Frontiers

Pattern-detection in the global automotive industry: A manufacturer-supplier-product network analysis

Massimiliano Fessina [a,*], Andrea Zaccaria [b,c], Giulio Cimini [d,c], Tiziano Squartini [a]

[a] *IMT School for Advanced Studies, Piazza San Francesco 19, 55100 Lucca, Italy*
[b] *Institute for Complex Systems (ISC), National Research Council, Piazzale Aldo Moro 2, 00185 Rome, Italy*
[c] *Enrico Fermi Research Center (CREF), Via Panisperna 89A, 00184 Rome, Italy*
[d] *Physics Department and INFN, University of Rome Tor Vergata, Via della Ricerca Scientifica 1, 00133 Rome, Italy*

ARTICLE INFO

ABSTRACT

Production networks arise from customer–supplier relationships between firms. These systems have gained increasing attention as a consequence of the frequent supply chain disruptions caused by the natural and man-made disasters occurred during the last years (e.g. the Covid-19 pandemic and the Russia-Ukraine war). Recent, empirical evidence has shown that production networks are shaped by 'functional' structures reflecting the complementarity of firms, i.e. their tendency to compete. However, data constraints force the few, available studies to consider only country-specific production networks. In order to fully capture the cross-country structure of modern supply chains, here we focus on the global, automotive industry as depicted by the 'MarkLines Automotive' dataset. After representing it as a network of manufacturers, suppliers and products, we look for the statistical significance of the aforementioned, 'functional' structures. Our exercise reveals the presence of several pairs of manufacturers sharing a significantly large number of suppliers, a result confirming that any two car companies are seldom engaged in a buyer–supplier relationship: rather, they compete although being connected to many, common neighbors. Interestingly, 'generalist' suppliers serving many manufacturers co-exist with 'specialist' suppliers serving few manufacturers. Additionally, we unveil the presence of patterns with a clearly spatial signature, with manufacturers clustering around groups of geographically close suppliers: for instance, Chinese firms constitute a disconnected community, likely an effect of the protectionist policies promoted by the Chinese government. We also show the tendency of suppliers to organize their production by targeting specific car systems, i.e. combinations of technological devices designed for specific tasks. Besides shedding light on the self-organizing principles shaping production networks, our findings open up the possibility of designing realistic generative models of supply chains, to be used for testing the resilience of the global economy.

1. Introduction

The growth of network science over the last twenty years has impacted several disciplines, by establishing new, empirical facts about the structural properties of those systems as well as novel methodologies for their analysis. Prominent examples are represented by economic and financial networks, such as the international trade [1–3], product exports by countries [4–7], transaction networks [8,9] and, after the 2008 crisis, interbank networks [10–14]. A class of systems that has recently gained attention is that of production networks emerging as (manufacturing) firms become dependent on other (supplier) firms for their own production. As a consequence of globalization [15] and of a constant strive towards efficiency [16], production networks have become increasingly interdependent - a feature lying at the basis of the

business interruptions that have occurred in consequence of the recent, natural and man-made disasters (e.g. the Covid-19 pandemic and the Russia-Ukraine war) [17–22].

The propagation of economic shocks has been traditionally studied using *industry-level* input–output tables [23], although the importance of *firm-level* supply data to provide an unbiased assessment of the consequences of individual failures [24,25] has been increasingly recognized [26–28]. Network theory is, thus, becoming an increasingly popular tool to analyze production systems [29–40], analogously to what has happened within the realm of financial systems in the aftermath of the 2008 crisis [41].

Firm-level datasets, however, are notoriously difficult to acquire because of both technical and privacy issues [42]. While early works

---

relied on aggregated flows of goods between countries [43–47], more detailed datasets have recently appeared: still, the number of firm-level datasets, with large-scale coverage, that are built from financial reports is limited; besides, they only cover the larger companies listed on the US stock exchanges and their main customers. This is the case of Factset [42,48], Compustat [36,49,50] and Capital IQ [51,52] data. Other, global datasets that have been analyzed in the Operations Research and Supply Chain Management literature cover specific, industrial sectors [53]: the most popular, and complete, one concerns the automotive sector, obtained from a private industry database (the 'MarkLines Automotive Information Platform') compiled through surveys sent to automotive supplier firms [54,55]. At the level of individual countries, instead, production networks can be constructed either from value-added tax (VAT) data (which exists for Belgium [56], Ecuador [57], Hungary [39] and Spain [58]) or from payment data mediated by central, or major, banks (which exists for Brazil [59], Japan [60,61] and The Netherlands [62]). National data, often characterized by a reporting threshold, typically provide a very good internal coverage.

The Japanese, inter-firm dataset provided by Tokyo Shoko Research (TSR) was the first, national, firm-level network to become available and, as such, has been extensively studied during the last decade [63–66]. It provides credit information of about one million firms, as well as the identity of the most relevant suppliers and customers for the business activity of each. Empirical analyses revealed that both firm-specific, structural quantities (e.g. the number of connections) and purely financial indicators (e.g. the total amount of sales, the total number of employees) are power-law distributed, while a firm degree grows along with its total sales in a non-linear fashion [63]. Regarding the topology of the network, it was shown to be disassortative by degree (as suppliers with few customers are preferentially connected with large firms) and characterized by a well-defined community structure, with clusters being shaped by geographical proximity and industrial sector similarity [66]. Notably, these clusters are characterized by bipartite structures with large companies not being directly linked but sharing many first-tier suppliers. This result was confirmed by a comprehensive study of triadic motifs [65], i.e. all, possible interaction patterns involving three nodes [67]. When compared with a benchmark constraining the degree of each firm, the TSR data features an under-representation of closed triangles and an over-representation of V-motifs [3], i.e. the patterns shaped by pairs of nodes (firms) that are not directly connected but share a common neighbor (be it a supplier or a customer).

The abundance of V-motifs in production networks can be interpreted as a sign that their self-organization is driven by *complementarity* ('similar nodes tend to share the same partners') rather than *homophily* ('similar nodes tend to be connected') [68]. From this perspective, production networks are more similar to protein-interaction networks than to social networks: in the first case, two proteins with similar binding sites do not interact directly but can be both connected with others having complementary binding properties [69]; in social networks, on the contrary, acquaintances, people with similar interests, etc. are likely to be connected (according to the motto 'the friend of my friend is my friend' [70]). Additionally, the over-representation of V-motifs is compatible with the over-representation of the square patterns known as X-motifs [3], given by two nodes (e.g. firms) that are not directly connected but share pairs of common neighbors (be they suppliers or customers). The findings of [65,66], thus, corroborate the picture according to which firms with similar outputs are often engaged in a competition, rather than in a buyer–supplier relationship — although they can have many suppliers and customers in common [55].

The aforementioned, over-representation of square motifs, however, has been explicitly shown only for the company-level, production network provided by CBS — Statistics Netherlands [68,71]: no results are available for global datasets, despite the well-documented cross-country structure of production networks [72]. The aim of this paper is bridging the gap, by carrying out a pattern-detection analysis on

production networks at the global scale, employing the 'MarkLines Automotive' dataset. In particular, after representing the system in terms of a manufacturer-supplier-product network, we carry out a preliminary description of its structural features. Then, we test their statistical relevance by adopting a validation framework based on the maximum-entropy principle. Specifically, we apply the method proposed in [73], which takes as input the bipartite, manufacturer-supplier representation of the 'MarkLines Automotive' dataset and produces monopartite projections whose connections are induced by the similarity of the corresponding nodes — in our case, the number of V-motifs, i.e. the common neighbors in the original, bipartite network. In statistical terms, this amounts to unveiling the presence of significant patterns as a consequence of the rejection of the null hypothesis that the node degrees are the only explanatory variables of the network [3,73–75]; filtering out the statistical noise allows us to discard the spurious patterns (e.g. the motifs emerging as a simple consequence of the node degrees [3]) while retaining those emerging as a genuine consequence of the interactions between nodes. Remarkably, the different kinds of structures we detect seem to clearly reflect national development strategies.

2. Methods

2.1. Network representation of the 'MarkLines Automotive' dataset

The 'MarkLines Automotive' dataset (https://www.marklines.com) identifies the 'manufacturers' with the companies making cars (e.g. *Honda*) and the 'suppliers' with the companies selling them products (e.g. *Hitachi*). After a thorough data-cleaning and harmonization procedure (see Appendix A), we consider $M = 301$ manufacturers, $S = 5.725$ suppliers, and $P = 291$ products, while the overall number of buyer–supplier relationships amounts to 26.535 (see Table 1). As there are few, 'internal' links between manufacturers (a car maker can happen to act as a supplier of another car maker) and no links between suppliers, the dataset admits a natural, bipartite representation once the 'internal' links are discarded.

According to the (technological) taxonomy provided by the platform, products are classified into five categories: Chassis/Body (CB), Electrical (EL), Powertrain (PW), Interior/Exterior (IE) and General parts (GP). Information about each category of products, indexed by $\gamma = 1 \ldots 5$, can be arranged into a biadjacency matrix $\mathbf{B}^\gamma$ whose dimensions read $M^\gamma \times S^\gamma$, where M^γ is the number of manufacturers and S^γ is the number of suppliers in category γ: naturally, $b_{ms}^\gamma = 1$ if supplier s provides manufacturer m with at least one product belonging to category γ and $b_{ms}^\gamma = 0$ otherwise. Overall, then, the 'MarkLines Automotive' dataset can be represented as a bipartite multiplex, each layer corresponding to a category of technological products (see also the left panel of Fig. 1).

Alternatively, the dataset can be represented as an $M \times S \times P$ tripartite network, i.e. a 'combination' of two bipartite networks sharing the set of suppliers (see also the right panel of Fig. 1): the generic element of the 'left' $M \times S$ biadjacency matrix $\mathbf{L}$ reads $l_{ms} = 1$ if manufacturer m buys from supplier s, otherwise $l_{ms} = 0$; analogously, the generic element of the 'right' $S \times P$ biadjacency matrix $\mathbf{R}$ reads $r_{sp} = 1$ if supplier s sells product p and $r_{sp} = 0$ otherwise. Although we could also link each manufacturer to the set of products purchased by its suppliers, the evidence that all of them are car producers, hence needing the same basket of products, let us opt for discarding this third set of connections.[1] Hereby, we will adopt the tripartite representation.

