DEMB Working Paper Series

N. 116

Clusters of specializations in the automotive supply chain in Italy. An empirical analysis using text mining

Pasquale Pavone*, Margherita Russo**

August 2017

* CAPP, Center for the Analysis of Public Policies
Email: pasquale.pavone@unipd.it

** University of Modena and Reggio Emilia
CAPP, Center for the Analysis of Public Policies
Address: Viale Berengario 51, 41121, Modena, Italy
Email: margherita.russo@unimore.it

ISSN: 2281-440X online

DEMB WORKING PAPER SERIES n. 116

Clusters of specializations in the automotive supply chain in Italy. An empirical analysis using text mining

Pasquale Pavone* e Margherita Russo♦

August 2017

*Centro di Analisi delle Politiche Pubbliche, Università di Modena e Reggio Emilia, pasquale.pavone@unipd.it

♦Dipartimento di Economia Marco Biagi and Centro di Analisi delle Politiche Pubbliche, Università di Modena e Reggio Emilia, margherita.russo@unimore.it

Abstract

The wide literature exploring supply chains is polarized on two perspectives: micro analyses focusing on management strategies of companies, and macro assessment of cross-country interdependences. In order to explore the ongoing innovation paths, this paper adopts a third perspective on the supply chain, focusing on the internal structure of specializations within the automotive supply chain in Italy. If we compare the degree of fragmentation across global value chains, the automotive supply chain has the highest degree of fragmentation. With regard to Italy, its structural characteristics (number, size of companies, location) and dynamics of change deserve attention both for its large share in domestic production and for its interconnections with other supply chains. In this paper, we explore a strategy to identify a classification of specializations within the automotive supply chain grounded on the textual description of activities provided by companies when they register their business. Pending the acquisition of the database for the other years of the Observatory, in this work the analysis refers only to 2017 data.

Keywords: automotive supply chains, industrial specialization, fragmentation, textual analysis, regional analysis, similarity analysis, IRaMuTeQ, Taltac2

JEL classification: L62, R12, Z13

Acknowledgments

The authors wish to thank ANFIA (the National Association of Automotive Industries) and Turin Chamber of Commerce for providing the database. The opinions expressed and arguments put forward are those of the authors. Comments and suggestions by Giovanni Bonifati and Michele Lalla contributed in enhancing interpretation of the results.

Introduction. Automotive supply chain in the changing industrial structure

As a complementary contribution to the 2017 edition of Italian Observatory of the automotive supply chain, this paper addresses some issues of classification issues at macro and meso level to help in analysing the spatial agglomeration and specific configuration of "snakes and spiders" in the Italian automotive supply chain. It suggests a methodology to single out the set of companies involved in the automotive supply chain and assesses the robustness of the empirical investigation undertaken with the 2017 Italian survey. It paves the way for a more generalized procedure of identification of companies involved in the supply chain by using big data drawn from the company description in the Registry of enterprises and from their website.

Supply chains are defined as the “full range of activities that firms and workers do to bring a product from its conception to its end use and beyond” (Gereffi and Fernandez-Stark, 2016). They are at the core of both microanalysis, focusing on strategic management of companies within their value chain, and macro assessment of cross-country interdependences. In particular, the automotive supply chain has a prominent role in the discussion on the changing industrial structure. By using the OECD inter-country input-output (ICIO) model[1], De Backer and Miroudot (De Backer and Miroudot, 2012) estimate the length of global value chains (GVCs), i.e. the highest number of production stages (as proposed by Fally, 2011; De Backer and Miroudot, 2012, p. 6; OECD, 2017). They show that, in the period 1995-2005, the average length of GVCs across all industries increases, mainly because of their internationalization. When industries are ranked according to the length of their GVC, we observe that the vehicle and motor industry has the highest value. In this supply chain there is a large share of domestic intermediate products (almost two thirds). In the cross country comparison, Korea, China, Turkey have a higher length of the motor vehicle supply chain (with relatively less vertically integrated production) but a greater importance of domestic production of intermediate goods, about 90%; Italy, France and Japan have a similar length, but Japan relies almost 100% on domestic production. As Sturgeon has pointed out (Sturgeon et al., 2008; Sturgeon et al., 2009) two main directions have characterized changes in the automotive industry: decreasing fragmentation and increased regionalization.

Baldwin and Venables (2013) propose a general framework to analyse the drivers for the spatial unbundling of the supply chains in terms of processes characterized by a mix of "snakes and spiders" configurations. Snakes are processes in which goods move in a sequential manner from upstream to downstream. Spiders are processes formed by multiple parts coming together to form the final product itself or a component (such as a module in the auto-industry). At country level, an

[1] The model covers 56 countries (55 OECD and non-OECD economies plus the “rest of the world”) and 37 industries, defined on the basis of the ISIC Rev. 3 classification and harmonized across countries (De Backer and Miroudot, 2012), p. 3).

empirical analysis of such a framework needs to single out the companies embedded in the supply chain. So far, its perimeter has been delimited by using input-output analysis or expert knowledge, but its internal structure also needs to be defined. In this paper we present those classifications and elaborate a methodology to enhance information available from official sources in order to generalize the classification of the internal structure of the automotive supply chain, beyond the hierarchical classifications, and aggregation by digit that can be adopted by using standard classification of economic activities.

