https://doi.org/10.14669/AM/214254

# NETWORK ALGORITHM TO MODEL AUTOMOTIVE SUPPLY CHAIN STRUCTURE

**JOSÉ BARROS[1], CHRISTOPHER TURNER[2]**

## Abstract

A network algorithm that models the structure of automotive supply chains, compiled from a proprietary database, is presented. An initial structural analysis was conducted using key performance indicators, including average path length, clustering coefficient, and degree distribution, to assess network configurations. The networks were then partitioned into subnetworks, with an emphasis on reflecting the operational dynamics of supply chain activities. Regression analysis was applied to each subnetwork, using the number of vertices as the independent variable, to develop an algorithm for generating synthetic networks. These synthetic constructs serve as benchmarks for the automotive sector and have shown a strong average correlation (0.94) with the structure of actual supply networks. This methodological contribution provides tools for analysing and optimising supply chain structures that underpin automotive engineering and manufacturing, ensuring robustness and efficiency in vehicle production systems. The prevalence of tree-like structures within supply networks challenge conventional beliefs regarding the complexity of automotive supply chains and prompts further investigation into the determinants of their resilience.

**Keywords:** Graph Theory; Networks; Supply Chain Management; Automotive Industry; Structural Analysis

## 1. Introduction

Graph theory provides a foundational framework for analysing distinct areas of knowledge that can be represented by networks, and is employed here to create a model that systematically reproduces automotive supply chain structure. In operations management, a key area of investigation is the analysis and optimisation of supply chains, which are networks

[1] Iscte-Sintra, Instituto Universitário de Lisboa, Portugal, email: jmmcb@iscte-iul.pt, ORCID: 0000-0002-9063-8360

[2] University of Surrey, University of Surrey, United Kingdom, email: christopher.turner@surrey.ac.uk, ORCID: 0000-0003-2812-3312

https://doi.org/10.14669/AM/214254

of interconnected companies engaged in buy-sell relationships. Within these networks, companies are represented as vertices (nodes, entities, …) and the commercial relationships as edges (links, arcs, …). One of the primary objectives of supply chain management (SCM) is to optimise their network structure and functionality together with internal and external variabilities. In the automotive sector, supply networks are often defined by a manufacturer-centric web of companies that contribute directly or indirectly to the production process [3, 13]. Given that the efficiency and resilience of supply networks directly influence the design, manufacturing, and operational continuity of automotive systems, this study situates network algorithms firmly within the domain of automotive engineering [2].

In this study, we compare the network structure of empirical data from automotive supply chains with established network models, such as Regular, Random, Small-World, and Scale-Free networks. This comparison aims to categorize real networks within established theoretical frameworks. Regular graphs have long average path length (APL), high Clustering Coefficient (CC), and constant degree distribution. Random networks have randomly distributed edges which implies a small CC. Small-world networks are built from switching random edges in a regular network in order to reduce the APL but keep the high CC. The scale-free networks have power-law distributions, which means they have small-world property with a low CC.

Our investigations into automotive supply network data, when measured against these models, failed to identify specific characteristics. Parameters such as APL, CC, and degree distribution did not match with any of the models. Hence, our research question is focused on the possibility to develop a network model that accurately represents the structural configurations of automotive supply chains, known to have specific characteristics in terms of local and global properties [5]. A mathematical model proposed in this work, designs the structure of synthetic automotive networks with correlation coefficient exceeding 90% when compared to real ones. This model serves as a potent tool in the decision-making process, bridging the gap between theoretical network models and the intricate realities of automotive supply chains [7].

## 2. Literature Review

Network science has become increasingly prominent, with research using network measures such as Centrality, which is a group of measures related to the distribution of edges between the vertices, to optimize supply chain structures. A high degree centrality company is a company with many suppliers and/or many clients, like a large assembler or a generalist distributor. A high closeness centrality company, on the other hand, is a company with important neighbours, and a high betweenness centrality company bridges different production lines. The automotive manufacturers are assemblers, which means that they have high in-degree centrality, and a correlation has been noted between high in-degree and performance of automotive companies [4].

https://doi.org/10.14669/AM/214254

Average path length is the average of the shortest paths between any two vertices in a network. The global CC is proportional to the ratio between the number of triangles and the number of paths of length equal to three, that are present in the network. Simulation on networks with small-world and scale-free characteristics has shown that lower path lengths and higher clustering are indicators of lower diffusion of risk. Percolation studies analyse the structural robustness of networks by removing vertices (site percolation) or edges (bond percolation), or both, usually until the network becomes disconnected. The higher the number of items needed to be removed, the higher the network robustness. In tree networks, the removal of any edge, or the removal of any vertex with degree higher than one, disconnects the network, showing that tree networks are not structurally robust. Degree distribution is a probability function of the degrees over the network, and its shape is associated with the structure: a constant distribution means a regular network, a Poisson distribution represents a large random network, and Power-Law distribution implies scale-free networks and are common in many fields. Each characteristic is associated with some structural and robustness attributes, and are usually analysed in studies with empirical data networks [20].

Null models are pattern-generating tools that are unbiased, allowing to formulate and test hypothesis against real data. In our study, we have applied these models, including random networks, scale-free networks, and small-world networks, to the context of automotive data, failing to observe a perfect correspondence. The literature is rich in different types of network modelling, such as growth models, providing evidence of the mechanisms of network growth, network dynamics and topology [1, 14]. Our research primarily focuses on the development of an algorithm that functions as a network model, producing what we denote as synthetic or artificial networks. The key advantage of this model is to create synthetic networks whose properties and measures are identical to the real networks, while having only one independent variable: the number of nodes of the network.

Building upon the foundational theories reviewed, our study introduces a novel methodological perspective that both aligns with and expands upon the recent advancements in network science within the automotive sector. Our findings contribute to the existing body of research on automotive supply chains by improving the understanding on their structural characteristics through network algorithms. In [11], a review is provided on reconstructing methods on supply networks from partial and indirect data observations, and our algorithm adds an approach for the consideration of complete empirical data. Additionally, the importance of reconstructing firm-level interactions is emphasized, which complements our method by significantly refining the knowledge on dynamic supply chain interactions and resilience in complex industrial environments [6, 11].

https://doi.org/10.14669/AM/214254

# 3. Methodology

To address our research question, we have outlined a series of objectives. The first was to conduct a thorough empirical data analysis that identifies and classifies the network based on its inherent properties. It will entail executing a suite of classical network measures, which are essential for a comprehensive understanding of the structural characteristics that are central to both network science and the evaluation of supply chains in automotive engineering.