---

[1] Intuition would suggest each manufacturer to need a minimum amount of products to make a car. This, however, does not seem to be the case as manufacturers exist that are linked to just one supplier but no supplier exists that provides all products. This, in turn, suggests that the 'MarkLines Automotive' dataset is (at least, partially) incomplete.
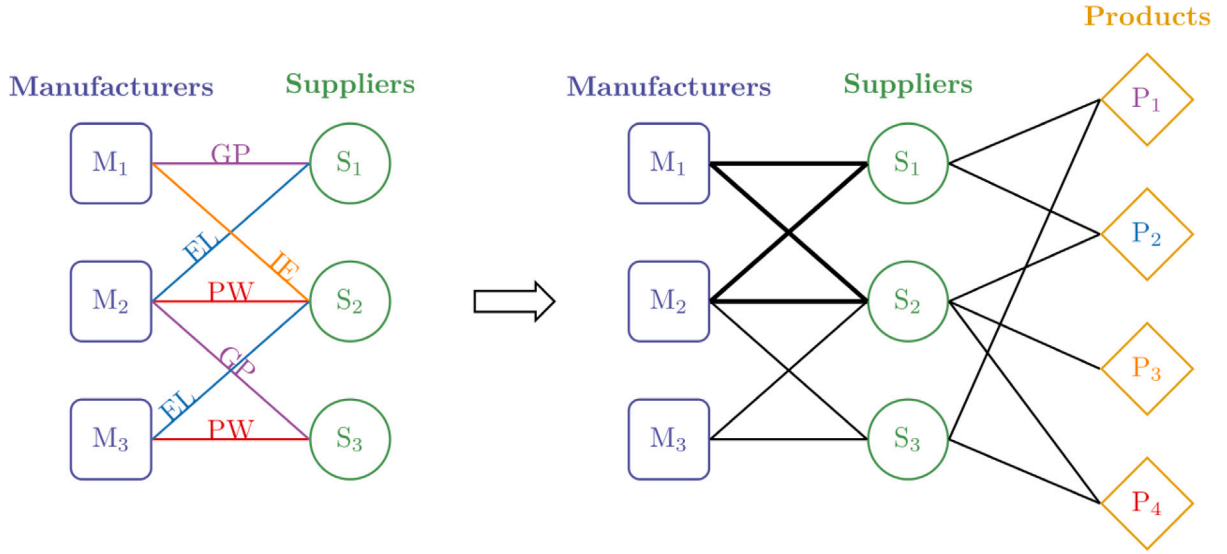
Fig. 1. The 'MarkLines Automotive' dataset, represented either as a bipartite multiplex (left) or as a tripartite network (right). In the multiplex representation, there are five layers, corresponding to the product categories of the data: a link between a manufacturer and a supplier corresponds to the exchange of a specific product, belonging to the corresponding category and colored accordingly. In the tripartite representation, a link between a manufacturer and a supplier is present as long as they are connected by, at least, one link across all layers; each supplier, instead, is connected to all the products it sells. Bold lines illustrate the motifs proxying competition between manufacturers and that we statistically validate..

Table 1
Basic statistics of the 'MarkLines Automotive' dataset for each category of products: Chassis/Body (CB), Electrical (EL), Powertrain (PW), Interior/Exterior (IE) and General parts (GP). The percentages of 'non-bipartite' links refer to the connections between manufacturers for each product category.

	CB	EL	PW	IE	GP	Full
# manufacturers	246	215	249	196	108	301
# suppliers	1.864	1.203	2.659	1.776	479	5.725
# links	7.937	5.390	10.978	6.502	1.687	26.535
% 'non-bipartite' links	0.4	0.94	2.6	0.83	0	1.3

2.2. Structural properties

Local connectivity. The most important network quantity at the local level is the degree, defined as the number of connections of a node. In order to properly describe our data, we need the following definitions:

- the number of suppliers, or product providers, of manufacturer m: $k_{m \to S} = \sum_{s=1}^{S} l_{ms}$;
- the number of customers, or client manufacturers, of supplier s: $k_{s \to \mathcal{M}} = \sum_{m=1}^{M} l_{ms}$;
- the 'diversification' of supplier s, i.e. the number of products it sells to its client manufacturers: $k_{s \to \mathcal{P}} = \sum_{p=1}^{P} r_{sp}$;
- the 'ubiquity' of product p, i.e. the number of its vendor suppliers: $k_{p \to S} = \sum_{s=1}^{S} r_{sp}$.

Assortativity. The presence of degree correlations is captured by the average nearest neighbors degree. In order to properly describe our data, we need the following definitions:

- the average number of customers of a manufacturer neighbors: $\text{ANND}_{m \to S} = \sum_{s=1}^{S} l_{ms} k_{s \to \mathcal{M}} / k_{m \to S}$;
- the average number of providers of a supplier neighbors: $\text{ANND}_{s \to \mathcal{M}} = \sum_{m=1}^{M} l_{ms} k_{m \to S} / k_{s \to \mathcal{M}}$;
- the average 'ubiquity' of a supplier products: $\text{ANND}_{s \to \mathcal{P}} = \sum_{p=1}^{P} r_{sp} k_{p \to S} / k_{s \to \mathcal{P}}$;
- the average 'diversification' of a product suppliers: $\text{ANND}_{p \to S} = \sum_{s=1}^{S} r_{sp} k_{s \to \mathcal{P}} / k_{p \to S}$.

Motifs. In order to capture the concept of a node 'highly connected neighborhood', we need to consider the bipartite analogue of the monopartite, square clustering coefficient [76]. To this aim, several definitions have been proposed so far (see also Appendix B): here, we adopt the one provided in [77] and reading[2]

$$\text{BCC}_m = \frac{\sum_{s<s'} q_m(s, s')}{\sum_{s<s'} [(k_{s \to \mathcal{M}} - 1) + (k_{s' \to \mathcal{M}} - 1) - q_m(s, s')]} \quad (1)$$

where $k_{s \to \mathcal{M}}$ is the number of customers, or client manufacturers, of supplier s and

$$q_m(s, s') = \sum_{\substack{m'=1 \\ (m' \neq m)}}^{M} l_{ms} l_{ms'} l_{m's} l_{m's'} = l_{ms} l_{ms'} \sum_{\substack{m'=1 \\ (m' \neq m)}}^{M} l_{m's} l_{m's'} \quad (2)$$

counts the number of cycles, composed by four links, involving the common neighbors of s and s', other than m. According to BCC_m, the total number of closed squares involving manufacturer m is given by the sum of the degrees of all pairs of its neighbors minus $q_m(s, s')$, i.e. the number of squares that are already closed. In other words, the total number of closed squares coincides with the number of squares that could become closed upon connecting, by adding *new* links, the neighbors of s with s' and the neighbors of s' with s, excluding m from both sets.

Upon considering that the generic addendum $l_{ms} l_{ms'} l_{m's} l_{m's'}$ equals 1 if a closed square involving nodes m, m', s and s' exists and 0 otherwise, we can compute the number of squares involving manufacturers m and

---

[2] In order to keep the discussion as simple as possible, we provide the definition of the bipartite clustering coefficient only for manufacturers. Whereas needed, it will be extended to suppliers and products.

m' by summing over s and s', i.e. as

$$X_{mm'} = \sum_{s<s'} l_{ms} l_{ms'} l_{m's} l_{m's'} = \binom{V_{mm'}}{2} \qquad (3)$$

where $V_{mm'} \equiv \sum_{s=1}^{S} l_{ms} l_{m's}$ is the number of common suppliers (V-motifs) to m and m' [3]. Hence, the number of cycles, composed by four links, involving node m equals the number of X-motifs involving it, i.e. $\sum_{s<s'} q_m(s,s') = \sum_{\substack{m'=1 \\ (m'\neq m)}}^{M} X_{mm'}$.

Subgraph centrality. In order to detect the presence of closed paths of higher order, we have considered the bipartite analogue of the monopartite subgraph centrality [78], whose definition reads

$$\mathrm{BSC}_m = \frac{\sum_{k=0}^{\infty} [\mathbf{L}^{2k}]_{mm}}{(2k)!} = [\cosh(\mathbf{L})]_{mm} \qquad (4)$$

where $\mathbf{L}^0 \equiv \mathbf{I}$, i.e. the zero-th power of $\mathbf{L}$ coincides with the $M \times M$ identity matrix, and only closed walks of even length are accounted for. Still, carrying out a meaningful comparison of the centrality of nodes across different configurations requires it to be properly normalized; hereby, we adopt the following definition

$$\overline{\mathrm{BSC}}_m = \frac{\mathrm{BSC}_m}{\sum_{m=1}^{M} \mathrm{BSC}_m} = \frac{[\cosh(\mathbf{L})]_{mm}}{\mathrm{Tr}[\cosh(\mathbf{L})]}. \qquad (5)$$

Nestedness. Finally, we inspect the mesoscale organization of our bipartite networks by measuring their (degree of) nestedness [79], i.e. the tendency of low-degree nodes to have neighbors that are a subset of the neighbors of high-degree nodes. This pattern leads to a 'triangular' structure that has been observed in ecological systems [80], economic systems [81], etc. To quantify (the degree of) nestedness, we employ the NODF [82], reading

$$\mathrm{NODF}(\mathbf{L}) = \frac{1}{K} \left[\sum_{m,m'=1}^{M} \left(\theta(k_{m\to S} - k_{m'\to S}) \frac{\sum_{s=1}^{S} l_{ms} l_{m's}}{k_{m'\to S}} \right) \right.$$
$$\left. + \sum_{s,s'=1}^{S} \left(\theta(k_{s\to\mathcal{M}} - k_{s'\to\mathcal{M}}) \frac{\sum_{m=1}^{M} l_{ms} l_{ms'}}{k_{s'\to\mathcal{M}}} \right) \right]. \qquad (6)$$

2.3. Statistical benchmarks

Answering the question of whether an empirical property represents a non-trivial signature of a network requires comparing it with a properly defined benchmark (or *null model*, as its role is analogous to the one played by a null hypothesis in traditional statistics).

Following the approach introduced in [83] and developed in [84], here we adopt the Exponential Random Graphs (ERG) formalism, characterizing maximum-entropy probability distributions that preserve a desired set of constraints on average while keeping everything else as random as possible [85]. Among the models that can be defined within this framework, we consider the Bipartite Configuration Model (BiCM), introduced in [3] and defined by constraining the degrees of the nodes belonging to both layers of a bipartite network. In other words, this model embodies the null hypothesis that (the numerical values of) empirical network patterns are induced by (the numerical values of) the degrees of the nodes.

Let us, now, briefly illustrate the basic results for our manufacturer-supplier network, redirecting the interested reader to [3] for the explicit derivation of the BiCM. In the case of the BiCM, constrained entropy-maximization leads to a factorized probability reading $P(\mathbf{L}) = \prod_{m,s} p_{ms}^{l_{ms}} (1-p_{ms})^{1-l_{ms}}$ where

$$p_{ms} = \frac{x_m y_s}{1 + x_m y_s} \qquad (7)$$

is the probability that manufacturer m and supplier s are connected (i.e. that $l_{ms} = 1$) and x_m and y_s are functions of the Lagrange multipliers, respectively controlling for the degrees $k_{m\to S}$ and $k_{s\to M}$. These parameters, sometimes called 'fitnesses', are usually related to the (economic, financial, etc.) importance of a node, as degrees are

typically proportional to fluxes (export/import, money lent/money borrowed, etc.) [13].

The vectors of parameters $\boldsymbol{x}$ and $\boldsymbol{y}$ must be numerically determined. Here, we employ the maximum-likelihood principle, prescribing to maximize the expression $\mathscr{L} = \ln P(\mathbf{L}|\boldsymbol{x},\boldsymbol{y})$ with respect to x_m, $\forall\, m$ and y_s, $\forall\, s$. Such a recipe leads us to find the system of equations

$$k_{m\to S}^* = \sum_{s=1}^{S} \frac{x_m y_s}{1 + x_m y_s}, \ \forall\, m \qquad (8)$$

$$k_{s\to M}^* = \sum_{m=1}^{M} \frac{x_m y_s}{1 + x_m y_s}, \ \forall\, s \qquad (9)$$

ensuring that the empirical value of each constraint matches its expectation. The system above has been solved by running the NEMTROPY package [75] available at https://github.com/nicoloval/NEMtropy.

2.4. Projection of the 'MarkLines Automotive' dataset

The projection of a bipartite network onto one of its layers yields a monopartite graph, where any two nodes are connected as the number of common neighbors, proxying their similarity [86,87], is significantly large. Here, we follow the approach proposed in [73,74] to validate the similarity of any two nodes with respect to the BiCM. Schematically, this method works by (A) focusing on a specific pair of nodes belonging to the layer of interest and counting the number of common neighbors; (B) quantifying its statistical significance with respect to the BiCM; (C) linking the two nodes if, and only if, the corresponding p-value is sufficiently low. Let us, now, describe these steps in detail.