The structure of the paper is the following. Section 1 presents the perimeter of the Italian automotive supply chain, as it emerges from the input-output tables and expert classification. Section 2 describes the methodology for analysing the corpus of texts. It adopts three complementary tools in a sequential way. After the text corpus analysis (pre-treatment and lexical processing), extraction of information implements a multidimensional analysis of words, in order to identify clusters of specializations, and a similarity analysis of words, to provide indications on clustering of specializations as they are described by companies. Section 3 characterizes the corpus composed of the documents describing the activities of the companies from the Observatory's archive. Section 4 reports the results of two multidimensional analyses: on spatial clustering of activities and on the specializations characterizing the supply chain; a complementary analysis of co-occurrences in a semantic network is discussed to frame the many products and processes characterizing this supply chain. Section 5 presents the main results of the empirical analysis by comparing the classification that emerged from the text analysis with Nace rev.2 classification at 4 digits, highlighting a significant clustering that otherwise would have been difficult to obtain just by a hierarchical aggregation of digits. In section 6, the semantic classification is implemented to compare the population of companies in the automotive supply chain and the respondents to the 2017 online survey, providing some indications on the quality of the survey, beyond the information acquired by analysing size and location of the companies. Section 7 concludes the paper outlining the development path of an automatic text analysis to identify the perimeter of automotive components over the last twenty years and to analyse the dynamics of changes of its internal structure. Supplementary materials are displayed in Annexes 1-7.

1. Boundaries and partitions of an industrial structure

The perimeter of the Italian automotive supply chain, as it emerges from the input-output tables

With regard to Italy, the symmetric input-output table 2010BxB (63 branches), elaborated by Istat [SIMM_TOT_63BxB.xls], shows that added value in the "Manufacture of motor vehicles, trailers and semi-trailers" (V29) industry is 22.7% of production; imports (cif) are 19.2% of intermediate consumption of goods and services. By grouping the branches according to the macro-categories of goods and services, see Table 1, we observe that about 32% is the share of intermediate goods produced by the auto industry (V29) and by the manufacturers

of metal products (except machinery and equipment) (V25), respectively 16.6% and 15.4%. A further 25% are products from four branches: machinery and equipment; metallurgical activities; manufacture of rubber and plastic materials; manufacture of electrical equipment (respectively: Branches: V28, V24, V22, V27). Design, research and scientific activities (grouping branches V62_63, V69_70, V71, V72, V74_75) account for 5.7%; 5% are intermediate goods from commercial activities.

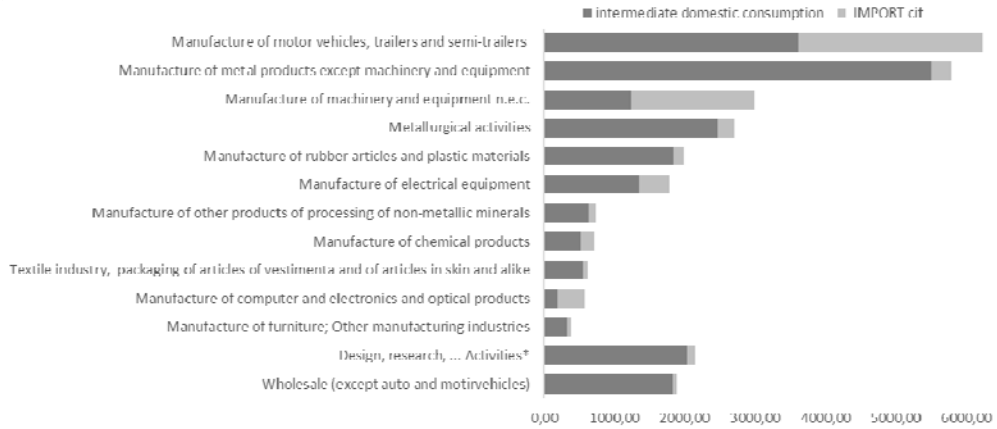
Figure 1 highlights the relative shares of domestic consumption of intermediate goods and services and imports in the automotive supply chain: the higher shares of import are those coming from branches V29 (Manufacture of motor vehicles, trailers and semi-trailers) and V28 (Manufacture of machinery and equipment n.e.c.), respectively 48% and 56% is produced abroad.