Secondly, upon verifying that automotive networks do not match any traditional theoretical model, the objective was to partition and study the intricate parts of supply chains separately. This reflects the operational reality in the automotive industry, where manufacturers assemble vehicles from modules delivered by Tier-1 suppliers, who in turn rely on Tier-2 and Tier-3, and further upstream. Within out framework, companies supplying directly to the manufacturer, hence the Tier-1 suppliers, constitute a subnetwork which is here defined as Layer 1 ($L_1$) of our network analysis. In addition, Tier-2 companies form what is here denoted as Layer 2 ($L_2$), and the same logic is applied to all tiers upstream. The partitioning method applied in this work remains directly interpretable within the tiered architecture that managers and engineers already use in practice. Hence, we can divide companies from the whole network into sub-networks, and in a first step estimate the number of companies included in each part. The number of edges or supply relationships between companies is also studied separately for each subnetwork, as this approach increases correlation accuracy. The goal was to have a simple model with one input variable, that could be easily obtained and used directly to perform a structural analysis of these layers, i.e., estimate the number of companies and their interconnections across the whole supply chain. As shown in this work, the number of vertices correlates strongly with the number of direct suppliers, a parameter usually known to companies.

Finally, the last objective was to ensure our processes and ideas could be verified. We developed a procedure which used part of the empirical data to create the model, and some other part of the data to be used as validation. In network terms, we have created 15 automotive supply networks, of which 14 were used to generate the model, and one was left aside for validation. A high correlation between the real network and synthetic networks is an indication of the validity of this model.

To achieve each of these objectives, the following steps were taken: ($i$) data collection and processing, including data cleaning and organisation into supply network format; ($ii$) selection of platform of analysis, chosen to be the Wolfram Mathematica software, with which we were able to read the files, create the graphs, apply the measures, and perform the partition process and consequent model creation; ($iii$) development of a systematic testing system based on random sampling of synthetic networks to be compared against the properties of real networks.

https://doi.org/10.14669/AM/214254

# 4. Data

Empirical data was collected from an independent database named Marklines and organised in the form of edge-lists (supplier-client), enabling the analysis of multiple automotive supply chains. Marklines has several data quality checking mechanisms that include the collection of the data through press releases, interviews, surveys, consultancy, and cooperation with other companies. Statistical indicators are displayed on the website regarding supplier information, Original Equipment Manufacturer (OEM), sales and production data, market and tech reports, prediction algorithms, and so on. The limitations of this database are at the extremities of the supply chain, i.e., the lack of raw materials' companies, and the clients of the automotive manufacturers. For this reason, the object of our analysis is the network of the product flow and assembling of automotive parts in order to build the final vehicles.

Data is organised as each company having a distinct profile webpage, so we have designed an algorithm that would capture the list of companies by collecting all company names and corresponding profile pages. Afterwards, another algorithm was developed to run through every company profile and collect the list of clients of each, along with other pieces of information. The raw data was stored in Excel files in tabular form with the following columns: company name, list of clients, list of products, headquarters' address, and country. Each company decides whether to share its information, and to which extent, hence some companies might be absent (or limited) by their own choice. In addition, data is entered manually in the database, so it is prone to have data quality issues, such as finding company names with "Ltd" and "Ltd,", resulting in two different vertices on the corresponding graph. We have performed extensive data cleaning to improve the quality of our results.

In the construction of the networks, each company was represented as a vertex and the edges were stored in comma-separated-value (CSV) files, where the first element is the supplier vertex and the second one is the client vertex. Directed edges are established from these files when reading them into our scripts. A weakly connected component refers to a group of vertices that are linked to one another, irrespectively of the direction of the edge, allowing for an analysis of how companies within a supply chain are interrelated. This connected component denoted in literature by Largest Connected Component (LCC) was found to have: 19262 companies, 89205 edges, 779 products and product categories or services, and 74 countries.

The LCC contains different production sites of the same company, such as *company A (Japan), company A (France)*, and so on, which makes the depth of the structural analysis at the production sites level. Nevertheless, all vertices and all edges were considered functionally identical, which means that the network is not weighted, because there was no information on the amount of product flow and flow times in each edge, or storage/assembling capacity in each vertex.

https://doi.org/10.14669/AM/214254

# 5. Results

Network partitioning is a process to divide a network into sub-networks, according to some logic. From the LCC, we identify one car brand, for example “Jaguar Land Rover” (JLR), and search for the companies that supply to JLR. All companies that are directly linked to JLR form what is called Tier-1 for JLR’s supply chain (which is a sub-network of the LCC). The next step is to search for the Tier-2 suppliers, i.e., the companies that are supplying to the Tier-1 suppliers of JLR. The process continues searching for consecutive tiers until there are no more suppliers. When this process ends, we get the sub-network from the LCC created for JLR, which we call the JLR Automotive Network (AN), and we save the information in a separate CSV file. Similarly, we have searched and stored the AN of the top 25 car brands by revenue according to Marklines database, and from those we have selected the top 15 networks by number of vertices, to guarantee statistical significance and reduce noise, as explained below.

One measure used to partition the networks was the (directed) distance. Every vertex other than the manufacturer (denoted by $M$), is origin of some path targeted to $M$. Therefore, we can identify the tiers, which here will be denoted Layers, as layer $L_k$ at a distance $k$ from $M$. The maximum distance to $M$ is the total number of layers, herein named *height* of the network (in graph theory is often called *diameter*). The average height found was 3 for larger networks, which means that the maximum layer is $L_3$. For smaller networks, the height found was 4, and when this layer $(L_4)$ exists it has at most 3 vertices. So, all networks analysed in the study had at most four layers, which means maximum of 4 steps between any company and the car manufacturer $M$.

In every network it is possible to observe a region of higher density in layers $L_1$ and $L_2$, which we define by the network core. Network Density, in graph theory terms, is defined as the number of edges divided by the number of vertices of a network, but the concept is borrowed to sub-networks and network regions. From the layer $L_2$ upwards $(L_3$ and $L_4)$ we find divided branches, such as linear chains with tree-like structures, i.e., groups of vertices that had no suppliers, and only had one client. In graph theory, trees are simple graphs, i.e., no loops or multiedges, with $n$ vertices and a maximum of $l = n+1$ edges. These tree-like regions are here denoted by the network periphery, due to their structure and radial positioning in the network. Since the periphery constitutes a large portion of the whole network (around 90%), a first conclusion can be derived: automotive supply chains are mostly tree-like, which according to percolation studies, are not robust structures, as explained before.

While our analysis highlights the prevalence of tree-like structures, it is crucial to consider other structural factors and strategic measures that may influence the overall robustness of these supply chains. Tree-like structures are known for their vulnerability in network theory, particularly due to their lack of redundancy: the failure of a single node or link can lead to significant disruptions in connectivity. However, automotive supply chains also exhibit

https://doi.org/10.14669/AM/214254

features that might counteract this inherent vulnerability. Features such as substitution and dual-sourcing strategies could potentially introduce redundancy and resilience into the system, mitigating the fragility typically associated with tree-like structures [16, 17]. In addition, in the automotive industry the reduction of redundancies and costs might favour streamline production flows, which are translated into tree-like structures, suggesting that these structures are found in supply chains by choice of the automotive management and engineering.