Quantifying nodes similarity. The simplest indicator of the similarity of two nodes belonging to the same layer of a bipartite network is provided by the number of their common neighbors. For the couple of manufacturers m and m' this number is given by

$$V_{mm'}^* = \sum_{s=1}^{S} l_{ms} l_{m's}, \qquad (10)$$

i.e. the number of V-motifs the two nodes originate (as m and m' cannot be directly connected, the presence of a common supplier, belonging to the opposite layer, draws a V-like shape [73]).

Statistical significance of nodes similarity. The BiCM, as any ERG model induced by linear constraints, treats links as independent random variables. Therefore, the presence of a common supplier for any two manufacturers m and m', i.e. $l_{ms} l_{m's} = 1$, can be described as the outcome of a Bernoulli trial whose probability coefficients read

$$f_{\mathrm{Ber}}(l_{ms} l_{m's}) = p_{ms} p_{m's}, \qquad (11)$$

$$f_{\mathrm{Ber}}(l_{ms} l_{m's} = 0) = 1 - p_{ms} p_{m's}. \qquad (12)$$

As $V_{mm'}$ is a sum of independent Bernoulli trials, each characterized by a different probability, the behavior of such a random variable is described by the Poisson-Binomial (PB) distribution [73]. Evaluating the statistical significance of the similarity of nodes m and m', thus, amounts to computing the p-value

$$\mathrm{p\text{-}value}(V_{mm'}) = \sum_{x \geq V_{mm'}^*} f_{\mathrm{PB}}(x). \qquad (13)$$

Validating the monopartite projection. P-values must, then, be validated by implementing a procedure for testing multiple hypotheses at a time. Several alternatives are viable, among which the Bonferroni correction [89], the Holm–Bonferroni correction and the Benjamini–Hochberg correction [90]. Here, we opt for the third one, controlling for the so-called False Discovery Rate (FDR), i.e. the expected proportion of false positives to appear within the set of tests that pass the
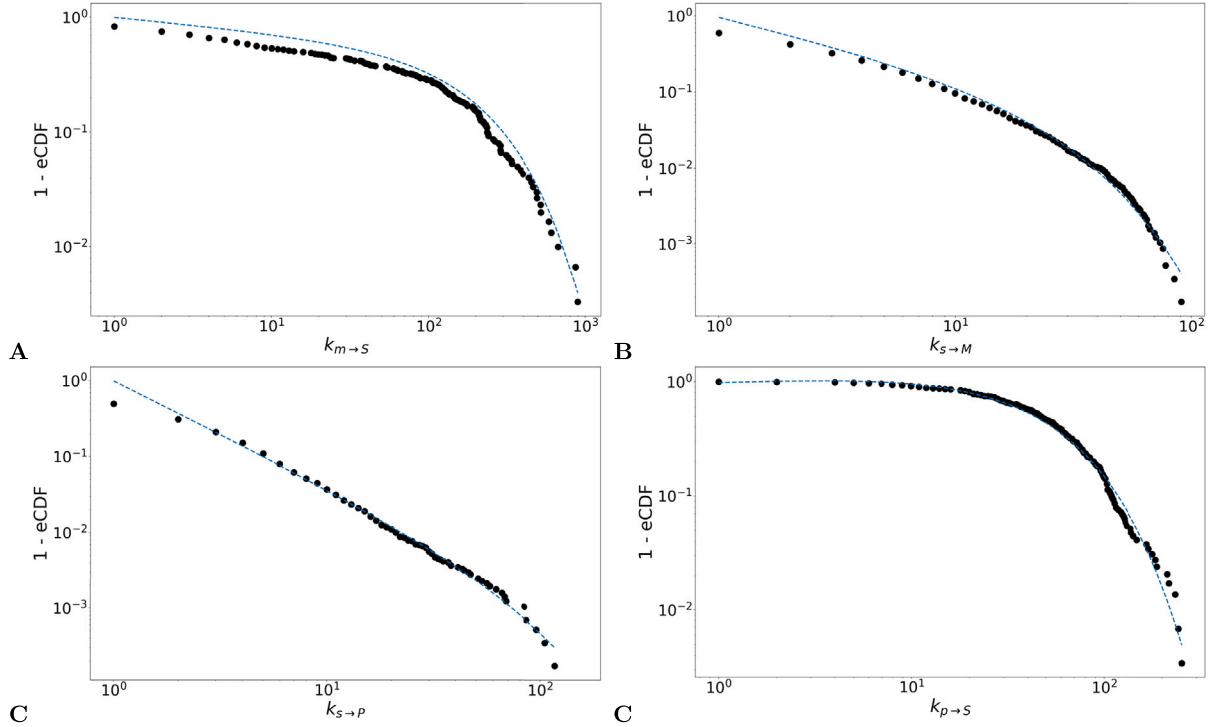
Fig. 2. Empirical cumulative density function of manufacturers' degrees (**A**), suppliers' 'left' degrees (**B**), suppliers' 'right' degrees (**C**) and products degrees (**D**). All of them are heavy-tailed and right-skewed, highlighting the large heterogeneity of the connectivity of nodes. More specifically, they all obey a power-law with exponential cutoff whose functional form reads $f(x) = x^{-\alpha} e^{-\beta x} \beta^{1-\alpha} / \Gamma[1-\alpha, \beta x_{\min}]$ and whose parameters read $\alpha = 0.134$, $\beta = 0.005$ (top left), $\alpha = 0.743$, $\beta = 0.049$ (top right), $\alpha = 1.276$, $\beta = 0.015$ (bottom left), $\alpha = 0.063$, $\beta = 0.022$ (bottom right). These values have been determined by implementing the Levenberg–Marquardt algorithm [88] for least squares optimization through the Python library SciPy. Overall, the above results indicate that 'generalist' suppliers co-exist with 'specialist' suppliers — noticeably, $\simeq 51\%$ of suppliers sells only one product.

validation. Thus, we sort the $n = M(M-1)/2$ p-values in increasing order

$$\text{p-value}_1 \leq \text{p-value}_2 \leq \cdots \leq \text{p-value}_n \tag{14}$$

and, then, individuate the largest integer i satisfying the condition

$$\text{p-value}_i \leq \frac{it}{n} \tag{15}$$

where t represents the single-test significance level, set to the value of 0.01. The FDR procedure prescribes to reject the null hypothesis for all pairs of nodes whose p-value is less than, or equal to, p-value$_i$, meaning that their similarity is considered statistically significant — hence, not explainable by constraining (just) the degrees. The corresponding nodes are, thus, linked in the resulting, monopartite projection.

2.5. Community detection

In order to detect the presence of communities partitioning the node set of our monopartite, validated projections, we employ the popular, modularity-based Louvain algorithm [91]. Modularity is a score function that assesses the quality of a given partition by comparing the number of internal links with the one expected under a given null model: the Louvain algorithm implements a heuristic exploration of the landscape of partitions, individuating the one maximizing it.

Although faster and more accurate than other methods, the Louvain algorithm is sensitive to the order in which nodes are selected [92,93]. This limitation can be overcome by running it several times, each time considering nodes in a different order: the best partition will be, again, the one attaining the largest value of modularity.

To detect bipartite communities, instead, we employ the BiLouvain algorithm, an extension of the Louvain algorithm individuating the partition maximizing Barber's bipartite modularity [94].

3. Results and discussion

3.1. Structural properties

Local connectivity. Fig. 2 shows the degree distributions of manufacturers, suppliers and products. All of them are heavy-tailed and right-skewed, an evidence pointing out the large heterogeneity of nodes connectivity: more specifically, they all obey a power-law with exponential cutoff [95].

Overall, the above results indicate the presence of 'generalist' suppliers (i.e. providing many products) co-existing with 'specialist' suppliers (i.e. providing few products; $\simeq 51\%$ of suppliers sells only one product). Fig. 3 further shows that the number of client manufacturers of a supplier and the number of different products it sells are positively correlated, indicating that 'generalist' suppliers tend to be connected with a large number of manufacturers. Still, many 'specialist' suppliers that are linked to a relatively large number of manufacturers exist; a noticeable exception is *Motor Super*, a Russian company that sells 56 different products to just 2 manufacturers, i.e. the Russian *AvtoVaz* and the American *Chevrolet*. Overall, $\simeq 41\%$ of suppliers is connected to only one manufacturer.

Assortativity. In order to inspect the presence of degree correlations, we have scattered the ANND$_{m \rightarrow S}$ values versus the $k_{m \rightarrow S}$ values (Fig. 4**A**) and the ANND$_{s \rightarrow \mathcal{M}}$ values versus the $k_{s \rightarrow \mathcal{M}}$ values (Fig. 4**B**). As the plots show, both trends are slightly decreasing, pointing out the presence of disassortative patterns: in other words, manufacturers with many suppliers tend to be connected with suppliers having few customers, while manufacturers with few suppliers tend to be connected with suppliers having many customers; similarly, suppliers with many clients tend to be connected with manufacturers having few suppliers, while suppliers with few clients tend to be connected with manufacturers having many suppliers. Overall, then, the automotive industry resembles an ecosystem where suppliers serving (only) 'bigger players', by providing
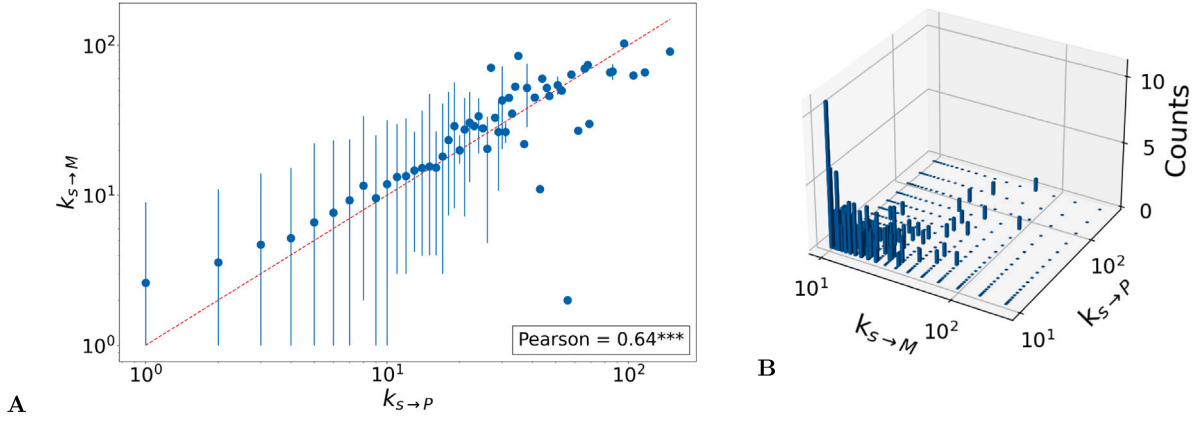
Fig. 3. A: average number of customers $k_{s\to\mathcal{M}}$ for suppliers with diversification $k_{s\to\mathcal{P}}$. The two quantities are positively correlated (the Pearson correlation coefficient reads $r \simeq 0.64$ with a p-value smaller than 0.001), indicating that 'generalist' suppliers tend to be connected with a large number of manufacturers; still, many suppliers providing few products while being connected with a relatively large number of manufacturers exist. The bars encode the 95% of the sample distribution while the red, dashed line is the identity. **B**: 3D histogram of $k_{s\to\mathcal{M}}$ versus $k_{s\to\mathcal{P}}$ (displayed for the values $k_{s\to\mathcal{M}} > 5$).

them few products – the so-called 'specialists' – co-exist with suppliers serving a larger number of clients, by providing them a larger basket of products — the so-called 'generalists'.