Table 1 Manufacture of motor vehicles, trailers and semi-trailers: total input and import of intermediate branches, Italy, 2010

		STOTBB_2010		SIMPBB_2010		
		consumption at basic prices	% on total consumption at basic pries	IMPORT cif	% on total import cif	share of import on intermediate consumption
	Branches (NACE*63)	branch V29		branch V29		
V29	Manufacture of motor vehicles, trailers and semi-trailers	6 232	16,55	2 620	36,15	42,04
V25	Manufacture of metal products except machinery and equipment	5 787	15,36	279	3,86	4,83
V28	Manufacture of machinery and equipment n.e.c.	2 994	7,95	1 755	24,21	58,61
V24	Metallurgical activities	2 707	7,19	244	3,37	9,01
V22	Manufacture of rubber articles and plastic materials	1 985	5,27	137	1,90	6,92
V27	Manufacture of electrical equipment	1 786	4,74	435	6,00	24,34
V23	Manufacture of other products of processing of non-metallic minerals	734	1,95	101	1,39	13,72
V20	Manufacture of chemical products	707	1,88	178	2,45	25,11
V13_15	Textile industry, packaging of articles of vestimenta and of articles in skin and alike	619	1,64	63	0,87	10,23
V26	Manufacture of computer and electronics and optical products	570	1,51	386	5,33	67,71
V31_32	Manufacture of furniture; Other manufacturing industries	382	1,01	58	0,80	15,11
*	Design, research, ... Activities*	2 151	5,71	121	1,68	5,64
V46	Wholesale (except auto and motirvehicles)	1 883	5,00	44	0,61	2,35
	Other branches	*9 128*	*24,23*	*825*	*11,39*	*9,04*
	Total	**37 667**	100,00	**7 246**	100,00	

* V62_63, V69_70, V71, V 72, V74_75
Source: our elaboration on Istat, Input-output symmetric tables (total, STOTBB, and import, SIMPBB) 63x63 branches, with branch technology, at basic prices (ISTAT_STOTBB_2010 and ISTAT_SIMPBB_2010), million euros

Figure 1 Manufacture of motor vehicles, trailers and semi-trailers: domestic intermediate consumption and import of intermediate goods and services, main branches, Italy, 2010



Source: our elaboration on Istat, Input-output symmetric tables (total, STOTBB, and import, SIMPBB) 63x63 branches, with branch technology, at basic prices (ISTAT_STOTBB_2010 and ISTAT_SIMPBB_2010), million euros

The picture emerging from input-output analysis allows us in delimiting which are the most important branches to be considered in the automotive supply chain, but

we cannot disentangle their internal structure, in terms of technologies, level of vertical integrations, size and location of the companies.

Using input-output matrix, a recent survey on the automotive industry in the main European countries (Prometeia and UnionCamere, 2015, pp, 39-42) proposes a classification of the core of the supply chain defined by focusing on seven areas: the upstream phase of supply of raw materials (steel, foundries, metal ingots), followed by first processing (of goods used in the downstream phases: paints and varnishes, glass, rubber and plastic, metalworking), intermediate machining (bodywork, motors, mechanical components, electronic components and batteries, tires), and final machining (assembly and production of vehicles). A complementary phase is provided by engineering and design services sourcing. Distribution includes dealers, aftermarkets, leasing and specialized hire, and vehicle repair. This classification is detailed in terms of Nace classification, highlighting the phases totally related to the automotive industry (p. 42). For the phases only partially related to the automotive supply chain, weights are implemented to estimate their total size (via input–output analysis); the number of companies and employees is estimated accordingly[2]. Although some estimates can be derived for a cross country analysis (as the Prometeia-Unioncamare study provides), the macro aggregation available, as well as input-output analyses in general, does not allow an enquiry of the internal industrial structure and this is the reason why that study cannot be used if we need to analyse the dynamics of changes in the automotive supply chain.

The perimeter of the Italian automotive supply chain, as defined by the Observatory

To monitor its peculiar characteristics and the ongoing transformations, since 1997 the automotive supply chain in Italy is studied through the Automotive Observatory, realized by the National Association of Automotive Industries (ANFIA), the Turin Chamber of Commerce[3]. According the Observatory, in 2016, there were about two thousand enterprises employing about 136,000 employees, with a turnover of 38.8 billion euros, of which nearly 20 billion euros of exports. Those companies do not include the car makers, nor auto dealers and maintenance services.

In the past twenty years, the Observatory has changed the boundaries of the automotive supply chain according to the specific focus on the main final product: from one on all transport vehicles to the most recent focusing only on cars. Data

[2] Some hypotheses, accepted in most of the input-output analyses (no economies of scale, same level of vertical integration of companies within each branch), are adopted also by the study conducted by Prometeia and UnionCamere (2015, see p. 204).

[3] Since 2016 the Observatory has the scientific direction of the Center for Automotive & Mobility Innovation (CAMI) of Ca' Foscari University of Venice-Department of Management. In the 2017 edition, the Observatory has the collaboration of the Center of Public Policy Analysis (CAPP-DEMB) of the University of Modena and Reggio Emilia-Economics Department Marco Biagi.

on the population of companies is individually checked across several sources (the Register of Companies, websites, the balance sheets) to update the list with entry/exit of companies. In particular, companies in the automotive supply chain are classified by Anfia in the following 2-digit Nace rev. 2 divisions: 20-30, 33, 71-72 and 74[4]. Those divisions include the ones listed in Table 1, but the specific list of companies is partly based on the list of members of Anfia, integrated by other secondary sources.