It is imperative to consider that while tree-like structures contribute to vulnerability, they are not the sole determinant of network robustness. The overall architecture of a supply chain network, including its diverse subgraph structures and the strategic integration of redundancy measures, plays a crucial role in defining its robustness. This perspective aligns with recent studies which suggest that complex networks can possess high resilience despite their structural vulnerabilities [15, 18]. Nevertheless, structure is not the sole parameter to be taken into account for the overall stability of networks. Even if the structure is not robust, other features suggest resilience of networks as seen through classical cases after floods, earthquakes, and other disasters. In [16], a review of nine parameters to analyse the resilience of supply networks is provided, with an overview over the past decades of the literature.

In addition, the database and data collection process could be inclined to search and store the data as a tree-like structure. However, Marklines has different parallel and redundant mechanisms of data collection, cleaning, and storage, to avoid implying these limitations into their dataset, making them a very reliable database. Finally, our algorithm was programmed to follow a list of companies and not any kind of sequential search which could output a tree-like dataset.

Since the network periphery is easily modelled by a tree-like model, the focus of our analysis was on the network core. Instead of completely exclude the network periphery, we have developed a reduction algorithm to simplify it to a single chain of vertices, in order to keep the information of the height (network diameter), and for measures such as average path length outputting the same results. The logic of the algorithm was to search the network for star-like motifs, and to reduce them to a dyad relationship. This process is reversible by keeping the motif size and location stored in separate variables. This also allowed us to focus on the complex part of the supply networks, which did not follow any particular standard structural features. It is important to underscore that the networks studied here are all in the reduced form, which largely changed the size of the network size but not its structural features.

Analysing Figure 1, we find that the group of vertices at distance 1 from $M$ (black vertex) is defined as layer $L_1$, and is shown as dark grey vertices. At distance 2 from $M$ are the $L_2$ vertices in a lighter grey colour, $L_3$ vertices are represented by a black, white-filled circles, while the only vertex in $L_4$ is in a blue, white-filled circle. Three sections a), b), and c) are

https://doi.org/10.14669/AM/214254

displayed: in section a) there is an edge between two $L_1$ vertices, which we defined as a horizontal edge; in b) there are three $L_3$ vertices that do not have suppliers and are all supplying the same $L_2$ company, forming a star-like sub-network; the vertices in $L_3$ in b) will be reduced to one vertex by the reduction algorithm we have designed; in c) a blue edge from $L_2$ to $L_3$, which we define an upward edge is displayed. The sub-network formed by $L_1$ vertices show two vertices sharing a horizontal edge, and four who do not. This aids in defining the horizontally connected vertices as forming a horizontal layer called $H_1$. In automotive networks, $H_1$ and $H_2$ are common, whereas they are absent for higher layers.

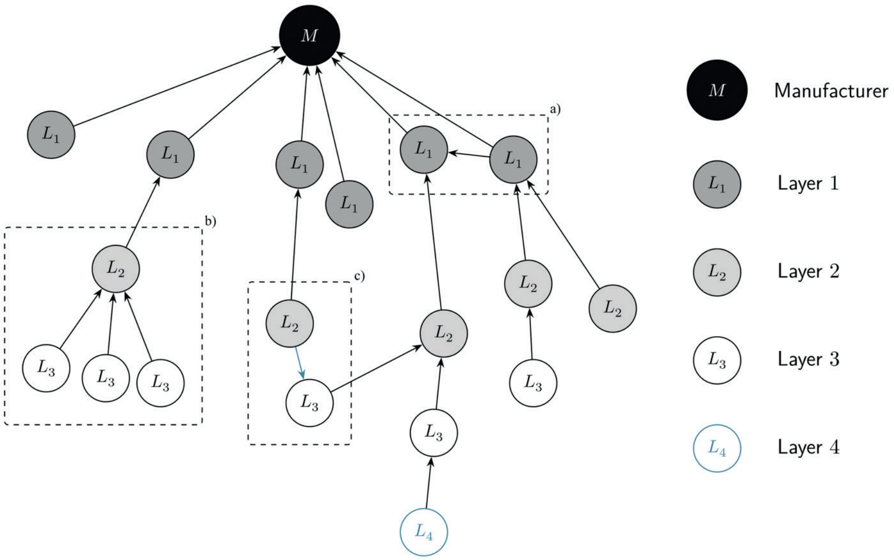


**Fig. 1. Graphical Representation of a Supply Chain: a) Horizontal edge, b) Star sub-network, c) Upward edge**

Studying the flow from $L_2$ to $L_1$, we call this sub-network an inter-layer and denote it by $L_2 \rightarrow L_1$, or $L_{21}$. The same logic will be applied to all regions. The configuration of $L_2 \rightarrow L_1$ depends on the out-degree of each vertex in $L_2$, and the in-degree of each vertex in $L_1$. However, not all vertices in $L_1$ belong to the sub-network $L_{21}$, since some of the vertices do not have suppliers, as we can see in Figure 1. Therefore, $L_1$ is divided into the following sub-layers:

- $L_1^1$ – the group of vertices in $L_1$ that have suppliers from $L_2$ – do not confuse with $H_1 = L_{11}$,
- $L_1^0$ – the group of vertices in $L_1$ which do not have suppliers from $L_2$.

With this definition, $H_1$ can also be divided into $H_1^1$ and $H_1^0$, dividing each of these into source and target vertices of the horizontal edges. It was found that the source vertices are almost

https://doi.org/10.14669/AM/214254

all from $L_1^0$ and the target vertices are almost all in $L_1^1$. These sub-layers (from the intra-layar $H_1$) are defined as:

- $H_1^1$ – the group of vertices in $H_1$ which are targets of the horizontal edges,
- $H_1^0$ – the group of vertices in $H_1$ which are sources of the horizontal edges.

The same ratio and logic were found and implemented for $H_2$. In summary, from the partition process, we described and identified a group of sections and properties to be modelled:

- Layer – group of vertices at the same distance to the manufacturer,
- Height – equivalent to the number of layers,
- Layer size – number of vertices in each layer,
- Inter-layer – sub-network formed by the vertices and edges of two different layers,
- Intra-layer – sub-network formed by the vertices and edges of the same layer,
- Upwards edges – edges connecting two vertices in opposite direction of the main flow.

# 6. Model Fitting

In this section we will describe the interpolating model developed where each of the network parts are correlated to the total network size, for example, how the number of vertices in $L_1$ correlates with the total number of vertices in the network. An initial test showed statistical noise for networks with less than 200 vertices, whereas for larger networks the correlations were all either linear or with small quadratic corrections. For this reason, smaller networks have been excluded from the analysis. The linear correlations are all higher than 0.85, and every time we added a quadratic correction, it was to obtain values above 0.9.