To spot differences between firms induced by their geographical localization, we have partitioned them into nine groups, i.e. African, Asean (i.e. Indonesia, Malaysia, South Korea, Thailand, and Vietnam), Chinese, Indian, Japanese, Middle Eastern (i.e. Egypt, Iran, and Turkey), Russian, Western (i.e. Australia, Europe, Israel, and US) and Joint Ventures (JVs) and colored them accordingly. Overall, the group of Western firms is quite well-distinguished from the group of Chinese firms and JVs. Among the manufacturers, Western ones display significantly large $\text{ANND}_{m\to S}$ values, lying in the top 0.5% of the ensemble distribution induced by the null model, while the $\text{ANND}_{m\to S}$ values for JVs, Chinese and Japanese manufacturers are either in line with the predictions of the null model or lie in the bottom 0.5% of the null distribution. For what concerns suppliers, instead, Chinese firms display significantly small $\text{ANND}_{s\to\mathcal{M}}$ values, lying in the bottom 0.5% of the null distribution while the $\text{ANND}_{s\to\mathcal{M}}$ values for Western and Japanese suppliers are in line with the predictions of the null model. This result suggests the Western and Japanese supply chains to be structurally different from the Chinese ones: Western and Japanese manufacturers tend to purchase products from suppliers whose degree is, on average, larger than the one of the suppliers serving Chinese manufacturers. In particular, the Japanese automotive industry revolves around few, big manufacturers purchasing products from many, low-degree suppliers.

For what concerns the supplier/product network, scattering the $\text{ANND}_{s\to\mathcal{P}}$ values versus the $k_{s\to\mathcal{P}}$ values and the $\text{ANND}_{p\to S}$ values versus the $k_{p\to S}$ values reveals its disassortative character, with suppliers selling many, less ubiquitous products and viceversa ((Fig. 4C–D)). This result is reminiscent of the one concerning the export of countries within the (bipartite representation of the) international trade. While firms do not seem to be partitioned according to any geographical criterion, products belonging to the Electrical sector display the larger $\text{ANND}_{p\to S}$ values.

The aforementioned, geographical difference is recovered when scattering the tripartite assortativity coefficient of each manufacturer, defined as $\text{ANND}_{m\to\mathcal{P}} = \sum_{s=1}^{S} l_{ms} k_{s\to\mathcal{P}}/k_{m\to S}$, versus its degree. As Fig. 4E confirms, Western and Japanese manufacturers tend to connect with suppliers providing a number of products that is larger than the one provided by the suppliers to which Chinese manufacturers and JVs tend to connect.

Motifs. For what concerns the bipartite clustering coefficient, both panels of Fig. 5 show an overall decreasing trend, signaling that manufacturers (suppliers) having a large degree are generally closing less

squares than manufacturers (suppliers) having a small degree. Again, the group of Western and Japanese firms is quite well-distinguished from the group of Chinese firms and JVs, as the former ones display significantly large values of the bipartite clustering coefficient, lying in the top 0.5% of the ensemble distribution induced by our null model. In other words, the presence of mesoscale structures characterizing the Western and Japanese subsets of firms cannot be simply traced back to the degrees of nodes: rather, it represents a peculiar feature of these areas whose automotive industry appears as a highly interconnected ecosystem. Chinese firms, on the contrary, constitute a seemingly fragmented environment with different (sets of) manufacturers purchasing products from different (sets of) suppliers, each one serving few clients. Our results complement the ones about the presence of statistically significant functional structures within the automotive industry [68], clarifying that they come along a geographical signature. These results are robust with respect to the definition of the bipartite clustering coefficient (see Appendix B).

Subgraph centrality. The results of the analysis of the bipartite subgraph centrality (BSC), illustrated in Fig. 6, show that nodes with a larger degree tend to be more central as well. More importantly, the BSC allows us to appreciate the different behavior displayed by firms located in different countries at best: Chinese firms and JVs are, in fact, characterized by values of the BSC that are much smaller (some hardly above zero) than the values of the BSC characterizing Western and Japanese firms. Specifically, the latter (former) have a larger-than-expected (smaller-than-expected) BSC on the layer of manufacturers. Similar trends are observed when considering the layer of suppliers, the only difference being that, now, the BSC of Western and Japanese firms is, overall, in line with the predictions of the BiCM. Once again, these findings reveal Western and Japanese firms to be crossed by a large number of patterns, an evidence confirming the plethora of interconnections leading from one node to another within this portion of the network. Chinese firms, instead, seemingly belong to less interconnected, if not completely segregated, supply chains.

Suppliers redundancy. We complement our analysis by calculating, for each manufacturer m, the average redundancy of its suppliers, defined as $\text{ASR}_m = \sum_{s=1}^{S} l_{ms}\text{RED}_s/k_{m\to S}$ where $\text{RED}_s = \sum_{p=1}^{P}\sum_{s'(\neq s)\in S_m} r_{sp}r_{s'p}/k_{s\to\mathcal{P}}$ and with S_m indicating the neighborhood of manufacturer m, i.e. the set of its suppliers. Scattering the ASR_m values versus the $k_{m\to S}$ values (see Fig. 7) reveals an increasing trend, pointing out that suppliers serving manufacturers with a larger degree tend to produce a larger number of similar products. Besides, the suppliers serving Western manufacturers display significantly large redundancy values,
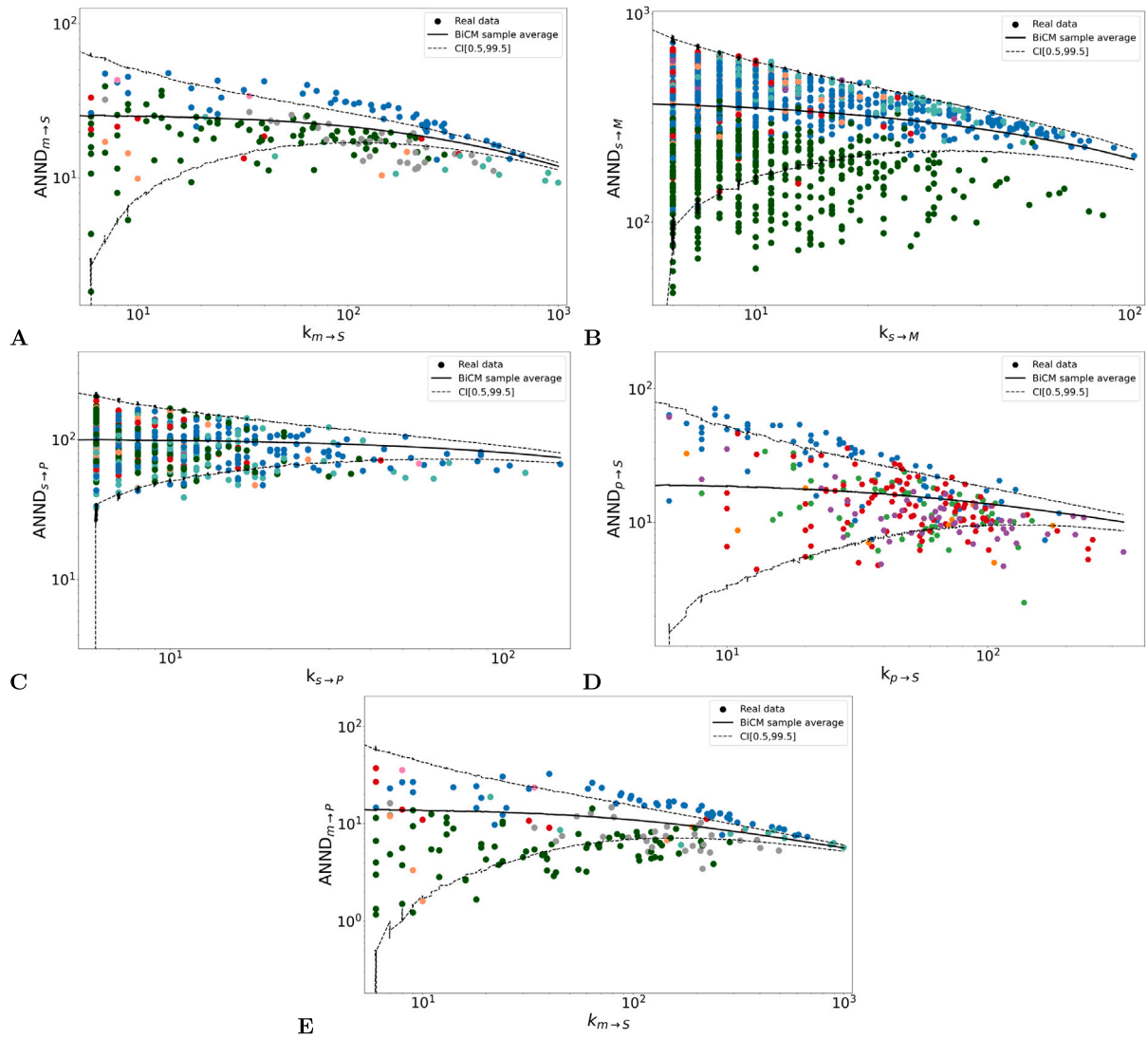
Fig. 4. Assortativity profiles of our networks. Top panels: $ANND_{m \to S}$ scattered versus $k_{m \to S}$ (**A**) and $ANND_{s \to \mathcal{M}}$ scattered versus $k_{s \to \mathcal{M}}$ (**B**). Both trends are slightly decreasing, hinting at a disassortative behavior, i.e. manufacturers (suppliers) having a large degree are generally connected with suppliers (manufacturers) having a small degree and viceversa. In other words, our results depict the automotive industry as an ecosystem populated by (1) suppliers that serve 'bigger players' by providing them few products and (2) suppliers that serve a larger number of clients by providing them a larger basket of products. Middle panels: $ANND_{s \to P}$ scattered versus $k_{s \to P}$ (**C**) and $ANND_{p \to S}$ scattered versus $k_{p \to S}$ (**D**). As for the manufacturer-supplier network, both trends are decreasing, pointing out a disassortative behavior, according to which suppliers providing many products tend to have exclusive products within their baskets, whereas suppliers providing few products tend to have a basket of ubiquitous products (in this case, however, no geographically-induced distinction between firms emerges). **E**: $ANND_{m \to P}$ scattered versus $k_{m \to S}$. The slightly decreasing trend confirms the picture provided by the figures above, i.e. Western and Japanese manufacturers with a larger degree tend to be connected with suppliers providing fewer products and viceversa. Nodes are colored according to their geographical localization: • Africa, • Asean, • Chinese, • Indian, • Japanese, • Latin America, • Middle East, • Russian, • Western and • Joint Ventures. Products are colored according to their technological sector, i.e. • Chassis/Body, • Electrical, • Powertrain, • Interior/Exterior, • General parts.

lying in the top 0.5% of the ensemble distribution induced by our null model. The size of the dots is proportional to the geographical membership of the suppliers of each manufacturer: while the suppliers serving Western and Japanese manufacturers are scattered across many different countries - e.g. 36 for *Toyota* and 34 for *Ford* - this is true to a much lesser extent for Chinese manufacturers and JVs - e.g. the number of countries hosting the suppliers of *Geely* and *FAW Volkswagen* is 10 and 15, respectively.

Nestedness. Nestedness has been shown to characterize the 'MarkLines Automotive' dataset [96–98]. Our analysis, however, returns a quite different picture, showing that **L** and **R** display smaller-than-expected values of the NODF (see Table 2). Left aside the differences concerning the data (the authors of [97] consider 2.474 manufacturers, 16.468 suppliers and 934 products), we argue this discrepancy to be traceable back to (1) the way nestedness is computed (namely, by picking random

subsamples composed of 50 rows [97]), (2) the different randomization strategy implemented.

