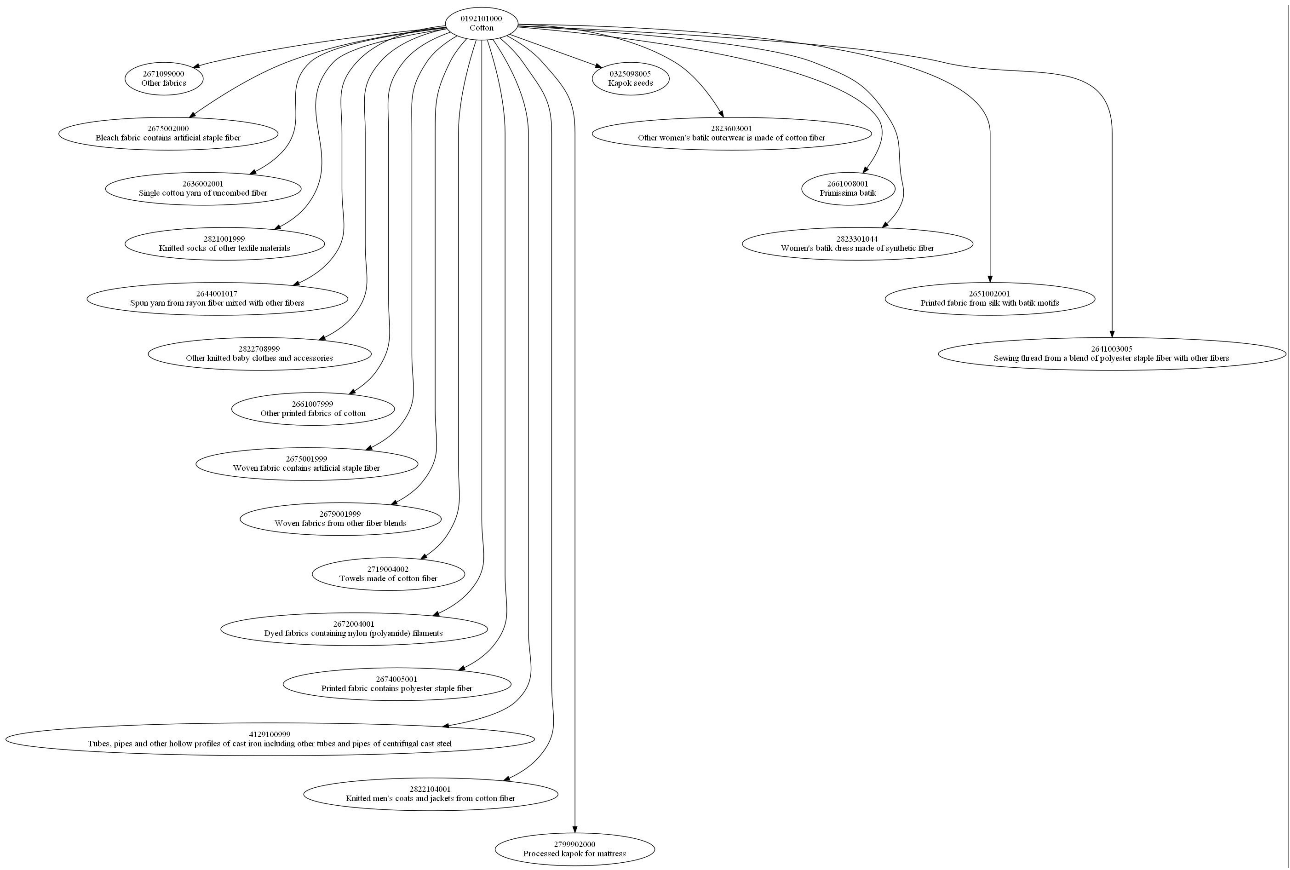
$$sign(u \longrightarrow v) = \begin{cases} 1, & \text{if } (u, v) \in E \\ 0, & \text{otherwise} \end{cases}$$

Similar to the indegree case, the outdegree centrality of the product (or node) $u \in V$ is the count of edges that starts from product u and ends at some other product $v \in V$ for any $v \neq u$. This denotes the product u's *usage by other products*, allowing one to address the question of whether product u is utilized by many other products or not. The products with the highest outdegree centrality scores are shown in Table 5.

A general observation from Table 5 on the 10-digit product-level highest outdegree is that, unsurprisingly, primary products still dominate. Chemical, plantation and basic manufacturing products are used by tens of other downstream products. Table 6 further presents the 2-digit level aggregation, and the 2-digit level observation suggests a similar story to the 10-digit level discussion. Agriculture, chemical, and extraction products have a high degree of usage by other products in the downstream processes.

Next, let us see a snapshot of one of the highly-utilized upstream products in the production network, namely *Cotton* (product code *"0192101000"*). Analogous to the previously-shown snapshot of Sweet bread's upstream and downstream neighbours, Figure 7 shows a similarly-constructed tree for Cotton.

Figure 7: **An illustration of *Cotton*'s immediate network**



Source: Statistics Indonesia-BPS (2017); network is calculated by authors
Note: The figure above presents the products (represented as nodes in a graph) with upstream and/or downstream linkages with Cotton (product code 0192101000). The figure uses a network specification that seeks the central node (cotton)'s downstream and upstream nodes both traversing by 1 stage of linkages into either directions, with pruning threshold of 50. Note that the edges presented in the table omits the non-SILE edges. High-precision fonts are used in the figure, detailed zoom-ins are enabled for reading.