## 6.1. Number of vertices per Subgraph

The height of the automotive networks have a threshold around network size of 350 vertices, dividing them into networks with height equal to 4 (size < 350 vertices), and networks with height equal to 3 (size > 350 vertices). There is only one network with more than 350 vertices, and 4 layers, but it contains only one vertex. The analysis of $L_4$ for all networks show that there are between 1 and 4 vertices in each, without any correlation to the network size, so we model this layer with a simple random integer generator with uniform distribution.

The number of vertices for $L_1$ and $L_2$ (denoted $n_{\mathrm{L}_1}$ and $n_{\mathrm{L}_2}$) are modelled with the total number of vertices $n$ (see Table 1). The fitting demonstrates a linear relationship between $n$ and both $n_{\mathrm{L}_1}$ and $n_{\mathrm{L}_2}$, and hence $n_{\mathrm{L}_3}$ is obtained as described in Equation 1:

$$n_{L_3} = n - n_{L_1} - n_{L_2} - n_{L_4} \tag{1}$$

https://doi.org/10.14669/AM/214254

As mentioned above, the $L_1$ and $L_2$ sub-networks are divided into $L_1^1$, $L_1^0$, $L_2^1$, and $L_2^0$. For each sub-network, the number of vertices is the sum of the vertices in each sub-part. The estimated sizes for $L_1^1$ and $L_2^1$ are also shown in Table 1.

**Tab. 1. Initial modelling functions and correlation coefficient for the number of vertices of the network sections**

| Parts | Estimated Function | Adjusted $R^2$ |
|---|---|---|
| $n_{L_1}$ | $n_{L_1} = 0.421n - 31.787$ | 0.943 |
| $n_{L_2}$ | $n_{L_2} = 0.638n - 20.487$ | 0.979 |
| $n_{L_1^1}$ | $n_{L_1^1} = 0.182n - 20.934$ | 0.980 |
| $n_{L_2^1}$ | $n_{L_2^1} = -0.011n + 23.249$ | 0.927 |
| $n_{L_3}$ | $n_{L_3} = n - n_{L_1} - n_{L_2} - n_{L_4}$ | |
| $n_{L_4}$ | $n_{L_4}$ = randomInteger [1,4] | |

## 6.2. Number of Edges per Subgraph

The number of edges $l$, as well as other parts, correlates with number of vertices $n$ (see Figure 2), and their estimated functions are displayed in Table 2. However, two additional rules are needed to distribute the $l$ edges in the network:

- One edge for every $L_1$ vertices towards the manufacturer – assembly rule,
- At least one edge for every vertex in the network – connectivity rule.

The edge distribution process for the intra- and inter-layers is more complex. For example, in $L_{21}$ there is the additional property of dividing the layers into $L_1^1$ and $L_1^0$ in $L_1$, and $L_2^0$ and $L_2^1$ in $L_2$. In addition, the horizontal layers $H_1$ and $H_2$ are also divided in the same manner. The total number of horizontal edges in the network ($l_H$), as well as the number of edges in ($l_{H_1}$), correlate with $n$. Hence, the number of edges in $H_2$ is found as $l_{H_2} = l_{\mathrm{H}} - l_{\mathrm{H}_1}$.

https://doi.org/10.14669/AM/214254

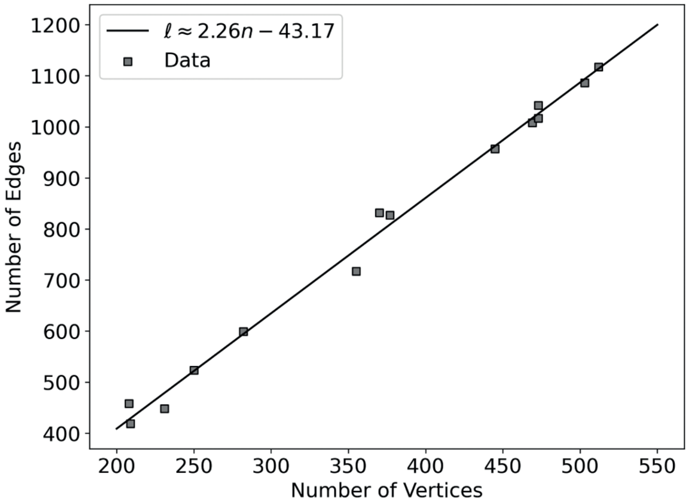


**Fig. 2. Data points and corresponding modelling plot for the number of vertices versus the number of edges, with adjust $R^2$ = 0.992**

Regarding the distribution of edges between layers, the following maximal quantities are defined to limit the degree $k$ in the edge attribution process:

- maximum out-degree in $L_{21}$: $\max k_{L_{21}}^{\text{out}}$,
- maximum in-degree in $L_{21}$ $\max k_{L_{21}}^{\text{in}}$,
- maximum out-degree in $H_1$: $\max k_{H_1}^{\text{out}}$,
- maximum in-degree in $H_1$: $\max k_{H_1}^{\text{in}}$,
- maximum in-degree in $H_2$: $\max k_{H_2}^{\text{in}}$,
- maximum in-degree in $L_{32}$ $\max k_{L_{32}}^{\text{in}}$,
- maximum out-degree for $H_2$ is defined equal to 3 since no correlation was found: $\max k_{H_2}^{\text{out}} = 3$,
- maximum out-degree for $L_{32}$ (recall this is an upward link, from layer $L_2$ to $L_3$) is defined equal to 2 or 3 since no correlation was found: $max\ k_{L_{32}}^{\text{out}}$ = randomInteger [2, 3],
- All $L_4$ vertices have exactly one out-edge each.

The upwards edges connect vertices with the opposite direction of the main flow of the network, for example, a company in $L_1$ that supplies a company in $L_2$. The number of upward edges is calculated with an estimated function (Table 2), but the way we have distributed them in the network follows an algorithm to increase the global clustering coefficient.

Our first analysis showed that the global CC of the synthetic networks was smaller than the one of the real networks. Hence, we distributed the upward edges in order to increase the number of triangles, by selecting a vertex in $L_2$ to be a client of a vertex from $L_1$, specifically,

https://doi.org/10.14669/AM/214254

one that is not a client already from the vertex in $L_2$, in order to avoid cycles of two vertices, since we only need loops only with three vertices (triangles). The following algorithm was implemented:

- randomly select a vertex from $L_1$ and identify the list of its neighbours,
- match the list of neighbours with the vertices in $L_2$, to know their suppliers,
- select a vertex in $L_2$, excluding the ones found in the previous step.