In order to further inspect the mesoscopic structure of our networks, we looked for the presence of nested communities. To this aim, we, first, performed a bipartite community detection on **L** and **R** and, then, measured the in-block nestedness. Both the manufacturer/supplier and the supplier/product networks display a marked, bipartite community structure, with bi-modularity amounting at $Q = 0.51$ and $Q = 0.57$, respectively. The former contains 10 communities, the five, biggest ones clearly identifying the Asean-Indian, Chinese, European, Japanese and US clusters (see Fig. 8); the latter contains 19 communities, representing coherent sets of products. In both cases, the in-block NODF is characterized by larger-than-expected values (see Table 2). If, instead, we test nestedness against the block-specific version of the BiCM, we recover smaller-than-expected values.
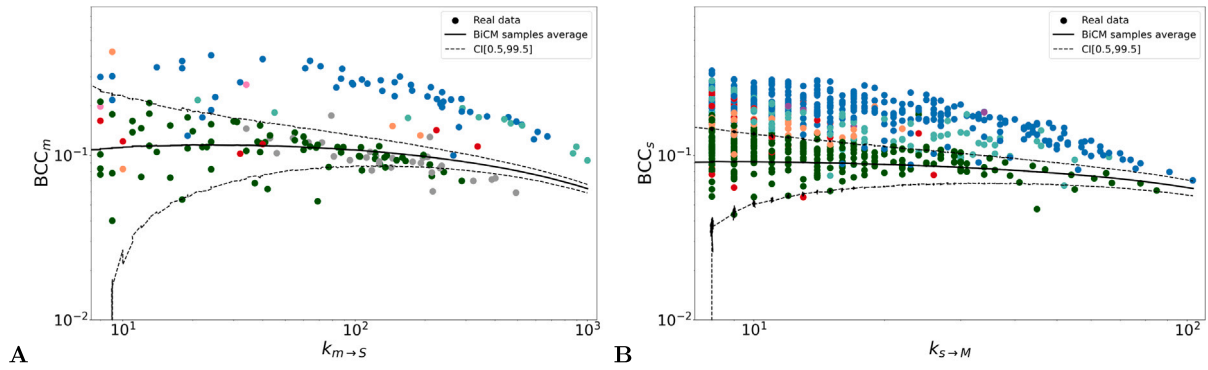
Fig. 5. Bipartite clustering coefficient of manufacturers BCC_m scattered versus the number of their suppliers $k_{m \to S}$ (**A**) and bipartite clustering coefficient of suppliers BCC_s scattered versus the number of their client manufacturers $k_{s \to M}$ (**B**). Both trends are overall decreasing, indicating that manufacturers (suppliers) having a large degree are generally closing less squares than manufacturers (suppliers) having a small degree. Besides, our results highlight that the automotive industry of different geographical locations is shaped by different organizing principles. The Western ecosystem seems to be highly interconnected, as manufacturers share many suppliers that, in turn, share many different manufacturers. Instead, the Chinese ecosystem seems to be rather fragmented, with different (sets of) manufacturers purchasing products from different (sets of) suppliers, each one serving few clients. Overall, our results confirm the presence of the (statistically significant) 'functional' structures first observed in [68], however highlighting their peculiarly geographical character. Nodes are colored according to their geographical localization: • Africa, • Asean, • Chinese, • Indian, • Japanese, • Latin America, • Middle East, • Russian, • Western and • Joint Ventures.
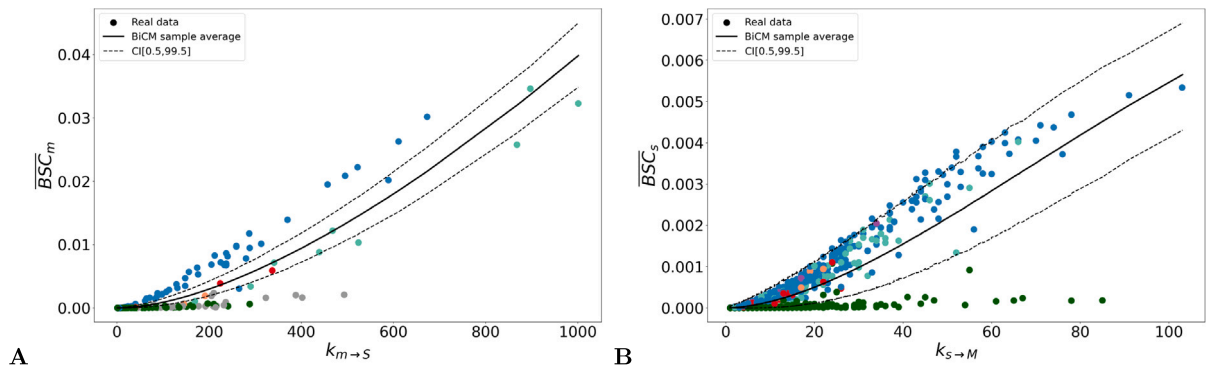
Fig. 6. Bipartite subgraph centrality of manufacturers $\overline{BSC}_m$ scattered versus the number of their suppliers $k_{m \to S}$ (**A**) and bipartite subgraph centrality of suppliers $\overline{BSC}_s$ scattered versus the number of their client manufacturers $k_{s \to M}$ (**B**). Both trends are increasing, pointing out that nodes with a larger degree are also more central. Chinese firms and JVs are characterized by values of the BSC that are much smaller than those characterizing Western and Japanese firms and always smaller-than-expected under the BiCM, a finding seemingly confirming that Chinese firms belong to less interconnected, if not completely segregated, supply chains. Western and Japanese firms, instead, appear to be crossed by a large number of patterns – actually, larger-than-expected under the BiCM, when considering the layer of manufacturers – an evidence confirming the plethora of interconnections leading from one node to another within this portion of the network. Nodes are colored according to their geographical localization: • Africa, • Asean, • Chinese, • Indian, • Japanese, • Latin America, • Middle East, • Russian, • Western and • Joint Ventures.

### 3.2. Projections of the 'MarkLines Automotive' production network

Community structure of the validated projections. Lastly, we focus on the validated projection onto each of our layers. Fig. 8A shows the validated network of manufacturers, where any two nodes are linked if sharing a significantly large number of suppliers. A markedly modular structure emerges (the value of modularity amounts to $Q = 0.50$), with clusters of manufacturers that are coherent with their geographical localization. Indeed the two, largest ones are those composed by Chinese and Western firms — which, however, are not interconnected. The Chinese cluster, instead, is linked only to the community of JVs, constituted by Joint Ventures involving one Chinese company. In turn, the Western cluster is connected with the cluster of JVs via the manufacturer *BMW Brilliance*, with the Japanese cluster via the manufacturer *Nissan* and with the Asean-Indian cluster via several links. Interestingly, the similarity of the two, small sets of nodes (recognized by the Louvain algorithm as individual communities) which are detached from the Chinese and Western clusters is not only due to geographical proximity but also to their technological characterization. Indeed, the three nodes lying on the left of the Chinese cluster (i.e. *GAC Aion*, *Weltmeister* and *Xpeng*) are all manufacturers of electric cars, while the four nodes lying below the Western cluster (i.e. *Alpina*, *Lotus*, *McLaren* and *SRT*) are all manufacturers of sportive cars. We stress once more that the Western cluster is much more internally-connected than the Chinese one, a finding further confirming that the closure of motifs, within the automotive industry, is strongly driven by geographical proximity.

Fig. 9 shows the validated projection of suppliers, where any two nodes are linked if sharing a significantly large number of manufacturers (see Appendix C). The network is markedly modular as well (the value of modularity amounts to $Q = 0.55$) with clusters embodying geographical information. As evident upon inspecting the figure, the Chinese cluster is not only isolated but also (internally) fragmented, being constituted by a plethora of smaller connected components. The Indian cluster is disconnected from the rest of the network as well, although its density is quite large. The largest components are constituted by two pairs of interconnected communities, i.e. the one gathering Asean and Japanese firms and the one gathering American and European firms. German suppliers give origin to a smaller community, lying on the right of the European subgraph.

For what concerns the supplier/product network, the validated projection of products (linked if sharing a significantly large number of suppliers) is shown in Fig. 10: it displays an interesting community structure (the value of modularity amounts to $Q = 0.45$), characterized by the emergence of clusters that do not overlap with the
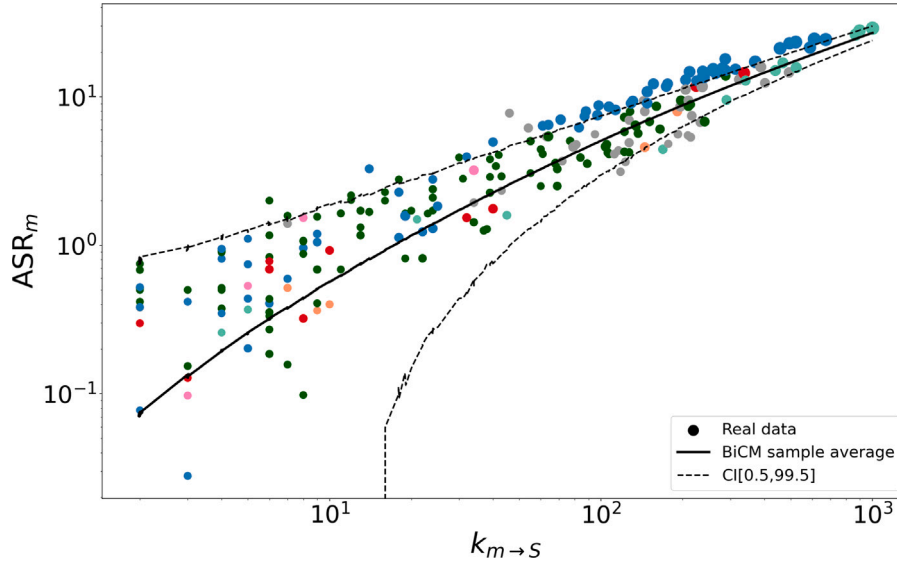
Fig. 7. Average suppliers redundancy of manufacturers ASR$_m$ scattered versus the number of their suppliers $k_{m \to S}$. The increasing trend signals that suppliers serving manufacturers with a larger degree tend to produce a larger number of similar products. The suppliers serving Western manufacturers display the larger redundancy values; the size of dots, proportional to the number of countries of the suppliers of each manufacturer, highlight a geographical signature: the suppliers serving Western and Japanese manufacturers are scattered across many, different countries (36 for *Toyota* and 34 for *Ford*) while this is true to a much lesser extent for what concerns the suppliers serving Chinese manufacturers and JVs (the number of countries hosting the suppliers of *Geely* and *FAW Volkswagen* are 10 and 15, respectively). Nodes are colored according to their geographical localization: • Africa, • Asean, • Chinese, • Indian, • Japanese, • Latin America, • Middle East, • Russian, • Western and • Joint Ventures.

Table 2
General and in-block nestedness of the manufacturer/supplier and supplier/product networks, along with their expected values, standard deviations and z-scores, computed under the BiCM. While the general NODF is significantly smaller-than-expected, the opposite is true for its in-block version. Communities l_6 to l_{10} and r_{15} to r_{19} are not shown as they are composed by just one or two nodes.