Figure 7 shows how cotton is well-connected, this time as an input to a high number of other downstream products, making it one of the products with the highest number of outdegree centrality scores. This study argues that also, by evaluating the related products, the network created by the algorithm again passes the sense check. Note that products such as fabrics, knitted socks, *batik*[9] cloths, among others, are listed as its downstream products.

Analogous to the sweet bread's case, this example aims to also traverse one stage in either direction (both upstream or downstream) starting from cotton in the graph. However, as cotton is one of the most upstream products, the figure only has its downstream products and no upstream product.

3.3.4 Products' betweenness: which products connect the most product chains?

While the previous discussions regarding the indegree and outdegree measures have shed light on the patterns revolving around the Indonesian manufacturing products, such snapshots can only arguably show "one side" of a production.

For instance, products with a high number of indegree might show that such a product uses many upstream products but does not say anything about whether that product is then, in turn, used by many others. On the other hand, products with a high number of outdegree might show that such a product is used by many downstream products but does not say how many other products such a product used to be produced.

If anything, both indicators might suggest good "upstream" (based on outdegree) and "downstream" (based on indegree) products, although imperfectly[10]. But, what if a policy question arises on which products use many inputs and are also used by many others as inputs? One might want to combine indegree and outdegree in some formal notion to investigate this kind of good "intermediate" or "connecting" products.

This study uses the notion of node betweenness to shed some light on this discussion. Technically, some node $b \in V$ is said to be "between" some other nodes $a, c \in V$ if node b is located in the shortest path between node a and c. Translating such a notion into this study's context, then we can say that, for instance, if cotton is used as an input to produce threads, threads are used to produce cloth, and then the cloth is used as an input to produce a t-shirt, then, cotton is "between" the cotton and the t-shirt.

The notion of betweenness comes with the aggregation of many shortest paths of some other products passing to a particular product x. If a product is between a high number of paths between other products, then one can argue

[9] *Batik* is one of the traditional Indonesian shirts.

[10] Note that a high number of indegree or outdegree might occur at any stage of a production network (except source nodes for indegree or sink nodes for outdegree).

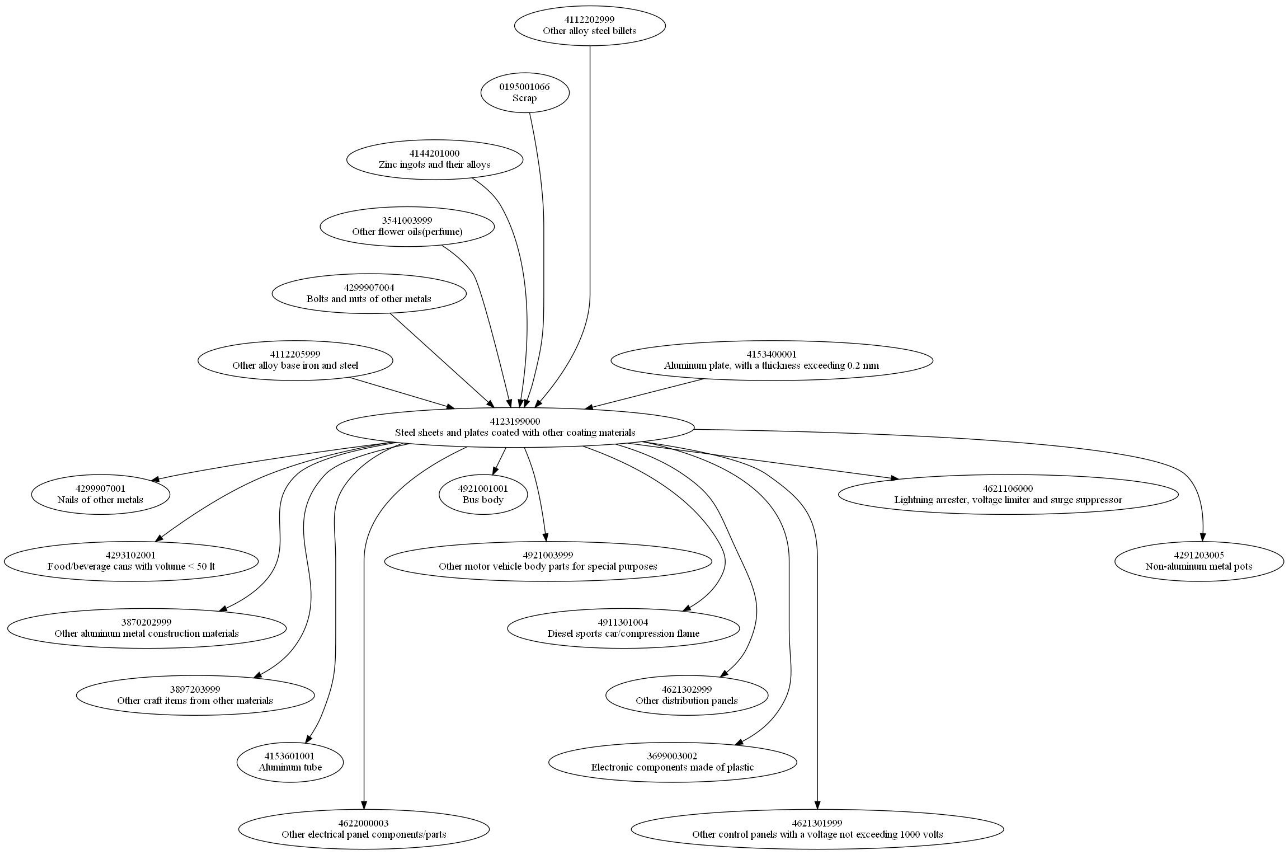
that such a product is a crucial intermediate product used in the production network. By definition, the betweenness centrality $BC(u)$ of a node in a graph (product $u \in V$ in our context) can be defined (see, for instance, Perez and Germon, 2016) as follow.