**Tab. 2. Full list of modelling functions and correlation coefficient for the number of edges of the network sections**

| Parts | Estimated Function | Adjusted $R^2$ |
|---|---|---|
| $l$ | $l = 2.259\ n - 43.166$ | 0.992 |
| $l_{L_{21}}$ | $l_{L_{21}} = 1.341n - 107.523$ | 0.992 |
| $l_H$ | $l_H = 0.54n + 24.14$ | 0.977 |
| $l_{up}$ | $l_{up} = 0.03n + 1.14$ | 0.901 |
| $1_{L_{32}}$ | $1_{L_{32}} = l - l_{L_{21}} - l_H - l_{up} - l_{L_{43}}$ | 0.910 |
| $1_{L_{43}}$ | $1_{L_{43}} = n_{L_4}$ | |
| $\max k_{L_{21}}^{\text{out}}$ | $\max k_{L_{21}}^{\text{out}} = 0.012n + 1.709$ | 0.990 |
| $\max k_{L_{21}}^{\text{in}}$ | $\max k_{L_{21}}^{\text{in}} = 0.018n + 31.622$ | 0.956 |
| $\max k_{H_1}^{\text{out}}$ | $\max k_{H_1}^{\text{out}} = 0.01n + 2.76$ | 0.938 |
| $\max k_{H_1}^{\text{in}}$ | $\max k_{H_1}^{\text{in}} = 0.04n + 7.51$ | |
| $\max k_{H_2}^{\text{out}}$ | $\max k_{H_2}^{\text{out}} = 3$ | |
| $\max k_{2}^{\text{in}}$ | $\max k_{2}^{\text{in}} = 0.03n + 1.33$ | 0.978 |
| $\max k_{L_{32}}^{\text{out}}$ | $\max k_{L_{32}}^{\text{out}} = \text{randomNumber}\ [2,3]$ | |
| $\max k_{L_{32}}^{\text{in}}$ | $\max k_{L_{32}}^{\text{in}} = \text{-}0.009n + 7.472$ | 0.911 |

Finally, the following parts needed specific rules to be estimated, either due to lack of data points, or to increase the match with the real networks:

- the proportion of origin vertices that are both in $H_1$ and in $L_1^1$ ($p_{H_1 \cap L_1^1} = 95.9 \pm 3.8\%$),
- the proportion of origin vertices that are both in $H_2$ and in $L_2^0$ ($p_{H_2 \cap \mathrm{L}_2^0} = 95.2 \pm 3.5\%$).

## 6.3. Inter-layer Out-Degree Distribution.

The out-degrees for the vertices in the inter-layers are well approximated by a random distribution of excess edges. To guarantee the connectivity rule, we need one out-edge for every vertex in the source layer, and one in-edge for every vertex in the target layer. Consid-

https://doi.org/10.14669/AM/214254

ering the $L_{21}$ sub-network, a total of $l_{L_{21}}$ edges are distributed, and $l^{\mathrm{r}}_{L_{21}} = l_{L_{21}} - \max\,[n_{L_2},\, n_{L_1^1}]$ are remaining, i.e., either $L_2$ or $L_1^1$ is filled. The maximum function $\max\,[n_{L_2},\, n_{L_1^1}]$ tell us the number of edges included in this process, by selecting the largest sub-layer, guaranteeing that the other sub-layer is completely connected to the first one. The same logic is used to build the distributions for $L_{32}$, $H_1$ and $H_2$. Remember, $H_1$ and $H_2$ are basically $H_1^0 \rightarrow H_1^1$ and $H_2^0 \rightarrow H_2^1$, respectively.

Mathematically, we group all networks into one, so we will have the sum of all edges, denoted here by $\tilde{\ell}$, to distribute for the sum of all vertices, denoted here by $\tilde{n}$. Hence, the probability $p^{\mathrm{out}}$ of having a vertex with out-degree $k^{\mathrm{out}}$ when $(\tilde{\ell} - \tilde{n})$ excess edges are randomly distributed among $\tilde{n}$ vertices, is described by the binomial distribution in Equation 2.

$$p^{\mathrm{out}}(k^{\mathrm{out}}) = \binom{\tilde{\ell} - \tilde{n}}{k^{\mathrm{out}} - 1} \left(\frac{1}{\tilde{n}}\right)^{k^{\mathrm{out}}-1} \left(1 - \frac{1}{\tilde{n}}\right)^{\tilde{\ell}-\tilde{n}-(k^{\mathrm{out}}-1)} \tag{2}$$

The solution of Equation 2 is a vector of probabilities for each degree up to the maximum possible degree, which is $\max k^{\mathrm{out}}$:

$$\overrightarrow{p^{\mathrm{out}}} = \ \{p^{\mathrm{out}}\,(1), p^{\mathrm{out}}\,(2), \dots, p^{\mathrm{out}}\,(\max k^{\mathrm{out}})\} \tag{3}$$

Since we truncate the distribution at $\max k^{\mathrm{out}}$, the sum of the probabilities in Equation 3 is not equal to 1. Hence, to normalize this probability vector, each of the elements is divided by the sum of all elements, obtaining a normalized vector.

$$\overrightarrow{p'} = \left\{p^{\mathrm{out}\,\prime}(1), p^{\mathrm{out}\,\prime}(2), \dots, p^{\mathrm{out}\,\prime}(\max k^{\mathrm{out}})\right\} \ \therefore \ \sum_{j=1}^{\max k^{\mathrm{out}}} p^{\mathrm{out}\,\prime}(j) = 1 \tag{4}$$

The components of the vector $\vec{p}$ are the frequencies of each degree. These are real numbers that need to be rounded to the nearest integer, representing the estimated number of vertices with degree $k$. Using the example of $L_{21}$, we are distributing out-degrees for the $L_2$ vertices, and the number of vertices with each distinct degree $k$ will be:

$$n_{k=1}, \mathrm{n}_{k=2}, \dots, \mathrm{n}_{k=\max k^{\mathrm{out}}} \tag{5}$$

Due to the rounding process, the total number of vertices $(n = n_{k=1} + n_{k=2} + \cdots)$ of this vector may differ from $n_{L_2}$. The plots in Figure 3 show the comparison of the out-degree distributions of the data (grey squares) and the model (black lines) obtained from Equation 2. Table 3 shows the parameters used for the calculation.

https://doi.org/10.14669/AM/214254

**Tab. 3. Number of vertices and edges used in the modelling of each sub-network**

| Sub-network | $\tilde{n}$ | $\tilde{\ell}$ |
|---|---|---|
| $L_{21}$ | 2996 | 5410 |
| $L_{32}$ | 403 | 532 |
| $H_1$ | 1173 | 2333 |
| $H_2$ | 702 | 820 |