	NODF	⟨NODF⟩	σ_{NODF}	z-score
L_{ms}	**0.096**	**0.120**	**0.001**	**−14.542**
l_1	0.119	0.077	0.002	17.941
l_2	0.407	0.190	0.010	22.288
l_3	0.424	0.244	0.007	24.462
l_4	0.293	0.137	0.013	11.742
l_5	0.219	0.076	0.017	8.484
R_{sp}	**0.023**	**0.031**	**0.0004**	**−19.859**
r_1	0.086	0.073	0.005	2.599
r_2	0.137	0.062	0.007	10.741
r_3	0.078	0.036	0.010	4.314
r_4	0.085	0.034	0.017	2.975
r_5	0.124	0.049	0.012	6.084
r_6	0.345	0.038	0.033	9.254
r_7	0.308	0.068	0.050	4.788
r_8	0.123	0.055	0.020	3.464
r_9	0.112	0.041	0.015	4.848
r_{10}	0.168	0.085	0.030	2.777
r_{11}	0.130	0.026	0.024	4.244
r_{12}	0.177	0.044	0.023	5.686
r_{13}	0.396	0.082	0.035	8.960
r_{14}	0.215	0.037	0.033	5.405

(technological) taxonomy provided by the MarkLines platform. Rather, they represent different car systems (i.e. groups of devices targeting specific tasks — see Appendix D), an evidence suggesting that suppliers tend to focus their production on 'coherent' sets of products [99,100]. Inspecting the nationality of the suppliers providing these products (see Appendix E) reveals that China, Germany, Japan and USA are the most represented countries. China, however, is extremely active in some communities (e.g. *Airbags, Engine, Exhaust System, Wheel*) while being completely absent from others (e.g. *Electronic Control Units*); the percentage of suppliers of Germany, Japan and USA, instead, is similar across all communities.

4. Conclusions

The recent pandemic has fostered research on the structure of global supply chains [18,22]. The use of tools routinely employed in economics [101] has allowed scholars to approach a number of relevant, economic and geopolitical issues, such as the consequences of different, national development strategies on the global value chain of low-carbon technologies [102] and the effects of the US–China trade war [103]. Yet, these discussions have been typically carried out from a qualitative, or merely statistical, point of view. Additionally, commonly use databases such as input–output tables [104] are unable to resolve interactions at firm-level, since focusing on whole sectors in specific, geographical areas. The aim of the present paper is bridging this gap, carrying out a quantitative investigation of the large-scale structure of a specific industry by employing methods rooted into statistical physics. Our analysis of the 'MarkLines Automotive' dataset allows us to draw a number of conclusions.

- As we obtain network projections by linking any two nodes in case the number of their common neighbors is significantly large and such a number is proportional to the number of X-motifs, we confirm the prominent role played by square patterns in shaping the structure of production networks [68]: firms with similar outputs are often engaged in a competition, rather than in a buyer–supplier relationship, although they can share many suppliers and customers.

- Degree distributions are heavy-tailed, hence describing an ecosystem where nodes are characterized by a highly heterogeneous number of neighbors. More explicitly, 'generalist' suppliers selling many products co-exist together with 'specialist' suppliers selling few products. Besides, 'generalist' suppliers tend to be connected with a large number of small-degree manufacturers while 'specialist' suppliers tend to be connected with a small number of large-degree manufacturers. Specifically, the suppliers selling only one product (amounting to $\simeq 51\%$ of the total) are connected with a subset of 166 manufacturers whose average degree amounts to $\simeq 151$, i.e. almost twice the manufacturers average degree - a result indicating that well-connected manufacturers are preferentially connected with suppliers seemingly providing 'exclusive' services.
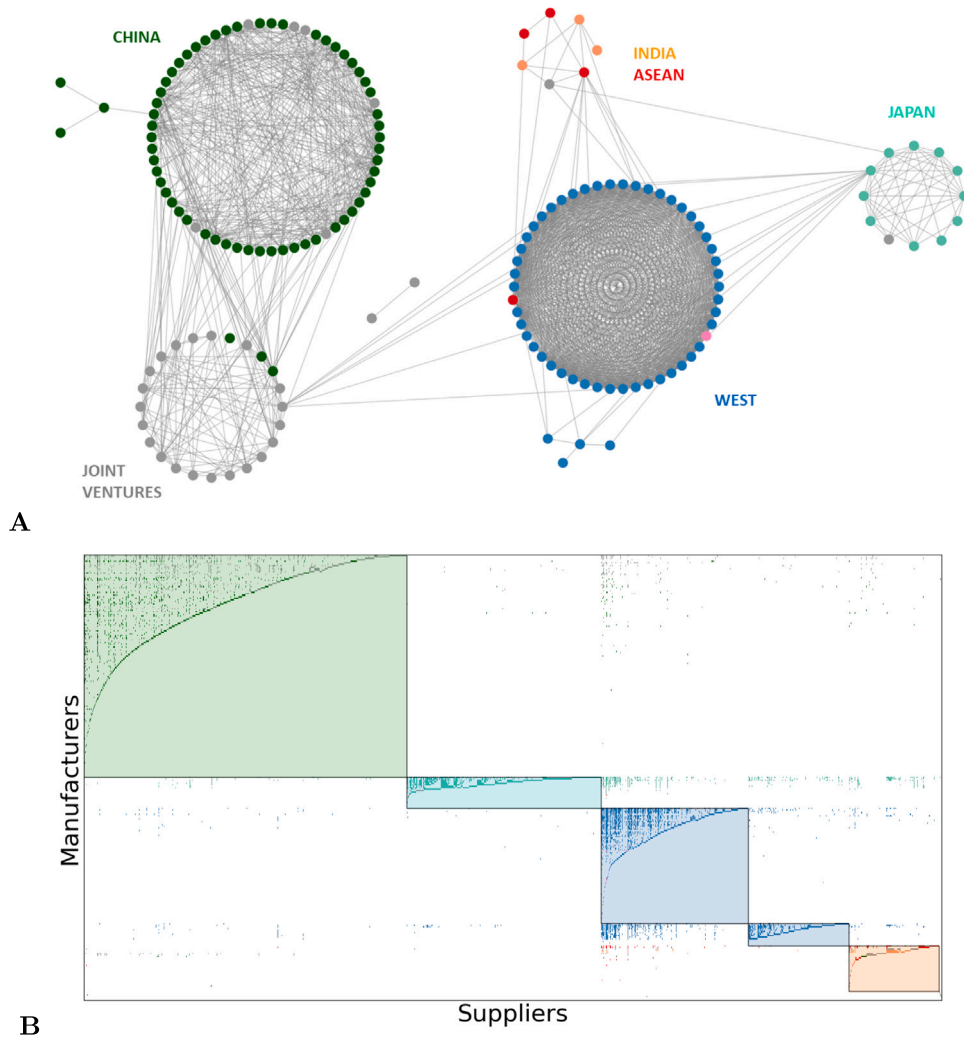
Fig. 8. A: validated projection onto the layer of manufacturers, linked if sharing a significantly large number of suppliers. Analyzing the presence of connections between communities reveals that the Chinese cluster is solely linked to the cluster of JVs which, in turn, is connected with the Western cluster via the manufacturer *BMW Brilliance*. As the latter is connected with the Asean-Indian and Japanese clusters as well, it is also the most central one of our validated projection. Lastly, the Western cluster is much more internally connected than the Chinese one, a finding further confirming the closure of motifs to be strongly driven by geographical proximity. **B**: biadjacency matrix of the manufacturer/supplier network, with nodes partitioned according to the communities spotted by the BiLouvain algorithm and reordered according to the 'Economic Fitness and Complexity' algorithm [81]. The triangularity of each module confirms the in-block nested structure of the network. Notice the overlap between the two sets of communities. Nodes are colored according to their geographical localization: • Africa, • Asean, • Chinese, • Indian, • Japanese, • Latin America, • Middle East, • Russian, • Western and • Joint Ventures.

- The analysis of the bipartite subgraph centrality points out that both manufacturers and suppliers with a larger degree tend to be more central, although the abundance of squares shows a decreasing trend when plotted as a function of the degree. These two quantities provide the neatest, geography-based partition of our basket of firms, separating Chinese from Western ones: the former close a smaller number of squares than the latter, hence providing an indication that the Chinese ecosystem is less integrated than the Western one. This is particularly evident when considering the set of Chinese suppliers: the limited, although statistically significant, number of neighbors they share induces a projection that is made of several disconnected components. Although this may be (at least, partly) induced by the peculiar way this projection has been obtained, our conclusions are supported by the (methodologically different) analysis carried out in [105], which finds that the Chinese automotive industry is characterized by a large degree of internal fragmentation - in play even at a territorial level, as the recovered components correspond to different, Chinese provinces. This evidence suggests

the Chinese industrial organization to be more distributed than the Western one, maybe to ensure all provinces a comparable level of economic development.

- The information provided by redundancy complements the one provided by geography, clarifying that the presence of (many) suppliers providing the same products to a manufacturer is compatible with a broader geographical distribution of the manufacturers plants. By converse, the presence of (fewer) suppliers selling different products to a manufacturer is compatible with a narrower geographical distribution of the manufacturers plants — in certain cases, a very local one. This result also suggests the manufacturers belonging to the first group are more resilient than the ones belonging to the second group, as their production is apparently less prone to interruptions due to supply shortages.

- Projecting on the layer of manufacturers reveals that the Chinese cluster is solely connected with the cluster of Chinese JVs, a result indicating that Chinese manufacturers do not share (a significantly large number of) suppliers with other manufacturers. Again, this is supported by the analysis carried out in [105],
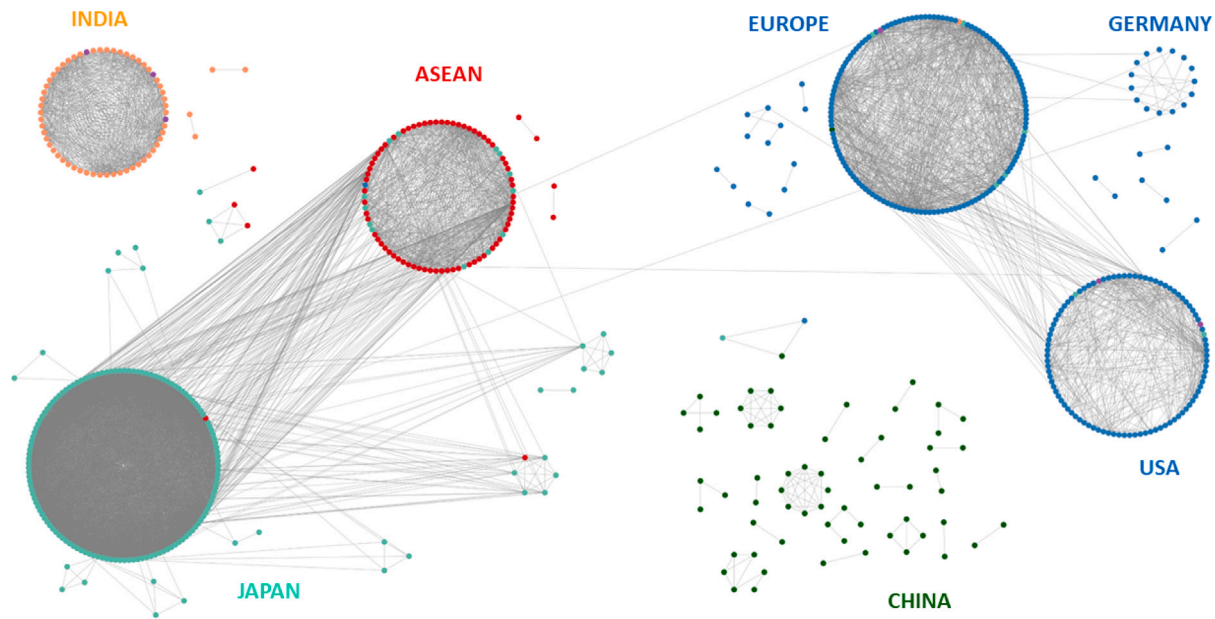
Fig. 9. Validated projection of suppliers, linked if sharing a significantly large number of manufacturers. Analyzing the presence of connections between communities reveals that the Chinese cluster is isolated and internally fragmented. The largest components are constituted by two pairs of interconnected communities, formed by Asean and Japanese firms and American and European firms. Nodes are colored according to their geographical localization: • Africa, • Asean, • Chinese, • Indian, • Japanese, • Latin America, • Middle East, • Russian, • Western and • Joint Ventures.
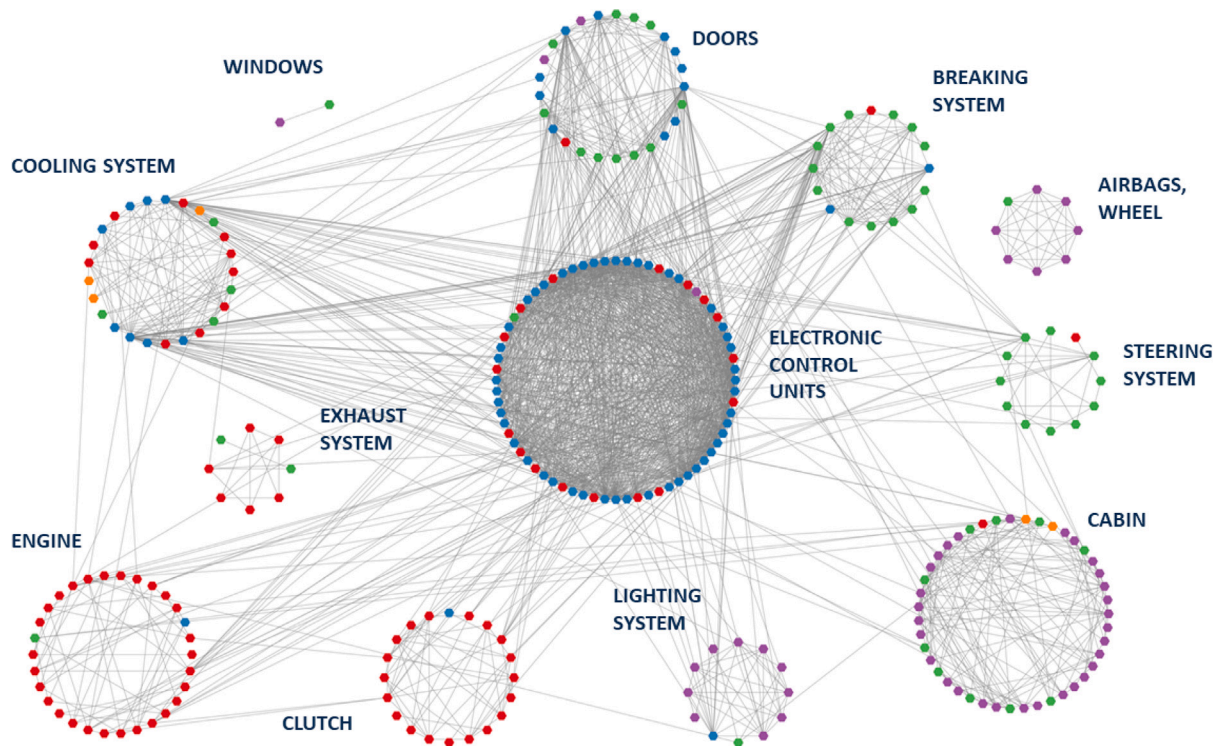