$$BC(u) = \sum_{s \neq u \neq t} \frac{\sigma_{s,t}(u)}{\sigma_{s,t}}$$

In the above equation, $BC(u)$ is calculated by taking a summation of the ratio between the count of shortest paths between any other two pairs of products $s, t \in V$ that is not u that passes through u, i.e., $\sigma_{s,t}(u)$ and the count of total shortest paths between the two products themselves s, t (regardless of whether the paths pass through u or not), i.e., $\sigma_{s,t}$. This denotes how product u *connects other products*, allowing one to address the question of interest. Table 7 presents the products (10-digit level) with the highest betweenness centrality scores.

One general observation from the above table is the domination of metal products in Indonesian production chains. We delve into the 2-digit level aggregation to check whether the same story remains in Table 8. In Table 8, it is shown that a similar story remains. Base metal products top the list, with chemical, food, and electrical products closely following.

Figure 8: **An illustration of *Steel sheet and other coating materials*'s immediate network**



Source: Statistics Indonesia-BPS (2017); network is calculated by authors
Note: The figure above presents the products (represented as nodes in a graph) with upstream and/or downstream linkages with *Steel sheet and other coating materials* (product code *4123199000*). The figure uses a network specification that seeks the central node (*Steel sheet and other coating materials*)'s downstream and upstream nodes both traversing by 1 stage of linkages into either directions, with pruning threshold of 50. Note that the edges presented in the table omits the non-SILE edges. High-precision fonts are used in the figure, detailed zoom-ins are enabled for reading.

Equipped with the betweenness centrality scores, let us now discuss another product-level snapshot, this time with one of the highest betweenness centrality scores. Figure 8 presents the immediate network (i.e., 1-stage upstream and downstream linkage) of *"Steel sheet and other coating materials"* (product code *4123199000*). In Figure 8, *steel sheet and other coating materials* is depicted as a well-connected product in the production network, in line with the high score of betweenness centrality provided in Table 7. By evaluating the descriptions of the related products, the study argues that the network produced by the algorithm also shows a sensible result.

3.3.5 Products' PageRank: which products are connected to quality products?

Suppose we have two products $a, b \in V$ with similarly high centrality scores. Our previous discussion suggests that, in that case, both products a and b are well-connected as products. This study attempts to take the previous questions into an even higher level of discussion: what if we consider the importance of the products' connections?

This study argues that such a question can be answered by solving for the popular algorithm developed by Google's founders (Page et al., 1999), widely referred to as the "PageRank" scores of each product (each node in our graph). In layman's terms, the PageRank scores of each node can be used to determine the "importance" of a node as it takes into account not only the nodes' connections but also the quality of such nodes' connections. Such a calculation is instrumental in Google's aim of providing relevant and accurate web-search results, but what if we apply such a calculation to this study's context: an industrial tree of product networks?

Let us describe the "importance" of some products in this manufacturing sector context. Suppose for every product p, it uses many inputs $p_1^1, p_2^1, ..., p_n^1$. Then, each of the inputs uses a varying number of other inputs themselves, say, $p_1^2, p_2^2, ..., p_n^2$. The process continues upwards iteratively. One can thus argue that a product that utilizes many inputs that utilize many (second order) inputs (and so on) is more "important" than some other product that uses few inputs that utilizes few (second order) inputs (and so on). The reason is that the former is *relied upon by* many other products, as if that said product is discontinued to be produced in the country, many other products will suffer.

Consider also that the linkages between product p and its inputs, and its inputs' inputs, and so forth might not be so straightforward. Imagine a simple case of a production chain between cotton, threads, fabric, and t-shirts. We know that cotton can be used to produce threads, threads as input to produce cloths, and cloths as input to produce t-shirts. In our network, here we have edges between products such as the path (cotton $\longrightarrow$ thread $\longrightarrow$ cloth $\longrightarrow$ t-shirt). However, what if some t-shirts are designed to have some cotton accents on them? Then, there will also be some edge (cotton $\longrightarrow$ t-shirt). Analogously, what if some t-shirts have nice shapes made out of threads?

This means that, for some product to be *relied upon by* many other products, we will want its network of input products *leading up to that product* to be as complex and overreaching as possible. We argue that this question can be answered by considering the PageRank scores of each node.

In calculating PageRank, one must solve the calculation on how likely a random walk in a network would end on each product. While the notion is highly technical, one can more easily imagine the problem as playing in a *maze* (not necessarily solving it). Consider the case where some corners of the labyrinth have more roads leading to it. Then, trivially, a confused maze player would more likely arrive at that corner rather than some other corner that has fewer roads leading to it[11]. Imagining each corner of the maze as a node in a graph, and the roads between the corners as the edges, then, a corner with a higher PageRank represents the higher likelihood that some maze player arrives (and ends her journey) at that corner. Can we say the same about products?