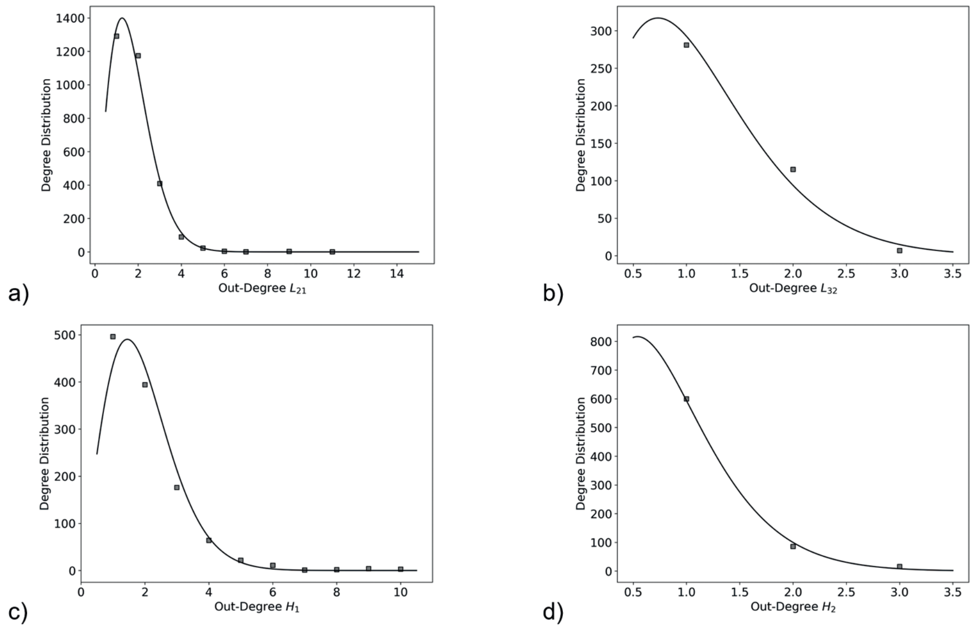


**Fig. 3. Out-degree distributions of the sub-networks (squares) against their model fit (line)**

## 6.4. Inter-layer In-Degree Distribution

For the in-degree distribution, the power-law distribution, displayed as black lines in Figure 4, follow with a certain degree of approximation the grey data points, without a clear overlap. The goodness of the fit was tested with the adjusted $R^2$ test, always obtaining correlation values above 0.9. The probability $p^{\text{in}}$ of having a vertex with in-degree $k^{\text{in}}$ is calculated, according to the power-law, by the expression:

$$p^{\text{in}}(k^{\text{in}}) = \alpha(k^{\text{in}})^{\beta} \tag{6}$$

https://doi.org/10.14669/AM/214254

where $\alpha$ and $\beta < 0$ are global fitting parameters. Due to the statistical noise resulted from the small size of the in-degree sub-layers, the data from all different networks were grouped together and the fitting was done to the resulting "big network". This means that all degrees $k^{in}$ and their corresponding frequencies, summed through all the networks, are used to find the best power-law distribution, adjusted with the method of the least-squares.

The in-degree distributions of the inter-layers $L_{21}$ and $L_{32}$ and the intra-layers $H_1$ and $H_2$ are obtained from the fitted distributions $p^{\text{in}}\ (k^{\text{in}})$, with the same considerations for the total edges and degrees, as explained above for the $p^{\text{out}}\ (k^{\text{out}})$. In Figure 4, the plots for the in-degree distributions of the data (grey squares) and the fitted model (black lines) are shown. Table 4 shows the $\alpha$ and $\beta$ parameters used for each sub-network.

**Tab. 4. Parameters $\alpha$ and $\beta$ used in the power-law modelling of each sub-network**

| Sub-network | $\alpha$ | $\beta$ |
|---|---|---|
| $L_{21}$ | 109.825 | -0.758 |
| $L_{32}$ | 129.713 | -1.246 |
| $H_1$ | 134.995 | -0.984 |
| $H_2$ | 90.682 | -1.231 |

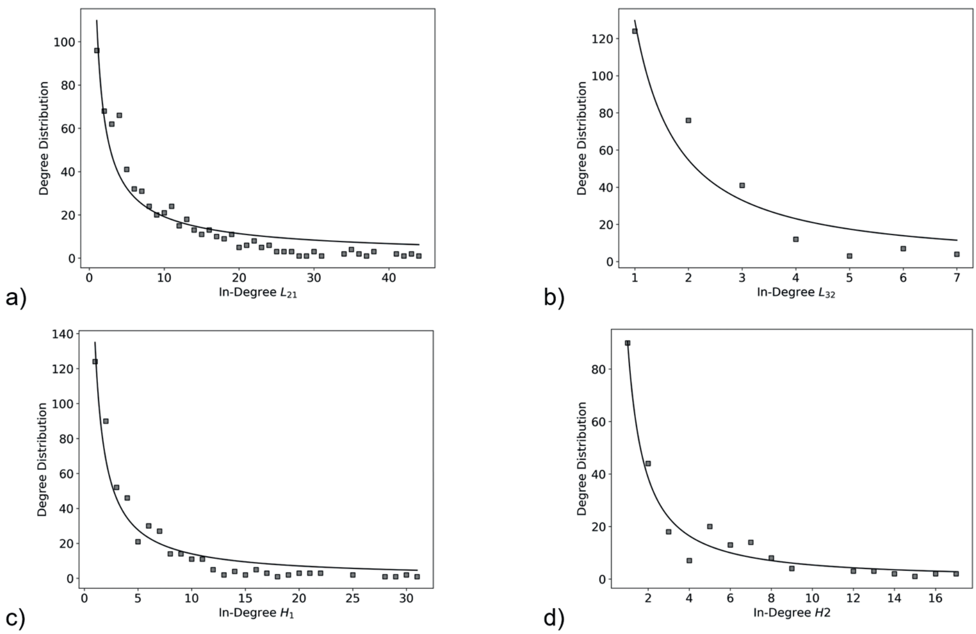


**Fig. 4. In-degree distributions of the sub-networks (squares) against their model fit (line)**

https://doi.org/10.14669/AM/214254

## 6.5. Network Algorithm

A network can be represented by an adjacency matrix $A$, which is square matrix of size $n$, where $n$ is the total number of vertices. Matrix elements $a_{ij}$ are defined as:

$$a_{ij} = \begin{cases} 1, & \text{if } i \text{ supplies } j \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

Vertex $j$ is placed on row $j$ and column $j$, which store the information on the clients and suppliers (respectively) of vertex $j$.

The algorithm developed in this work is a single input algorithm, whose input variable is the number of vertices of the network. Initially, it creates a matrix of size $n$ filled with zeros (no edges). There is only one vertex in the whole network that does not have suppliers (out-degree equal to zero), which is the manufacturer $M$, and by default it is placed at position $i = 1$ of the matrix (row 1, column 1). Each $L_1$ vertex has exactly one out-edge to $M$ and for that reason they are placed consecutively after index 1, i.e., the $n_{L_1}$ vertices are linked to $M$ via $a_{i1} = 1$, for $i = 2,...,n_{L_1} + 1$, and hence they are positioned in rows and columns of the same $i$. The remaining sections are ordered sequentially in the matrix, such that $L_2$ is after $L_1$, $L_3$ after $L_2$, and $L_4$ after $L_3$. After defining the position of the vertices in the matrix, and placing the $L_1 \rightarrow M$ edges, the algorithm continues into the $L_{21}$ section (the same logic will apply for $L_{32}$, $H_1$ and $H_2$). In the sub-network $L_{43}$, each vertex of $L_4$ has exactly one edge to a distinct vertex of $L_3$, guaranteeing the connectivity rule.