Fig. 10. Validated projection of products, linked if sharing a significantly large number of suppliers. The network displays a clearly defined community structure whose clusters represent different car systems. Nodes are colored according to their technological sector, i.e. • Chassis/Body, • Electrical, • Powertrain, • Interior/Exterior, • General parts.

where it has been observed that the Chinese government adopted protectionist policies to boost the development of an indigenous automotive industry, forcing Chinese JVs to buy the $40\% - 80\%$ of their components from Chinese suppliers.

Of course, our study comes with limitations: we have only considered the data of a single industry (which may be incomplete as a result of the data collection procedure); our dataset has been represented in a bipartite fashion while, in general, production networks have a less neat structure; we have employed a null model embodying the information encoded into the node degrees although more refined benchmarks can be used; we have employed the Louvain algorithm which is one out of many alternative choices for the task of detecting communities (however, known to perform satisfactorily in presence of few, large communities).

As a last remark, we would like to stress that the investigation of the economic, historical, and social motivations at the origin of the geography-related structures we found is beyond the scope of the present paper. Here, we have shown that adopting tools from network theory can lead to the discovery of clear, structural features (in this case, concerning the automotive sector) that should not be overlooked during the model-building phase. These patterns, in fact, could inform methods for reconstructing production networks [106] - which aim at overcoming the limitations affecting available datasets by adopting statistically-grounded techniques [62], machine learning tools [55,57, 107,108] or proper data proxies [109] - to be later employed for stress testing [39,40,110].

CRediT authorship contribution statement

Massimiliano Fessina: Conceptualization, Data curation, Formal analysis, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Investigation. **Andrea Zaccaria:** Conceptualization, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing, Formal analysis, Visualization, Funding acquisition. **Giulio Cimini:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing, Visualization. **Tiziano Squartini:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chaos.2024.114630.

References

[1] Serrano MÁ, Boguñá M. Topology of the world trade web. Phys Rev E 2003;68:015101.

[2] Garlaschelli D, Loffredo MI. Structure and evolution of the world trade network. Physica A 2005;355(1):138–44.

[3] Saracco F, Di Clemente R, Gabrielli A, Squartini T. Randomizing bipartite networks: the case of the world trade web. Sci Rep 2015;5(1):10595.

[4] Hidalgo CA, Hausmann R. The building blocks of economic complexity. Proc Natl Acad Sci 2009;106(26):10570–5.

[5] Tacchella A, Mazzilli D, Pietronero L. A dynamical systems approach to gross domestic product forecasting. Nat Phys 2018;14(8):861–5.

[6] Pugliese E, Cimini G, Patelli A, Zaccaria A, Pietronero L, Gabrielli A. Unfolding the innovation system for the development of countries: coevolution of science, technology and production. Sci Rep 2019;9(1):16440.

[7] Balland P-A, Broekel T, Diodato D, Giuliani E, Hausmann R, O'Clery N, et al. The new paradigm of economic complexity. Res Policy 2022;51(3):104450.

[8] Lin J-H, Primicerio K, Squartini T, Decker C, Tessone CJ. Lightning network: a second path towards centralisation of the bitcoin economy. New J Phys 2020;22(8):083022.

[9] Mattsson CES, Criscione T, Takes FW. Circulation of a digital community currency. Sci Rep 2023;13(1):5864.

[10] Boss M, Elsinger H, Summer M, Thurner 4 S. Network topology of the interbank market. Quant Finance 2004;4(6):677–84.

[11] Battiston S, Puliga M, Kaushik R, Tasca P, Caldarelli G. DebtRank: Too central to fail? Financial networks, the FED and systemic risk. Sci Rep 2012;2(1):541.

[12] Squartini T, Van Lelyveld I, Garlaschelli D. Early-warning signals of topological collapse in interbank networks. Sci Rep 2013;3(1):3357.

[13] Cimini G, Squartini T, Garlaschelli D, Gabrielli A. Systemic risk analysis on reconstructed economic and financial networks. Sci Rep 2015;5(15758).

[14] Bargigli L, di Iasio G, Infante L, Lillo F, Pierobon F. The multiplex structure of interbank networks. Quant Finance 2015;15(4):673–91.

[15] Chase-Dunn C, Kawano Y, Brewer BD. Trade globalization since 1795: Waves of integration in the world-system. Am Sociol Rev 2000;65(1):77–95.

[16] Craighead CW, Blackhurst J, Rungtusanatham MJ, Handfield RB. The severity of supply chain disruptions: Design characteristics and mitigation capabilities. Decis Sci 2007;38(1):131–56.

[17] Inoue H, Todo Y. Firm-level propagation of shocks through supply-chain networks. Nat Sustain 2019;2(9):841–7.

[18] Guan D, Wang D, Hallegatte S, Davis SJ, Huo J, Li S, et al. Global supply-chain effects of COVID-19 control measures. Nat Hum Behav 2020;4(6):577–87.

[19] Aldrighetti R, Battini D, Ivanov D, Zennaro I. Costs of resilience and disruptions in supply chain network design models: A review and future research directions. Int J Prod Econ 2021;235:108103.

[20] Carvalho VM, Nirei M, Saito YU, Tahbaz-Salehi A. Supply chain disruptions: Evidence from the great east Japan earthquake*. Q J Econ 2021;136(2):1255–321.

[21] Ivanov D, Tsipoulanidis A, Schönberger J. Supply chain risk management and resilience. In: Global supply chain and operations management: a decision-oriented introduction to the creation of value. Cham: Springer International Publishing; 2021, p. 485–520.

[22] Chowdhury P, Paul SK, Kaisar S, Moktadir MA. COVID-19 pandemic related supply chain studies: A systematic review. Transp Res E 2021;148:102271.

[23] Miller RE, Blair PD. Input-output analysis: foundations and extensions. Cambridge University Press; 2009.

[24] Morimoto Y. On aggregation problems in input-output analysis. Rev Econom Stud 1970;37(1):119–26.

[25] Diem C, Borsos A, Reisch T, Kertesz J, Thurner S. Estimating the loss of economic predictability from aggregating firm-level production networks. 2023, arXiv preprint arXiv:2302.11451.

[26] Bak P, Chen K, Scheinkman J, Woodford M. Aggregate fluctuations from independent sectoral shocks: self-organized criticality in a model of production and inventory dynamics. Ricerche Econ 1993;47(1):3–30.

[27] Gabaix X. The granular origins of aggregate fluctuations. Econometrica 2011;79(3):733–72.

[28] Acemoglu D, Carvalho VM, Ozdaglar A, Tahbaz-Salehi A. The network origins of aggregate fluctuations. Econometrica 2012;80(5):1977–2016.

[29] Choi TY, Dooley KJ, Rungtusanatham M. Supply networks and complex adaptive systems: control versus emergence. J Oper Manage 2001;19(3):351–66.

[30] Surana A, Kumara S, Greaves M, Raghavan UN. Supply-chain networks: a complex adaptive systems perspective. Int J Prod Res 2005;43(20):4235–65.

[31] Wycisk C, McKelvey B, Hülsmann M. "Smart parts"supply networks as complex adaptive systems: analysis and implications. Int J Phys Distrib Logist Manage 2008;38(2):108–25.

[32] Kim Y, Chen Y-S, Linderman K. Supply network disruption and resilience: A network structural perspective. J Oper Manage 2015;33–34(1):43–59.

[33] Brintrup A, Wang Y, Tiwari A. Supply networks as complex systems: A network-science-based characterization. IEEE Syst J 2017;11(4):2170–81.

[34] Perera S, Bell MGH, Bliemer MCJ. Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective. Appl Netw Sci 2017;2(1):33.

[35] Brintrup A, Ledwoch A. Supply network science: Emergence of a new perspective on a classical field. Chaos 2018;28(3):033120.

[36] Barrot J-N, Sauvagnat J. Input specificity and the propagation of idiosyncratic shocks in production networks *. Q J Econ 2016;131(3):1543–92.

[37] Demirel G, MacCarthy BL, Ritterskamp D, Champneys AR, Gross T. Identifying dynamical instabilities in supply networks using generalized modeling. J Oper Manage 2019;65(2):136–59.

[38] Demir B, Javorcik B, Michalski TK, Ors E. Financial constraints and propagation of shocks in production networks. Rev Econ Stat 2022;1–46.

[39] Diem C, Borsos A, Reisch T, Kertész J, Thurner S. Quantifying firm-level economic systemic risk from nation-wide supply networks. Sci Rep 2022;12(1):1–13.

[40] Fujiwara Y, Terai M, Fujita Y, Souma W. Debtrank analysis of financial distress propagation on a production network in Japan. 2016, RIETI Discussion Paper Series 16-E-046.