Much like the case of that maze, imagine each corner as one product and each road as one input-output linkage (this time around, the roads might not always be reversible, as input-output linkages are more one-directional rather than bi-directional). Then, products with higher PageRank scores would correspond to products where *more roads* lead to them, and thus more reliance by other products (even considering the complexities of the network).

Thus, products with high PageRank scores can arguably be classified as products that are not only well-connected but also well-connected to other products that are used by a high number of firms in production. More formally, PageRank of a product u, $PR(u)$ can be defined (see, for instance, De Keyser, 2012) as follow.

$$PR(u) = (1-d) + d\left(\sum_{v \neq u \in N(u)} \frac{PR(v)}{OC(v)}\right)$$

Above, $PR(u)$ is defined as the addition between $(1-d)$ where d is the "damping factor"[12], and the summation of the division between the PageRanks of every other nodes $v \in N(u)$ linked to u, i.e. $PR(v)$, and such nodes' out-degree $OC(v)$, multiplied again by the damping factor d. Table 9 presents the products (at the 10-digit level) with the highest PageRank scores.

A general observation on Table 9 shows that intermediate and more downstream products are primarily present in the table. This is shown by products such as bottled water, sweet bread, and trousers, among others being the products with the highest PageRank scores. Such an occurrence is expected as the graph is mostly acyclic. The 2-digit level aggregation is shown in Table 10.

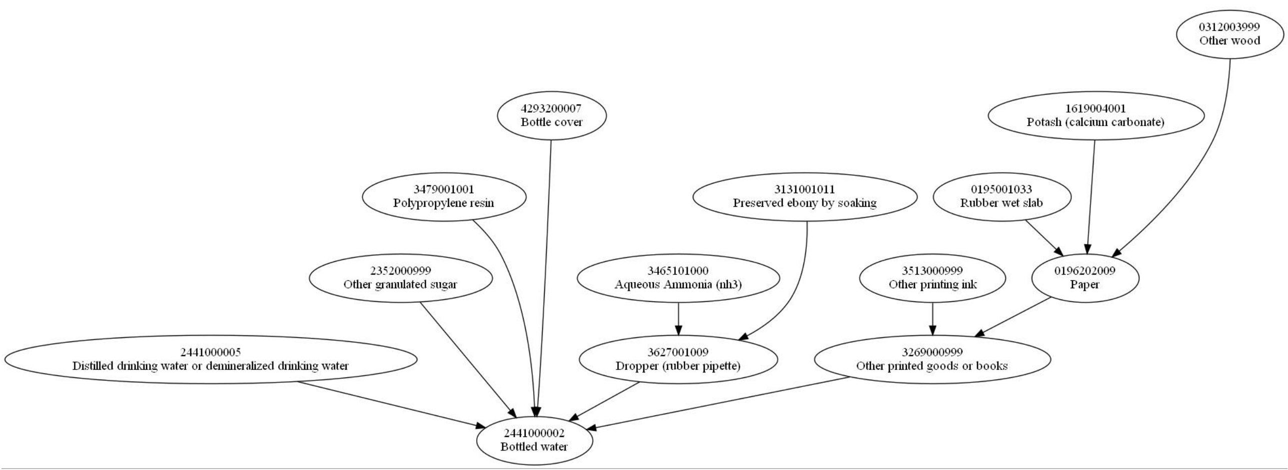
[11]Technically, the PageRank calculation also involves that the maze player will *end* her journey on that corner (due to the damping factor).

[12](Essentially the "boredom / continuation factor", i.e., how likely the random walker will/will not continue to the next iteration)

From Table 10, food and tobacco generally top the list of sectors possessing the highest number of presence in the list of 50 highest-PageRank products. Such sectors can be interpreted as having the most products where the said products are well-connected with other highly-used products.

Finally, let us also discuss one of the product-level snapshots that possess a high score of PageRank (relative to the complete network). This study uses a case study of *"Bottled water"* (product code *2441000002*). The Figure 9 presents the adjacent network – now with 3-stages of upstream and downstream linkages, if applicable – of *"Bottled water"* (product code *2441000002*).

Figure 9: **An illustration of *Bottled water*'s immediate network**



Source: Statistics Indonesia-BPS (2017); network is calculated by authors
Note: The figure above presents the products (represented as nodes in a graph) with upstream and/or downstream linkages with *Bottled water* (product code *2441000002*). The figure uses a network specification that seeks the central node (*Bottled water*)'s downstream and upstream nodes both traversing by 3 stages of linkages into either directions, with pruning threshold of 50. Note that the edges presented in the table omits the non-SILE edges. High-precision fonts are used in the figure, detailed zoom-ins are enabled for reading.

The visualization in Figure 9 is deliberately shown with a higher than one stage of adjacency, as the figure aims to show the depth and connectedness of *Bottled water* with the multitude of products it is connected to. As seen on the network, the most immediate predecessors of *Bottled water* are the drinking water itself (*Distilled drinking water or demineralized drinking water*), the bottle cover, printing on the cover (*Other printed goods or books*), among others. This study again argues that the findings in this example pass sense checks, as the mentioned products *do* form bottled waters.