Specializations and co-location of suppliers and innovation processes

In the discussion of the changes in the supply chains, Helper and Sako (Helper and Sako, 2010) underline that interdependent decision-making, allowing the achievement of economies of scale and scope, require "managerial structures that make decisions based on hierarchical commands and also on discussion among peers". They observe that "customer firms increasingly rely on their suppliers' specialized skills, but have less private incentive to invest in improving those skills, since suppliers are shared with their rivals" (p. 424). To explore this perspective in the automotive supply chain in Italy, we need to assess the specialization and location of the suppliers and their degree of dependence on customers. With regard to Italy, co-location of suppliers - a phenomenon investigated by Sako (Sako, 2005) in a cross-country analysis of 'automotive supplier parks' - has a path dependence that can be traced back to the competence networks in the mechanical industry concentrated in a wide area of Northern regions (Piedmont, Lombardy, Emilia-Romagna), as discussed by Rinaldi and Russo (Rinaldi, 2008; Russo, 2008). In 2016, the VW and FCA decisions to increase in Emilia-Romagna - respectively, production of a car model in Lamborghini and R&D of Maserati - goes in the direction of expanding the presence of car manufacturing and design in ecosystems where both car manufacturers of sport and racing cars and their suppliers are located, benefiting from a still remarkable variety of high level competence networks and employees' skills.

The structural characteristics (number, size of companies, location) and dynamics of change of the Italian automotive supply chain deserve attention not only for the high share of domestic production, but also for the interconnections with other supply chains. In Italy, producers of automotive components are intertwined in the mechanical production system characterizing Italian manufacturing of special machinery of the Northern regions. From the analysis of the internal structure of

[4] The Nace rev. 2. divisions considered by Anfia refer to manufacturing of: chemicals and chemical products (20); basic pharmaceutical products and pharmaceutical preparations (21); rubber and plastic products (22); other non-metallic mineral products (23); basic metals (24); fabricated metal products, except machinery and equipment (25); computer, electronic and optical products (26); electrical equipment (27); machinery and equipment n.e.c. (28); motor vehicles, trailers and semi-trailers (29); other transport equipment (30).
Other divisions group: repair and installation of machinery and equipment (33); professional services, encompassing, in particular, architectural and engineering activities, technical testing and analysis (71); scientific research and development (72); other professional, scientific and technical activities (74).

the machinery supply chain, Russo (Russo, 2015) concluded that after the crisis there was an increase in the vertical integration of companies in those regions, with many sub contractors integrating downward their production processes and extending their aftermarket services. She observes that increased vertical integration in production stages would impact on innovation processes. In fact, the process of vertical disintegration of companies, specializing in one or few phases of the process of producing goods and services in the mechanical sector is a key element that fuels technological convergence (Rosenberg, 1963) between firms and production chains, and enhances the degeneracy of the system, i.e. a process in which structurally different elements provide features that overlap the ones already offered, and often support temporary adjustments from which innovations emerge (Bonifati, 2013; Lane, 2010). Therefore, an increase in the level of vertical integration affects the path of innovation within and across the supply chains, and the organization of the industry.

Henceforth, by identifying the structure of specializations and locations in the automotive supply chain we could ground the structural configuration across this chain to analyse the dynamics of changes in the last decades and to highlight productive complementarities and territorial specializations, supporting the study of innovation paths.

A technical and relational classification within the supply chain

The expert perspective of Anfia on the automotive supply chain has been complemented, through the Observatory, by an in-depth investigation of its internal structure through an empirical investigation, using an online questionnaire. Among other questions, companies are asked to classify themselves[5]. The classification proposed by Anfia, presented in Annex 1, is grounded on a detailed knowledge on the automotive supply chain, in terms of technical specialization and relational conditions that further specifies what is proposed by Memedovic (Sturgeon et al., 2009, p 7; Memedovic, 2007)[6].

Anfia characterizes companies in two major categories: suppliers of complex systems and/or entire vehicle parts; suppliers of specific components. The latter encompasses both companies that sell products in the OEM (Original Equipment

[5] "The following companies are excluded from the analysis: Automobile Manufacturers or Industrial Vehicles (companies engaged in the production of cars, light commercial vehicles, limited-edition sports cars, industrial vehicles, buses, military vehicles, armoured cars); Special Vehicle Manufacturers / Standards (Companies specializing in the production of special vehicles such as ambulances, cold storage vehicles, disabled vehicles, crane trucks and elevating platforms, sweepers, waste transportation vehicles, drainage vehicles, etc…). In addition, Trailers and Semi-trailer producers are not considered part of the supply chain. Equipment Manufacturers, Machine Tools, Robots, Automation Systems for the Industry are not included, too". Souce: Anfia 2017

[6] Memedovic (2007) groups specializations in five main areas: (a) design/product development; (b) raw materials (rubber, glass, steel, plastic and aluminium), (c) OEM parts (including bodies, mechanical and electronic components, tire and rubber hoses, seats, windshield, air bags, lighting, batteries, engines, transmissions ad replacement parts, OEM of parts), (d) OEM assembly, (e) after sale services.