Each vertex in $L_2$ has a specific out-degree, and each vertex in $L_1$ has a specific in-degree. The highest degree vertex is selected, independently if it is from $L_1$ or $L_2$, and in the case of several vertices having the same degree, one is randomly selected into variable $v_s$. If $v_s$ is from $L_1$, edges are added randomly to vertices in $L_2$ towards $v_s$, and if the $v_s$ is from $L_2$, edges are added randomly to vertices in $L_1$ coming from $v_s$. After attributing to the selected vertex $v_s$ all its edges, $v_s$ is removed from the algorithm since its expected degree is matched. The remaining vertices that also verify their expected degrees, at any step, are immediately excluded from the algorithm. For example, there are a few vertices expected to have degree equal to one, and at any point they can be assigned an edge, and if so, those vertices are removed from the rest of the attribution process. The final step is to add the upward edges.

The output is an adjacency matrix, which is a mathematical representation of the network. The adjacency matrix is used to calculate network measures and compare empirical and synthetic networks. The algorithm can create any number of synthetic networks with the same size, respecting the logic to build them, but slightly different from each other, allowing to create samples of networks, in this case respecting the Automotive Industry characteristics.

https://doi.org/10.14669/AM/214254

# 7. Conclusions

The network algorithm we have developed has the goal of partitioning and analysing the empirical networks from our collected sample by averaging their properties. Consequently, this algorithm is not entirely independent of these networks. Hence, one real network was excluded from the network sample used in the creation of the model. This network will be used to compare with the results from the model, as an efficiency mechanism, and to be part of the validation process, which encompasses testing the networks for network measures such as average path length, clustering coefficients, and degree distributions. Similar results between real and synthetic networks, i.e., closer than the results found in random networks, are indicators of efficiency of the algorithm.

To systematically compare the synthetic networks with the selected real network $G_S$, we have run the algorithm 30 times to create 30 synthetic networks of the same size of $G_S$, $n$ = 425. In addition, 30 random networks were also created, with 425 vertices each, to work as null model. In Table 5 we show the results for the local and global CC, showing that the real and synthetic networks are within the same magnitude, while the random networks output considerably lower values, which is in accordance with random graph theory: the CC is approximated by the ratio between the average degree and the number of edges, which is predicted to be close to 0.01 in this case. This means that the structure of the synthetic and real networks are similar and not a result of randomness.

Considering the average path lengths, we have studied a twofold: the usual metric, and one specific for the APL from any vertex to the manufacturer $M$. The value of 1.725 for the latter is an indicator of an efficient supply chain because it takes in average less than two edges to reach $M$, and because it is lower than the logarithm of the number of vertices (the small-world test). The synthetic networks have an average path length to $M$ of 1.713. The random networks do not have the same overall structure, such as the existence of a focal vertex like the manufacturer, which makes the calculation of the average path length to $M$ impossible. Regarding the usual average path length, the results show a higher value than the small-world test, however, these measures are performed in the reduced networks and not the actual empirical ones, where the number of vertices is considerably larger and the logarithm becomes larger than the average path length.

**Tab. 5. Comparison of the local and global clustering coefficients of real, synthetic, and random networks**

| Quantities | Real | Synthetic | Random |
|---|---|---|---|
| Average local CC | 0.225 | 0.206 ± 0.009 | 0.010 ± 0.003 |
| Global CC | 0.043 | 0.036 ± 0.001 | 0.011 ± 0.002 |

https://doi.org/10.14669/AM/214254

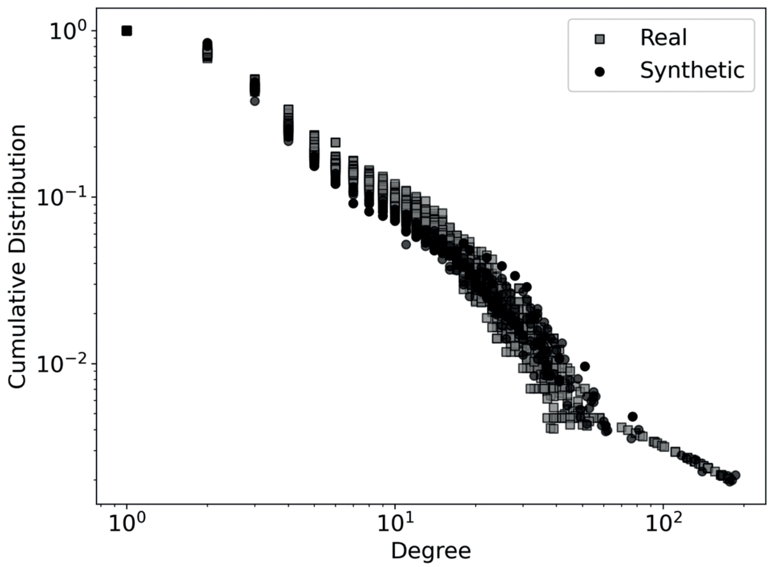


**Fig. 5. Cumulative degree distribution of the real and synthetic networks**

The degree distribution is related to the probability of finding a vertex with a certain degree. The specific function mapping the degree distribution is said to contain important information regarding the networks' structure and dynamics. For example, it has been argued that degree distributions with power-law shapes imply that the networks are more stable due to its heterogeneity of degrees, creating frequent supply relationships between high-degree vertices and low-degree vertices. Testing the in-degree and out-degree distribution of the real and synthetic networks, we verify the presence of similar main trends in both, guaranteeing the existence of common structural and functional features (see Figure 5).

The summary of the network measures comparing real and synthetic networks can be analysed in Table 6, where the $\alpha$ and $\beta$ are the power-law parameters from the fitting of the degree distributions. The results show that the synthetic network output from the network algorithm we have developed respect the main structural characteristics of the real networks, reproducing the identical features verified by the network measures. This answers our research question, since it reproduces with a high degree of accuracy, the structural features of automotive supply chains.

The creation of such a network algorithm raises the debate on the complexity and robustness of the networks. It has been noted that tree-like networks are not stable against random attacks, and that the link between complexity and stability is not as evident as thought before [19]. However, literature defines the supply chain's ability to recover from disruptions

https://doi.org/10.14669/AM/214254

or catastrophes, such as the COVID-19, as resilience, and the evidence on supply chain resilience may contradict the lack of robustness we claim in this work, so further investigation is needed [8, 10, 17].

The suggestions for future work and here presented:

- Partitioning process

The partitioning and modelling process should be repeated in different fields to understand if this mechanism is restricted to networks in a supply-chain configuration, or not. Food webs, for example, are structurally different and do not possess a focal vertex, and adjustments shall be made.