3.4 Discussions

This section discusses how this study's framework and the characterisations mentioned above of the country's manufacturing sector using this rather different lens by utilising the Indonesian manufacturing survey as a network compared to the existing literature. Before delving further, however, this study argues that the number of studies compared to our results might be limited, mainly due to two reasons. The first reason is that the construction of firm-level and product-level data as networks and, consequently, its network characteristics exploration is still limited for the Indonesian context – if any exists. A second related reason would be that as such an analysis is still scarce, the topics discussed by past studies when using the Indonesian manufacturing survey data are still limited[13], as noted by (Márquez-Ramos, 2020). Using the framework of this study, we argue that new issues can be addressed by utilising the network theory framework as well as different characteristics that can be overlaid on the network structure.

First, we note that our findings generally point to the products contained within the scope of the food and beverages sector to be well-connected in relation to the whole production network. Our results further support one of the findings in a joint study by ADB and Indonesia's Ministry of Planning (*BAPPENAS*) in 2019 (Zhongming et al., 2019), as the study notes that one-fourth of the total number of manufacturing firms, one-sixth of the total employment, and just under one-fifth of total value added in the Indonesian manufacturing comes from the food sector[14]. Our study shows that products such as "Bottled water" and "Sweet bread", among others, are mentioned numerous times in terms of products with high network connectivity indicators, as well as updating calculations with the latest 2017 data available (Zhongming et al., 2019 used 2014 numbers).

Taking a broader context in the (Zhongming et al., 2019) study, a combination between food, textile, and wearing apparel manufacturing comprises 44 % of the overall number of firms and 39 % of the overall employment in Indonesia. Along with that finding, our study provides further evidence that products originating from such sectors are also well-connected products in our network depicting the manufacturing sector. For instance, our findings show that food products, as well as weaving/knitting products and other textiles, are well represented in the top indegree (most users of other sectors, see Table 4), outdegree (most used by other sectors, see Table 6), betweenness (most connecting of other products, see Table 8), and PageRank (see Table 10).

Using (Zhongming et al., 2019)'s classifications of technological groups of

[13]Based on the survey conducted on 33 studies that utilised the Indonesian manufacturing survey (as used in this study), past studies have generally focused on three topics, namely 1) trade liberalisation, foreign ownership, and firms' performance; 2) social issues (e.g., employment, poverty, skills, and wage); 3) other issues such as corruption and environment (see Márquez-Ramos, 2020).

[14]See the study's (Zhongming et al., 2019) Section 6.2 and its related description and discussions.

products in the Indonesian manufacturing[15], our findings suggest that the central products of Indonesian manufacturing are indeed more on the low-to-medium technological products. For instance, products from food and beverages, tobacco, textiles and apparel, as well as other low technological products, dominate the sectors with the highest centrality scores in Tables 4, 6, 8, and 10, if not the medium technology products such as rubber, basic metal, and others.

Further, our study also relates to the sectoral and product-specific studies previously conducted on Indonesian manufacturing, as the framework in our study allows the integration and characterization of how the said sector stands in the middle of the whole production network of Indonesian manufacturing. Such studies include, for instance, the sectoral study on the furniture industry (see Clements et al., 2019), chemical and pharmaceutical industry (see Suyanto and Salim, 2013), apparel industry (see Hayakawa et al., 2017), automotive industry (see Okamoto and Sjöholm, 2000), among others.

4 Conclusion

The body of literature utilizing models with sectors or products as nodes in networks have been rapidly expanding, starting from the early works by (W. W. Leontief, 1936), the traditional stream of literature utilizing the input-output based modeling (W. Leontief, 1987; Richardson, 1985) and the recent general equilibrium derivations based on IO tables (Acemoglu et al., 2012; Bartelme and Gorodnichenko, 2015; Baqaee, 2018; Acemoglu and Azar, 2020). However, the most granular level such studies can provide is still mostly limited to the sectoral level in most economies (except the U.S. tables where 4-digit level SIC is used).

Recent works in the Product Space framework literature provides a comprehensive and granular look of products-as-networks (Hidalgo et al., 2007; Hidalgo and Hausmann, 2008) and allows important policy questions to be addressed in different economies (Hausmann and Klinger, 2008a; Hausmann and Klinger, 2008b; De La Cruz and Riker, 2012). However, its method in constructing the product network is mainly based on the *results* of the production landscape in a country, that is, the trade and competitiveness patterns of countries around the world.

This study attempts to bring the product-level granularity of analysis that closely resembles the Product Space literature into the traditionally used input-output-based modelling. Specifically, this study explored and provided an approximation of a manufacturing sector 10-digit product-level unweighted input-output network using firm-level and product-level data from Indonesia's annual survey of the manufacturing sector.

[15] See Figure 6.2 in Zhongming et al., 2019.

Our results suggest that, generally, low-to-medium technology goods in Indonesia are the products that are the most central in the Indonesian production network, using several centrality indicators from network theory. We also discussed the different centrality scores that might be useful for different types of exploratory questions.

With all the above being said, this study still suffers from several limitations. First, the inherent shape of the Indonesian raw material (*Rawin*) and output product (*Proin*) datasets that are only able to be joined by firm IDs limit our study to only being an approximation of the true network. We also limit this study's network to an unweighted version due to this condition, despite the value-added figures being available in the data. A future study that can robustly estimate and properly weigh the edge values in this network will thus be able to simulate a proper 10-digit product-level input-output table.