Manufacturer) or in the spare parts market (aftermarket) and companies that sell their products essentially in the aftermarket. In the supply chain there are also subcontractors, second/third tiers that supply components to complex systems or to specialized producers. In this category are also classified companies that produce non-high-performance fastening systems. Complementary specializations are those of companies engaged in the design, engineering, design/style phases of the realization of some vehicle systems (e.g. powertrain, interior, etc.).

This classification, essentially a technical and relational one, could support an analysis of the automotive supply chain in terms of "spiders and snakes". However, there is no way to derive it from the Nace classification. One could outline that structural and relational composition of the supply chain only from the empirical investigation.

Since the 1990s, international classification of business companies has been harmonized across taxonomies of technologies and economic activities (Isic, Nace), products are classified as well (Hs, Cn, Cpc, Cpa, Prodcom, Sitc) and tables of conversion are officially recognized by Eurostat (RAMON service). Those classifications do not highlight the relational dimension of the activities undertaken by companies, nor is that dimension generally detected by ordinary investigations of the industrial structure. Although activities and product taxonomies are crucial for several aspects of cross-country analysis, when we analyse supply chains we need to implement complementary strategies in creating significant clusters of activities that could not be obtained just aggregating digits.

To overcome the limitation of both input-output and expert classifications, to enhance information available from official sources and to generalize the classification of its internal structure, in this paper we propose a text analysis and a check of the results of the emerging classification against the respondents of the 2017 online survey.

The methodology is grounded on a multidimensional text analysis of the description of activities provided by companies when they register their business. It can be generalized to outline the perimeter of the supply chain and to encompass the integration of text data from the companies' websites. Pending the acquisition of the original database for the other years of the Observatory and of information from websites, in this paper the analysis refers only to 2017 data received from Anfia of their archive on the Italian automotive supply chain (as described above), elaborated by an expert selection.

2. Methodology

For the exploration of the corpus and the presentation of the results, the analysis was carried out by integrating a number of software tools: Taltac2[7] (Bolasco, 2010) to identify lexemes (Lexicon Analysis Units), both as simple and multi-word forms, occurring in the corpus; Spad, for Multidimensional Analysis and

[7] http://www.taltac.it/it/index.shtml

Cluster Analysis; Iramuteq[8], for the analysis of co-occurrences of words and similarity analysis; Gephi, to display the graph resulting from the similarity analysis.

The analysis is characterized by three stages: pre-treatment and lexical processing; multidimensional analysis and network of co-occurrences.

Pre-Treatment of texts and lexical analysis

In the present analysis, texts are the description of the activities and products of the companies. Pre-Treatment of texts consists in the acquisition of text by numerical indexing of words (their tokenization) for the recognition of the units of analysis, and in the creation of the document database.

The lexical analysis was carried out using Taltac2 software. This analysis provides: (a) the identification of lexical analysis units (the lexemes, "words"), intended both as single words and multi-words; (b) the attribution of grammatical meta-information to the "words" constituting the Vocabulary; (c) the selection of content words (nouns, verbs, adjectives) as the keywords to be analysed; (d) the probability-based analysis of the meaningful terms characterizing the specific language used in describing the activities and products mentioned by individual companies.

Text mining: specific language to single out activities and products

A specific language representation was elaborated on a factorial plan[9], which graphically presents the combinations of a matrix of *Characteristic x Documents*. In our corpus, a document is the sub-text associated with each company. The position of words on the factorial plan is a function of the association of their occurrences in the sub-texts, thus expressing their similarity or diversity: two words are close because they are present in the same sub-texts. At the centre of the factorial plan are the most common terms between the different languages of the various sub-texts.

Through a correspondence analysis (CA), the row and column elements of the matrix are mathematically formalized as vectors, and the above profiles are represented by points in a multidimensional space. The distances between the lexical profiles are measured using a weighted Euclidean metric (chi-square metric). The complex multidimensional space of the variables [words, in our case] is then reduced to a few key factors that can represent, on dimensions named "factorial axes", the relationships between the elements of the data matrix. CA produces the best simultaneous representation of row profiles vs. column profiles in each factorial plan, and on each of its axes (Bolasco, 2013). To do this, consider a matrix *Documents x Keywords* (n × p). The matrix profiles the different documents to be classified with the occurrences of the keywords contained therein.

[8] http://www.iramuteq.org/
[9] Factorial analysis was done with Spad.