- Structural and functional stability

A global discussion on the structural analysis should be done, especially in automotive supply chains. It is established that a first level analysis on stability (robustness or resilience) comes from removing vertices or edges. If automotive supply chains are mostly tree-like, almost any removal will disconnect the network, which will halt the overall production. However, there are plenty of examples of disruptions, such as tsunamis, earthquakes, economic drops, and others, followed by many different recovery rates which depend on a variation of factors [9, 12]. Hence, further investigation into the role of structural stability into overall robustness and resilience, should be addressed.

The achievement of this algorithm is significant for decision-makers in supply chain management, as it enables them to formulate and test hypotheses using synthetic networks prior to applying them to actual supply chains. Moreover, the creation of synthetic analogues of real networks opens avenues for exploring structural and dynamic modifications, such as assessing risk disruptions in model networks, thereby offering insights for continuous enhancements in real-world applications.

**Tab. 6. Comparison of Network Measures Between Real and Model Networks**

| Quantities | Real Network | Model Network |
|---|---|---|
| Number of vertices | 425 | 425 |
| Average local CC | 0.23 | 0.20 |
| Global CC | 0.04 | 0.04 |
| APL | 3.17 | 3.17 |
| APL to Manufacturer | 1.72 | 1.71 |
| $\alpha$ | 1.12 | 1.08 |
| $\beta$ | -1.00 | -1.01 |
| Adjusted $R^2$ | 0.93 | 0.97 |

https://doi.org/10.14669/AM/214254

# 8. Nomenclature

| | |
|---|---|
| $a_{ij}$ | Matrix element of the adjacency matrix A, indicating the row $i$ and column $j$ |
| AN | Automotive Networks |
| APL | Average Path Length |
| CC | Clustering Coefficient |
| H | Horizontal layer |
| JLR | Jaguar Land Rover |
| $k$ | Degree |
| $l$ | Number of edges |
| $L$ | Layer |
| LCC | Largest Connected Component |
| M | Manufacturer |
| $n$ | Number of nodes |
| OEM | Original Equipment Manufacturer |
| SCM | Supply Chain Management |
| $v_s$ | Selected vertex |

# 9. References


[1] Bell M, Perera S, Piraveenan M, Bliemer M, Latty T, Reid C. Network growth models: A behavioural basis for attachment proportional to fitness. Scientific Reports. 2017;7(1):42431. https://doi.org/10.1038/srep42431.

[2] Brahma A, Wynn DC. Concepts of change propagation analysis in engineering design. Research in Engineering Design. 2023;34(1):117–151. https://doi.org/10.1007/s00163-022-00395-y.

[3] Brintrup A, Ledwoch A, Barros J. Topological robustness of the global automotive industry. Logistics Research. 2016;9(1):1–17. https://doi.org/10.1007/s12159-015-0128-1.

[4] Ferreira AS, Sacomano Neto M, Candido SEA, Ferratti GM. Network Centrality and Performance: Effects in the Automotive Industry. Revista Brasileira de Gestao de Negocios. 2021;23(4):677–695. https://doi.org/10.7819/rbgn.v23i4.4132.

[5] Fessina M, Zaccaria A, Cimini G, Squartini T. Pattern-detection in the global automotive industry: A manufacturer-supplier-product network analysis. Chaos, Solitons and Fractals. 2024;181:114630. https://doi.org/10.1016/j.chaos.2024.114630.

[6] 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. Scientific Reports. 2022;12(1):11847. https://doi.org/10.1038/s41598-022-13996-3.

[7] Kang A, Oh J. The configuration and evolution of Korean automotive supply network: an empirical study based on k-core network analysis. Operations Management Research. 2023;16(3):1251–1270. https://doi.org/10.1007/s12063-023-00349-0.

[8] Karamoozian A, Tan CA, Wu D, Karamoozian A, Pirasteh S. COVID-19 automotive supply chain risks: A manufacturer-supplier development approach. Journal of Industrial Information Integration. 2024;38:100576. https://doi.org/10.1016/j.jii.2024.100576.

[9] Kim Y, Chen YS, Linderman K. Supply network disruption and resilience: A network structural perspective. Journal of Operations Management. 2015;33–34:43–59. https://doi.org/10.1016/j.jom.2014.10.006.

https://doi.org/10.14669/AM/214254


[10] Ledwoch A, Brintrup A, Mehnen J, Tiwari A. Systemic Risk Assessment in Complex Supply Networks. IEEE Systems Journal. 2018;12(2):1826–1837. https://doi.org/10.1109/JSYST.2016.2596999.

[11] Mungo L, Brintrup A, Garlaschelli D, Lafond F. Reconstructing supply networks. Journal of Physics: Complexity. 2024;5(1):12001. https://doi.org/10.1088/2632-072X/ad30bf.

[12] Ogawa Y, Akiyama Y, Yokomatsu M, Sekimoto Y, Shibasaki R. Estimation of supply chain network disruption of companies across the country affected by the nankai trough earthquake Tsunami in Kochi city. Journal of Disaster Research. 2019;14(3):508–520. https://doi.org/10.20965/jdr.2019.p0508.

[13] Perera S, Bell MGH, Bliemer MCJ. Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective. Applied Network Science. 2017;2(1):33. https://doi.org/10.1007/s41109-017-0053-0.

[14] Perera S, Perera HN, Kasthurirathna D. Structural characteristics of complex supply chain networks. 3rd International Moratuwa Engineering Research Conference, MERCon 2017. 2017:135–140. https://doi.org/10.1109/MERCon.2017.7980470.

[15] 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. Scientific Reports. 2022;12(1):13347. https://doi.org/10.1038/s41598-022-13104-5.

[16] Shishodia A, Sharma R, Rajesh R, Munim ZH. Supply chain resilience: A review, conceptual framework and future research. International Journal of Logistics Management. 2023;34(4):879–908. https://doi.org/10.1108/IJLM-03-2021-0169.

[17] Singh J, Hamid ABA, Garza-Reyes JA. Supply chain resilience strategies and their impact on sustainability: an investigation from the automobile sector. Supply Chain Management. 2023;28(4):787–802. https://doi.org/10.1108/SCM-06-2022-0225.

[18] Um J, Han N. Understanding the relationships between global supply chain risk and supply chain resilience: the role of mitigating strategies. Supply Chain Management. 2021;26(2):240–255. https://doi.org/10.1108/SCM-06-2020-0248.

[19] Wiedmer R, Rogers ZS, Polyviou M, Mena C, Chae S. The Dark and Bright Sides of Complexity: A Dual Perspective on Supply Network Resilience. Journal of Business Logistics. 2021;42(3):336–359. https://doi.org/10.1111/jbl.12264.

[20] Yang D, Tang M, Ni Y. Robustness of automotive supply chain networks based on complex network analysis. Electronic Commerce Research. 2025;25(5):3609–3636. https://doi.org/10.1007/s10660-024-09814-9.