Second, this study only covers the Indonesian manufacturing context. In the absence of other countries' networks, the only meaningful analysis derived would be limited to the Indonesian-only cases. This study is also only depicting the domestic market linkages, as a connection between the network to other countries (e.g., through trade) requires concordance between the Indonesian product-level code (KBKI) and the standardized trade codes (e.g., HS codes). Future works that incorporate such analysis will provide an even richer analysis with the addition of regional and international trade aspects.

However, despite the limitations, this study argues that the importance of incorporating a more granular level of analysis in the input-output based modelling, as well as the richness of analysis that can be provided if one is to combine it with the product-level trade-related networks (e.g., product space frameworks), cannot be understated. With the complexities and intricacies of global trade, economies with better intelligence of the existing (input-output models) and potential (PS models) markets will gain an advantage. This study attempts to give Indonesia its first step in the said direction.

References

Acemoglu, D., & Azar, P. D. (2020). Endogenous production networks. *Econometrica*, *88*(1), 33–82.

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

Baqaee, D. R. (2018). Cascading failures in production networks. *Econometrica*, *86*(5), 1819–1838.

Bartelme, D., & Gorodnichenko, Y. (2015). *Linkages and economic development* (tech. rep.). National Bureau of Economic Research.

Clements, C., Alwang, J., & Achdiawan, R. (2019). Value chain approaches in a stagnant industry: The case of furniture production in jepara, indonesia. *Bulletin of Indonesian Economic Studies*, *55*(3), 341–365.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). 22.2 breadth-first search. *Introduction to algorithms*.

De Keyser, P. (2012). *Indexing: From thesauri to the semantic web*. Elsevier.

De La Cruz, J., & Riker, D. (2012). *Product space analysis of the exports of brazil*. US Internat. Trade Commission, Office of Economics.

Hansen, D. L., Shneiderman, B., Smith, M., & Himelboim, I. (2011). Social network analysis: Measuring, mapping, and modeling collections of connections. analyzing social media networks with nodexl: Insights from a connected world.

Hausmann, R., & Klinger, B. (2008a). Growth diagnostics in peru. *CID Working Paper Series*.

Hausmann, R., & Klinger, B. (2008b). South africa's export predicament. *Economics of transition*, *16*(4), 609–637.

Hayakawa, K., Matsuura, T., & Takii, S. (2017). Does trade liberalization boost quality upgrading? evidence from indonesian plant-product-level data. *The Developing Economies*, *55*(3), 171–188.

Hidalgo, C. A., & Hausmann, R. (2008). A network view of economic development. *Developing alternatives*, *12*(1), 5–10.

Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, *317*(5837), 482–487.

Leontief, W. (1987). Input-output analysis. *The new palgrave. A dictionary of economics*, *2*(1), 860–864.

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

Márquez-Ramos, L. (2020). A survey of papers using indonesian firm-level data: Research questions and insights for novel policy-relevant research in economics. *Bulletin of Indonesian Economic Studies*, (just-accepted), 1–49.

Moore, E. F. (1959). The shortest path through a maze. *Proc. Int. Symp. Switching Theory, 1959*, 285–292.

Okamoto, Y., & Sjöholm, F. (2000). Productivity in the indonesian automotive industry. *ASEAN Economic Bulletin*, 60–73.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The pagerank citation ranking: Bringing order to the web.* (tech. rep.). Stanford InfoLab.

Perez, C., & Germon, R. (2016). Graph creation and analysis for linking actors: Application to social data. In *Automating open source intelligence* (pp. 103–129). Elsevier.

Poot, H. (1991). Interindustry linkages in indonesian manufacturing. *Bulletin of Indonesian Economic Studies*, *27*(2), 61–89.

Richardson, H. W. (1985). Input-output and economic base multipliers: Looking backward and forward. *Journal of Regional science*, *25*(4), 607–661.

Stepner, M. (2013). Binscatter: Stata module to generate binned scatterplots.

Suyanto & Salim, R. (2013). Foreign direct investment spillovers and technical efficiency in the indonesian pharmaceutical sector: Firm level evidence. *Applied Economics*, *45*(3), 383–395.

Tui, R. N. S., & Adachi, T. (2021). An input-output approach in analyzing indonesia's mineral export policy. *Mineral Economics*, *34*(1), 105–112.

Zhongming, Z., Linong, L., Xiaona, Y., Wangqiang, Z., Wei, L., et al. (2019). Policies to support the development of indonesia's manufacturing sector during 2020–2024: A joint adb–bappenas report.

Zuhdi, U. (2015). The dynamics of indonesian creative industry sectors: An analysis using input–output approach. *Journal of the Knowledge Economy*, *6*(4), 1177–1190.

Zuse, K. (1972). Der plankalkül, konrad zuse internet archive.

Appendices

Extra Tables

Table 3: **Products utilizing most number of inputs in production (sorted by indegree centrality scores)**

No	KBKI code (10 digits)	Product description	Indegree
1	2349001002	Sweet bread	9
2	3811201999	Other wooden tables/chairs/benches	9
3	2153500000	Crude oil/raw oil palm	8
4	4129100999	Vessels, pipes and other hollow profiles of cast iron including vessels and other pipes from centrifugal cast steel	8
5	2922002999	Genuine leather goods for other personal purposes	7
6	4123199000	Steel sheet and other coating materials	7
7	2509001001	Cigarette tobacco	7
8	2399909999	Other crackers	7
9	4912904999	Other power train components	7
10	3699003002	Plastic electronic components	6

Source: Statistics Indonesia-BPS (2017); network is calculated by authors

Note: The table above presents the products (nodes in a graph) that use the most number of other products as inputs (denoted as a strong edge in the graph). Indegree calculations explain how many other products are used as inputs by a certain product. Note that the number of edges presented in the table omits the non-SILE edges.