By treating this matrix to the classic sequence of multidimensional statistical analysis (simple match analysis and cluster analysis), you will get a partition in *K* document groups, homogeneous within them and heterogeneous among them: cluster analysis consists of grouping sub-texts according to their maximum lexical similarity, and thus characterizing them in terms of activities/products declared by enterprises. The semantic field of expression of this homogeneity can be visualized by the proximity of the corresponding terms on the factorial plan. "The resulting *K* groups [of business companies, in our case] represent disaggregated classes and the word lists associated with each group define the theme or activity of the group" (Bolasco, 2013).

Text mining: semantic network of co-occurrences to single out specializations

The last phase of our exploratory analysis consists in a word similarity analysis (Flament, 1962; Flament, 1981; Marchand and Ratinaud, 2012). The purpose of this analysis is to study the proximity and relationship between the elements of a set. In the textual case, therefore, it allows us to clarify the relationships between words, based on their co-occurrences within the fragments of texts (the length of a fragment is a sentence or at most 40 words). By co-occurrence we mean the existence of a pair of terms within fragments. Through the Iramuteq software, it is possible to perform this analysis based on an algorithm that measures how many times words are two-by-two "next" in the fragments of texts. As a result, a co-occurrence matrix is obtained, whose generic term contains the number of co-occurrences between each pair of words. A graph analysis of this matrix produces some synthesis of the relationships between the lexical units considered. In particular, "semantic communities" are generated through the modularity algorithm[10], which identifies sub-networks of keywords within the corpus and outlines the structure of the co-occurrence network. The graphical representation of co-occurrence networks is elaborated with Gephi[11] (Bastian et al., 2009). In the co-occurrence matrix there are keywords with more than 10 occurrences, but the graph represents all co-occurrences. The optimization algorithm for visualizing the graph is the Fruchterman-Reingold one. To each semantic sub-network is associated a colour. The vertices' size is proportional to the word's frequency and the edges' width indicate the word's co-occurrence strength.

3. Corpus of texts

Description of the corpus

The corpus describing the activities of companies consists of 2,026 documents, one for each company in the automotive supply chain in Italy, selected by Anfia. Each document is made up of two sections of text, one with the "Description of Activity" the second one with the "Verified Description of the Activity", carried out by Anfia's experts. Each document is classified according to six main

[10] See Fortunato and Hric (2016), for details on this community detection algorithm.
[11] Gephi is a software for displaying and exploring graphs and networks, https://gephi.org/.

categories, respectively: two for the location of the enterprise (the region, at NUTS 2 level, and the province, at NUTS 3 level), the employee class (with 10 classes[12]), and three categories for the activity classification (at 2, 3, and 4 digits of the Ateco 2007 codes, corresponding to the Nace rev. 2).

After a first structuring of textual information within the Vocabulary database and the Documents database, the Corpus of 2,026 documents consists of 4,834 different words, for a total of 53,625 occurrences, subdivided into the 2 sections (29,126 occurrences in the Activity Description and 24,499 occurrences in the Verified Activity Description). The amount of textual information contained in each document is very varied, ranging from a minimum of 2 words to a maximum of 271 words used to describe the activity. In 127 documents, textual information is only available in one of two text sections (123 times only Description Activity and 4 times Verified Activity Only) and in 295 cases the text in the two sections is identical. In 809 instances, text length in the Verified section is longer that the one in the Activity description. Three examples of Documents are presented in Annex 2. Table 2 and Table 3 show the size of the corpus in terms of documents, according to two categorical variables, respectively, regions and classes of employees. These details on the Corpus characteristics underline a non-systematic manipulation of texts that, in fact, should not be done, to fully exploit the potentialities of a completely automatic analysis.

Table 2 Number of documents for the category "Region"

Region	N. of documents	Region	N. of documents
Abruzzo	70	Molise	2
Basilicata	20	Piedmont	724
Calabria	4	Puglia	29
Campania	82	Sardinia	3
Emilia Romagna	209	Sicily	13
Friuli Venezia Giulia	21	Tuscany	42
Lazio	50	Trentino Alto Adige	24
Liguria	11	Umbria	13
Lombardy	528	Valle d'Aosta	4
Marche	29	Veneto	148

Table 3 Number of documents for the category "Class of employees"

Class of employees	N. of documents	Class of employees	N. of documents
1	77	100-249	179
2-5	177	250-499	82
6-9	203	500-999	52
10-19	343	1000 and more	26
20-49	424	n.a.	230
50-99	233		

Identification of lexical analysis units

Through the grammatical annotation and the application of a textual-lexicon model (Bolasco and Pavone, 2007; Pavone, 2010) for the search for syntactic structures, it was possible to identify 186 multi-words (a selection of multi-words with 30 and more occurrences is listed in Table 4). Multi-word recognition has

[12] Classes are assigned to 2015 employees: 1; 2-5; 6-9; 10-19; 20-49; 50-99; 100-249;250-499; 500-999; 1000 and more.

occurred through the application of a series of Regular Expressions that by searching in the Corpus the syntactic structures of the most common lexemes allowed us to identify 4,298 entities. Among these, 186 multi-words were automatically singled out by selecting only those with more than 5 occurrences[13].