[41] Bardoscia M, Barucca P, Battiston S, Caccioli F, Cimini G, Garlaschelli D, et al. The physics of financial networks. Nat Rev Phys 2021;3(7):490–507.

[42] Bacilieri A, Borsos A, Astudillo-Estevez P, Lafond F. Firm-level production networks: what do we (really) know?. Tech. rep. 2023–08, INET Oxford Working Paper; 2023.

[43] Lee K-M, Yang J-S, Kim G, Lee J, Goh K-I, Kim I-m. Impact of the topology of global macroeconomic network on the spreading of economic crises. PLoS One 2011;6(3):e18443.

[44] Mizuno T, Ohnishi T, Watanabe T. Structure of global buyer-supplier networks and its implications for conflict minerals regulations. EPJ Data Sci 2016;5(1):2.

[45] Gephart JA, Rovenskaya E, Dieckmann U, Pace ML, Brännström Å. Vulnerability to shocks in the global seafood trade network. Environ Res Lett 2016;11(3):035008.

[46] Klimek P, Poledna S, Thurner S. Quantifying economic resilience from input–output susceptibility to improve predictions of economic growth and recovery. Nature Commun 2019;10(1):1677.

[47] Starnini M, Boguñá M, Serrano MÁ. The interconnected wealth of nations: Shock propagation on global trade-investment multiplex networks. Sci Rep 2019;9(1):13079.

[48] König MD, Levchenko A, Rogers T, Zilibotti F. Aggregate fluctuations in adaptive production networks. Proc Natl Acad Sci 2022;119(38):e2203730119.

[49] Atalay E, Hortaçsu A, Roberts J, Syverson C. Network structure of production. Proc Natl Acad Sci 2011;108(13):5199–202.

[50] Cohen L, Frazzini A. Economic links and predictable returns. J Finance 2008;63(4):1977–2011.

[51] Chakraborty A, Ikeda Y. Testing "efficient supply chain propositions"using topological characterization of the global supply chain network. PLoS One 2020;15(10):e0239669.

[52] Chakraborty A, Reisch T, Diem C, Thurner S. Inequality in economic shock exposures across the global firm-level supply network. 2021, arXiv preprint arXiv:2112.00415.

[53] Wiedmer R, Griffis SE. Structural characteristics of complex supply chain networks. J Bus Logist 2021;42(2):264–90.

[54] Brintrup A, Ledwoch A, Barros J. Topological robustness of the global automotive industry. Logist Res 2015;9(1):1.

[55] Brintrup A, Wichmann P, Woodall P, McFarlane D, Nicks E, Krechel W. Predicting hidden links in supply networks. Complexity 2018;2018.

[56] Dhyne E, Magerman G, Rubínová S. The Belgian production network 2002–2012. Tech. rep 288, National Bank of Belgium Working Paper; 2015.

[57] Mungo L, Lafond F, Astudillo-Estévez P, Farmer JD. Reconstructing production networks using machine learning. J Econom Dynam Control 2023;148:104607.

[58] Peydró J, Jiménez G, Huremovic K, Moral-Benito E, Vega-Redondo F. Production and financial networks in interplay: Crisis evidence from supplier-customer and credit registers. Tech. rep 15277, CEPR Press Discussion Paper; 2020.

[59] Silva TC, Amancio DR, Tabak BM. Modeling supply-chain networks with firm-to-firm wire transfers. Expert Syst Appl 2022;190:116162.

[60] Tamura K, Miura W, Takayasu M, Takayasu H, Kitajima S, Goto H. Estimation of flux between interacting nodes on huge inter-firm networks. Int J Modern Phys Conf Series 2012;16:93–104.

[61] Fujiwara Y, Inoue H, Yamaguchi T, Aoyama H, Tanaka T, Kikuchi K. Money flow network among firms' accounts in a regional bank of Japan. EPJ Data Sci 2021;10(1):19.

[62] Ialongo LN, de Valk C, Marchese E, Jansen F, Zmarrou H, Squartini T, et al. Reconstructing firm-level interactions in the dutch input–output network from production constraints. Sci Rep 2022;12(1):1–12.

[63] Saito YU, Watanabe T, Iwamura M. Do larger firms have more interfirm relationships? Physica A 2007;383(1):158–63.

[64] Ohnishi T, Takayasu H, Takayasu M. Hubs and authorities on Japanese inter-firm network: Characterization of nodes in very large directed networks. Progr Theoret Phys Suppl 2009;179:157–66.

[65] Ohnishi T, Takayasu H, Takayasu M. Network motifs in an inter-firm network. J Econ Interact Coord 2010;5(2):171–80.

[66] Fujiwara Y, Aoyama H. Large-scale structure of a nation-wide production network. Eur Phys J B 2010;77(4):565–80.

[67] Alon U. Network motifs: theory and experimental approaches. Nature Rev Genet 2007;8(6):450–61.

[68] Mattsson CE, Takes FW, Heemskerk EM, Diks C, Buiten G, Faber A, et al. Functional structure in production networks. Front Big Data 2021;4:666712.

[69] Kovács IA, Luck K, Spirohn K, Wang Y, Pollis C, Schlabach S, et al. Network-based prediction of protein interactions. Nature Commun 2019;10(1):1240.

[70] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks. J Am Soc Inf Sci Technol 2007;58(7):1019–31.

[71] Hooijmaaijers S, Buiten G. A methodology for estimating the dutch interfirm trade network, including a breakdown by commodity. Tech. rep, Technical report, Statistics Netherlands; 2019.

[72] Antràs P, Chor D. Global value chains. In: Handbook of international economics, vol. 5, Elsevier; 2022, p. 297–376.

[73] Saracco F, Straka MJ, Di Clemente R, Gabrielli A, Caldarelli G, Squartini T. Inferring monopartite projections of bipartite networks: an entropy-based approach. New J Phys 2017;19(5):053022.

[74] Gualdi S, Cimini G, Primicerio K, Di Clemente R, Challet D. Statistically validated network of portfolio overlaps and systemic risk. Sci Rep 2016;6(1):39467.

[75] Vallarano N, Bruno M, Marchese E, Trapani G, Saracco F, Cimini G, et al. Fast and scalable likelihood maximization for exponential random graph models with local constraints. Sci Rep 2021;11(1):15227.

[76] Lind PG, Gonzalez MC, Herrmann HJ. Cycles and clustering in bipartite networks. Phys Rev E 2005;72(5):056127.

[77] Zhang P, Wang J, Li X, Li M, Di Z, Fan Y. Clustering coefficient and community structure of bipartite networks. Physica A 2008;387(27):6869–75.

[78] Estrada E, Rodriguez-Velazquez JA. Subgraph centrality in complex networks. Phys Rev E 2005;71(5):056103.

[79] Mariani MS, Ren Z-M, Bascompte J, Tessone CJ. Nestedness in complex networks: Observation, emergence, and implications. Phys Rep 2019;813:1–90.

[80] Bascompte J, Jordano P, Melián CJ, Olesen JM. The nested assembly of plant–animal mutualistic networks. Proc Natl Acad Sci 2003;100(16):9383–7.

[81] Tacchella A, Cristelli M, Caldarelli G, Gabrielli A, Pietronero L. A new metrics for countries' fitness and products' complexity. Sci Rep 2012;2(1):723.

[82] Almeida-Neto M, Guimaraes P, Guimaraes Jr PR, Loyola RD, Ulrich W. A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. Oikos 2008;117(8):1227–39.

[83] Park J, Newman ME. Statistical mechanics of networks. Phys Rev E 2004;70(6):066117.

[84] Squartini T, Garlaschelli D. Analytical maximum-likelihood method to detect patterns in real networks. New J Phys 2011;13(8):083001.

[85] Cimini G, Squartini T, Saracco F, Garlaschelli D, Gabrielli A, Caldarelli G. The statistical physics of real-world networks. Nat Rev Phys 2019;1:58–71.

[86] Zhou T, Ren J, Medo M, Zhang Y-C. Bipartite network projection and personal recommendation. Phys Rev E 2007;76(4):046115.

[87] Cimini G, Carra A, Didomenicantonio L, Zaccaria A. Meta-validation of bipartite network projections. Commun Phys 2022;5(1):76.

[88] Moré JJ. The levenberg-marquardt algorithm: implementation and theory. In: Numerical analysis: proceedings of the biennial conference held at dundee, June 28–July 1, 1977. Springer; 2006, p. 105–16.

[89] Tumminello M, Micciche S, Lillo F, Piilo J, Mantegna RN. Statistically validated networks in bipartite complex systems. PLoS One 2011;6(3):e17994.

[90] Thissen D, Steinberg L, Kuang D. Quick and easy implementation of the benjamini-hochberg procedure for controlling the false positive rate in multiple comparisons. J Educ Behav Stat 2002;27(1):77–83.

[91] Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Statist Mech Theory Exper 2008;2008(10):P10008.

[92] Lancichinetti A, Fortunato S. Community detection algorithms: a comparative analysis. Phys Rev E 2009;80(5):056117.

[93] Fortunato S. Community detection in graphs. Phys Rep 2010;486(3–5):75–174.

[94] Barber MJ. Modularity and community detection in bipartite networks. Phys Rev E 2007;76(6):066102.

[95] Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. SIAM Rev 2009;51(4):661–703.

[96] Brintrup A, Kito T, Alzayed A, Meyer M. Nested patterns in large-scale automotive supply networks. Captur Value Int Manuf Supply Networks Inst Manuf 2012.

[97] Brintrup A, Barros J, Tiwari A. The nested structure of emergent supply networks. IEEE Syst J 2018;12(2):1803–12.

[98] Chauhan VK, Perera S, Brintrup A. The relationship between nested patterns and the ripple effect in complex supply networks. Int J Prod Res 2021;59(1):325–41.

[99] Laudati D, Mariani MS, Pietronero L, Zaccaria A. The different structure of economic ecosystems at the scales of companies and countries. J Phys Complex 2023;4(2):025011.

[100] Albora G, Zaccaria A. Machine learning to assess relatedness: the advantage of using firm-level data. Complexity 2022;2022.

[101] Taglioni D, Winkler D. Making global value chains work for development. World Bank Publications; 2016.

[102] Goldthau A, Hughes L. Protect global supply chains for low-carbon technologies. Nature 2020;585(7823):28–30.

[103] Fajgelbaum P, Goldberg PK, Kennedy PJ, Khandelwal A, Taglioni D. The US-China trade war and global reallocations. Tech. rep, National Bureau of Economic Research; 2021.

[104] Timmer MP, Dietzenbacher E, Los B, Stehrer R, De Vries GJ. An illustrated user guide to the world input–output database: the case of global automotive production. Rev Int Econ 2015;23(3):575–605.

[105] Veloso F, Kumar R. The automotive supply chain: Global trends and Asian perspectives. Int J Bus Soc 2003;4(2):27–70.

[106] Cimini G, Mastrandrea R, Squartini T. Reconstructing networks. Elements in structure and dynamics of complex networks, Cambridge University Press; 2021.

[107] Wichmann P, Brintrup A, Baker S, Woodall P, McFarlane D. Extracting supply chain maps from news articles using deep neural networks. Int J Prod Res 2020;58(17):5320–36.

[108] Kosasih EE, Brintrup A. A machine learning approach for predicting hidden links in supply chain with graph neural networks. Int J Prod Res 2021;1–14.

[109] Reisch T, Heiler G, Diem C, Klimek P, Thurner S. Monitoring supply networks from mobile phone data for estimating the systemic risk of an economy. Sci Rep 2022;12(1):1–10.

[110] Shao BB, Shi ZM, Choi TY, Chae S. A data-analytics approach to identifying hidden critical suppliers in supply networks: Development of nexus supplier index. Decis Support Syst 2018;114:37–48.