Table 4: **Sectors (2-digits KBKI) utilizing most number of inputs in production (sorted by the highest contributors to top 50 indegree calculations)**

No	KBKI code (2 digits)	Product description	Number of products (10 digits) in top 50 indegree	%
1	23	Paddy milling, kanji and other kanji products; other food products	6	12%
2	21	Meat, fish, fruits, vegetables, oil and fat	6	12%
3	36	Rubber and plastic products	4	8%
4	31	Products from wood, cork, straw and woven materials	4	8%
5	26	Yarn and yarn weaving/knitting; Woven fabric and tagged textile fabric	3	6%
6	25	Tobacco products	3	6%
7	37	Glass and glass products and other unclassified non-metal products	3	6%
8	38	Household furniture; other unclassfied items that can be moved	3	6%
9	41	Basic metal	3	6%
10	28	Knitted cloth or link; apparel	2	4%
11	29	Leather and product from the skin; footwear	2	4%
12	42	Manufacturing metal products, except machineries and equipment	2	4%
13	32	Pulp, paper and paper products; printed goods and related items	2	4%
14	03	Forestry products and logging	1	2%
15	35	Other chemical products; artificial fiber	1	2%
16	49	Transportation equipment	1	2%
17	46	Electric machines and components	1	2%
18	33	Coke oven products, processed petroleum products, nuclear fuels	1	2%
19	24	Drink	1	2%
20	01	The results of agriculture, horticulture and plantations	1	2%

Source: Statistics Indonesia-BPS (2017); network is calculated by authors

Note: The table above presents the products (nodes in a graph), aggregated to the 2-digit level, that use the most number of other products as inputs (denoted as a strong edge in the graph). Note that the number of edges presented in the table omits the non-SILE edges.

Table 5: **Products utilized the most by other products as inputs in production (sorted by outdegree centrality scores)**

No	KBKI code (10 digits)	Product description	Outdegree
1	2641001002	Synthetic filament sewing yarns from poliester	35
2	3465101000	Non-liquid ammonia (NH3)	33
3	0192101000	Cotton	21
4	2311000001	Wheat flour	18
5	4123199000	Steel sheet and other coating materials	14
6	4153400001	Aluminum plates, with a thickness exceeding 0.2 mm	14
7	0411101999	Other fish	13
8	4621302999	Other distribution panels	12
9	4112205999	Iron and other alloy base steel	12
10	0196202009	Paper	11

Source: Statistics Indonesia-BPS (2017); network is calculated by authors
Note: The table above presents the products (nodes in a graph) that are used by the most number of other products as inputs (denoted as a strong edge in the graph). Outdegree calculations explains how many other products use the product as an input. Note that the number of edges presented in the table omits the non-SILE edges.

Table 6: **Sectors (2-digits KBKI) utilized the most by other products as inputs in production (sorted by the highest contributors to top 50 outdegree calculations)**

No	**KBKI code (2 digits)**	**Product description**	**Number of products (10 digits) in top 50 outdegree**	**%**
1	23	Paddy milling, kanji and other kanji products; other food products	6	12%
2	41	Basic metal	6	12%
3	01	The results of agriculture, horticulture and plantations	5	10%
4	26	Yarn and yarn weaving/knitting; Woven fabric and tagged textile fabric	4	8%
5	21	Meat, fish, fruits, vegetables, oil and fat	3	6%
6	34	Basic chemistry	3	6%
7	35	Other chemical products; artificial fiber	3	6%
8	42	Manufacturing metal products, except machineries and equipment	3	6%
9	03	Forestry products and logging	2	4%
10	38	Household furniture; other unclassified items that can be moved	2	4%
11	27	Items from textiles other than clothing	2	4%
12	16	Other minerals	2	4%
13	25	Tobacco products	1	2%
14	49	Transportation equipment	1	2%
15	45	Office, Accounting and Computing Machines	1	2%
16	46	Electric machines and components	1	2%
17	29	Leather and product from the skin; footwear	1	2%
18	33	Coke oven products, processed petroleum products, nuclear fuels	1	2%
19	04	Fish and other fishery products	1	2%
20	02	Live animals and animal products (not including meat)	1	2%
21	32	Pulp, paper and paper products; printed goods and related items	1	2%

Source: Statistics Indonesia-BPS (2017); network is calculated by authors

Note: The table above presents the products (nodes in a graph) that are used by the most number of other products as inputs (denoted as a strong edge in the graph). Note that the number of edges presented in the table omits the non-SILE edges.