Through the grammatical classification of all the graphic forms of the Vocabulary database it is also possible to identify the keywords, i.e. 'active graphic forms' (understood as the content words: nouns, verbs, adjectives), and the 'additional forms' (speech structure words: conjunctions, prepositions, articles, etc.). A selection of active and additional forms is listed in Table 5).

Table 4 Most frequent multi-words in the Corpus

Graphic form	**English translation**	**Occurrences**
veicoli industriali	*Industrial vehicles*	111
conto terzi	*third parties*	100
lavorazioni meccaniche	*mechanical machining*	96
apparecchiature elettriche	*electrical equipment*	89
materie plastiche	*plastic materials*	83
oggetto sociale	*corporate purpose*	79
settore automotive	*automotive sector*	66
rivestimento dei metalli	*metal coating*	55
settore auto	*car industry*	44
minuterie metalliche	*small metal parts*	43
officina meccanica	*mechanical workshop*	41
settore automobilistico	*automotive industry*	37
lavori di meccanica	*works of mechanics*	37
pezzi di ricambio	*spare parts*	37
articoli tecnici	*technical articles*	35
parti meccaniche	*mechanical parts*	34
macchine utensili	*machine tools*	33
parti di ricambio	*spare parts*	32
industria automobilistica	*automotive industry*	30

Table 5 DB Vocabulary: active forms and additional forms

Content forms			**Supplementary forms**		
Graphic forms	**English translation**	**Occurrences**	**Graphic forms**	**English translation**	**Occorrences**
produzione	*production*	1158	di	*of*	4094
fabbricazione	*manufacture*	774	e	*is*	3954
autoveicoli	*vehicles*	681	per	*for*	2458
accessori	*accessories*	584	in	*in*	1198
parti	*parts*	509	ed	*and*	1128
in genere	*generally*	416	la	*there*	820
costruzione	*construction*	396	a	*to*	421
componenti	*components*	356	l'	*the*	380
progettazione	*design*	348	loro	*their*	372
lavorazione	*processing*	303	il	*the*	329

Active forms, ranked by decreasing number of occurrences, provide the general theme of the Corpus under analysis, for example: *produzione, fabbricazione, autoveicoli, accessori (production, manufacture, vehicles, accessories)*. Among these active forms, with a threshold of 5 occurrences, 1,002 keywords were selected for the subsequent phases of the representation of the information. In the

[13] The complete list of multi-words and hapax of all the active forms will be examined when the complete corpus is available.

analysis, however, we consider all the active forms.

4. Results

The corpus′ analysis was developed in two directions. The first one seeks to identify - by a multidimensional analysis - a companies' classification based on the description of their business. The second one proposes a classification of the areas of specialization emerging from co-occurrences.

Multidimensional analysis

The results of multidimensional analysis refer to two different data arrays/sets of data: *Keywords x Categories (mode)* and *Documents x Keywords*. Both matrices are of Boolean type, i.e. internal elements indicate the presence or absence of terms within modes or documents. Usually, for the purposes of this analysis, the occurrence of a word is used, but the high variability of the size of text documents and, above all, redundancy of the terms in each document (such as manufacturing, processing, car vehicle) would produce a significant distortion of the results.

Which activities characterize companies by region and size?

Through multidimensional analysis we first look for which activities characterize companies by location and size. We ground this analysis on lexical content of the documents, by first elaborating the matrix *Keywords x Categories* (1,002 × 30) with the *keywords* selected on a grammatical basis and *categorical variables* (modes) of 20 regions and ten classes of employees. From the distribution of the modalities on the factorial plan (Figure 2) it can be observed that along the first factor there is a polarization between a group of three regions (Lombardy, Piedmont and Emilia-Romagna) - which constitute an industrial area with more similar characteristics – and the rest of the regions (Sardinia, Basilicata, Trentino-Alto Adige, Sicily, Calabria, Molise, Valle d'Aosta). Factor 2, on the other hand, is mainly characterized by the polarization on the size of the company, highlighting SMEs on a pole and the large companies on the opposite side, with classes 10-19, 20-49, 50-99 closer to the average (and typically of companies in Lombardy, Piedmont and Emilia-Romagna).

To interpret the semantics that characterises the categorical variables, one can observe both the keywords distribution in the factorial plan (Figure 3) and the Simple Match Analysis tabs in Annex 3, with the list of categorical variables. The first factor contrasts the production of mechanical and metallurgical parts - in the left quadrants of the factorial plan - with the construction and assembly of accessories and the production of electronic components – characterizing the quadrants on the right in the factorial plan. Factor 2 shows a polarization between replacement, feasibility studies, consultancy services, ideation industrial equipment, at the top, and emissions, sheet metal stamping, extrusion, exchange boxes, transmission systems, glass, conditioners, at the bottom.