Table 7: **Products that connects the most production chains (sorted by betweenness centrality scores)**

No	KBKI code (10 digits)	Product description	Betweenness
1	4123199000	Steel sheet and other coating materials	490.3
2	4299907004	Other metal bolts and nuts	340.0
3	3870202999	Other aluminum metal construction materials	294.3
4	4621302999	Other distribution panels	240.0
5	4153400001	Aluminum plates, with a thickness exceeding 0.2 mm	171.0
6	4112202999	Billet steel alloys	121.0
7	2153500000	Crude oil/raw oil palm	77.0
8	4129100999	Vessels, pipes and other hollow profiles of cast iron including vessels and other pipes from centrifugal cast steel	49.0
9	0196202009	Paper	48.0
10	2509001001	Cigarette tobacco	48.0

Source: Statistics Indonesia-BPS (2017); network is calculated by authors

Note: The table above presents the products (nodes in a graph) that connect the most production chains by being between shortest paths of other products (betweenness). Betweenness centrality calculations explain how "central" a product is by showing how many shortest paths among every other products passes through the node. Note that the number of edges considered in producing the table omits the non-SILE edges.

Table 8: **Sectors (2-digits KBKI) utilized the most by other products as inputs in production (sorted by the highest contributors to top 50 betweenness calculations)**

No	KBKI code (2 digits)	Product description	Number of products (10 digits) in top 50 betweenness	%
1	41	Basic metal	5	20%
2	23	Paddy milling, kanji and other kanji products; other food products	3	12%
3	34	Basic chemistry	3	12%
4	26	Yarn and yarn weaving/knitting; Woven fabric and tagged textile fabric	2	8%
5	16	Other minerals	1	4%
6	42	Manufacturing metal products, except machineries and equipment	1	4%
7	33	Coke oven products, processed petroleum products, nuclear fuels	1	4%
8	25	Tobacco products	1	4%
9	36	Rubber and plastic products	1	4%
10	38	Household furniture; other unclassified items that can be moved	1	4%
11	46	Electric machines and components	1	4%
12	21	Meat, fish, fruits, vegetables, oil and fat	1	4%
13	49	Transportation equipment	1	4%
14	01	The results of agriculture, horticulture and plantations	1	4%
15	32	Pulp, paper and paper products; printed goods and related items	1	4%
16	35	Other chemical products; artificial fiber	1	4%

Source: Statistics Indonesia-BPS (2017); network is calculated by authors

Note: The table above presents the products (nodes in a graph) that connect the most production chains by being between shortest paths of other products (betweenness). Note that the number of edges considered in producing the table omits the non-SILE edges.

Table 9: **Products connected to other quality products (sorted by PageRank scores)**

No	KBKI code (10 digits)	Product description	PageRank
1	2349001002	Sweet bread	0.00961
2	3811201999	Other wooden tables/chairs/benches	0.00928
3	2441000002	Bottled water	0.00862
4	2399909999	Other crackers	0.00761
5	2509001001	Cigarette tobacco	0.00599
6	2502002001	*Kretek* cigarette without filter	0.00590
7	2153900001	Crude oil/raw oil palm seeds	0.00579
8	4654100999	Other direct components of electric lights	0.00571
9	4641000999	Other batteries	0.00493
10	2153500000	Crude oil/raw oil palm	0.00483

Source: Statistics Indonesia-BPS (2017); network is calculated by authors

Note: The table above presents the products (nodes in a graph) sorted by the weighted PageRank scores. The PageRank calculations explain how "important" a product is by scoring each product based on their connections and the connections' connections. Edge weights is defined as the number of firms reporting to use the predecessor node as an input to produce the successor node. Note that the number of edges considered in producing the table omits the non-SILE edges.

Table 10: **Sectors (2-digit level) by number of presence in top-50 PageRank scores**

No	KBKI code (2 digits)	Product description	Number of products (10 digits) in top 50 PageRank	%
1	23	Paddy milling, kanji and other kanji products; other food products	7	14%
2	36	Rubber and plastic products	5	10%
3	41	Basic metal	5	10%
4	25	Tobacco products	5	10%
5	38	Household furniture; other unclassified items that can be moved	4	8%
6	21	Meat, fish, fruits, vegetables, oil and fat	4	8%
7	15	Stones, sand and clay	2	4%
8	46	Electric machines and components	2	4%
9	24	Drink	2	4%
10	01	The results of agriculture, horticulture and plantations	1	2%
11	49	Transportation equipment	1	2%
12	28	Knitted cloth or link; apparel	1	2%
13	42	Manufacturing metal products, except machineries and equipment	1	2%
14	32	Pulp, paper and paper products; printed goods and related items	1	2%
15	29	Leather and product from the skin; footwear	1	2%
16	34	Basic chemistry	1	2%
17	26	Yarn and yarn weaving/knitting; Woven fabric and tagged textile fabric	1	2%
18	31	Products from wood, cork, straw and woven materials	1	2%
19	35	Other chemical products; artificial fiber	1	2%

(continued on the next page)

Table 10: **Sectors (2-digit level) by number of presence in top-50 PageRank scores**

No	**KBKI code (2 digits)**	**Product description**	**Number of products (10 digits) in top 50 PageRank**	**%**
20	16	Other minerals	1	2%
21	33	Coke oven products, processed petroleum products, nuclear fuels	1	2%
22	47	Radio equipment, television and communication tools and equipment	1	2%
23	37	Glass and glass products and other unclassfied non-metal products	1	2%

Source: Statistics Indonesia-BPS (2017); network is calculated by authors

Note: The table above presents the frequency of 10-digit level products in each sector (2-digit level aggregation) appearing in the top-50 products with highest PageRank scores. Edge weights is defined as the number of firms reporting to use the predecessor node as an input to produce the successor node. Note that the number of edges considered in producing the table omits the non-SILE edges.