Figure 2 Distribution of categorical variables: regions and classes of employees

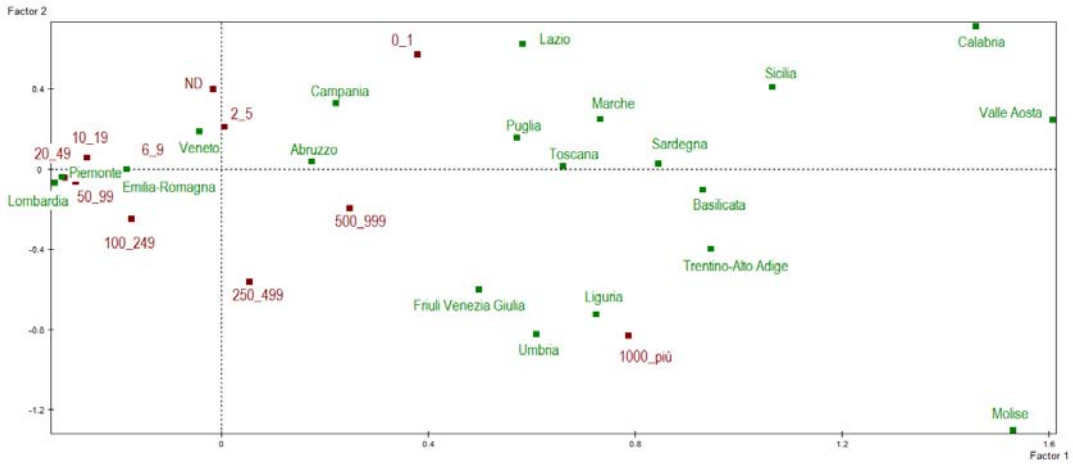
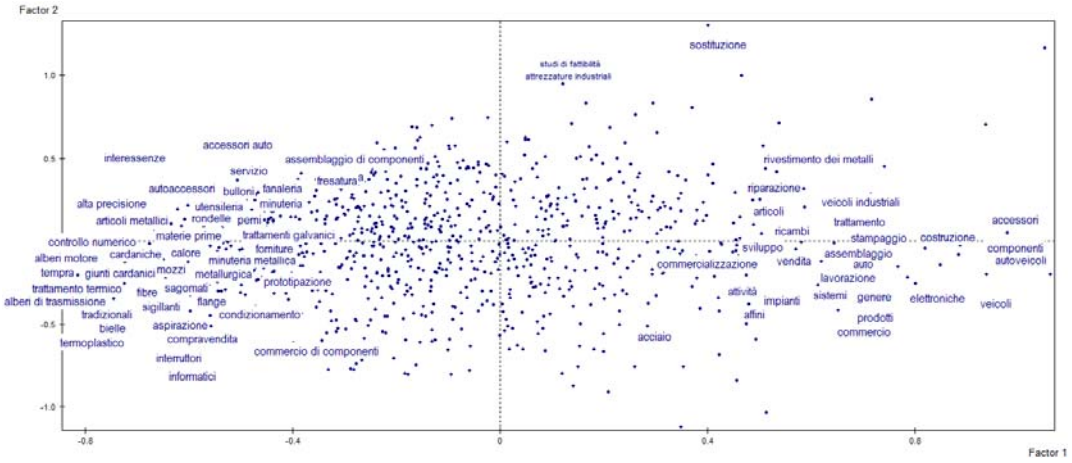


Figure 3 Distribution of the Vocabulary (lexicon) with a selection of characteristic terms (in Italian)



Which specializations characterize the automotive supply chain?

Multidimensional analysis and Cluster Analysis applied to the matrix *Documents x Keywords* (2,026 × 1,002) allow grouping of documents based on lexical content. This clustering phase on the single keyword base is an unattended classification that reflects the semantic level of similarities between the documents. This conceptual homogeneity expresses the semantic theme, or semantic trait, prevalent in that group of fragments, which can be summed up in a category, not defined before but obtained through this analysis[14]. As can be seen in Figure 4, there is a concentration around the centre of the bulk of documents, from which three tails depart. By highlighting the words of those tails, one can detect their different semantic domains: the upper-left quadrant of the factorial plan is characterized by companies specialized in metal coating and surface treatment; the top right quadrant is characterized by companies engaged in

[14] This operation was performed with the Spad software. Out of the 2,026 documents, only 1,733 documents were considered as active cases, consisting of at least 5 words. The remaining 293 documents, having less than 5 words, were considered as additional cases. In this way, the latter documents do not affect the formation of factor plans, but can be categorized into clusters within the ones obtained from the analysis of 1,733 active cases.

research and development and in engineering activities; while the lower right quadrant is characterized by companies specialized in metal profiling (forging and metallurgy).

Using a cluster analysis applied to a multidimensional analysis, with 10 factors, we obtain 20 clusters of documents. Figure 5 shows the distribution of documents in the 20 partitions and the number of documents for each of them.

Figure 4 Distribution of the Vocabulary by Documents

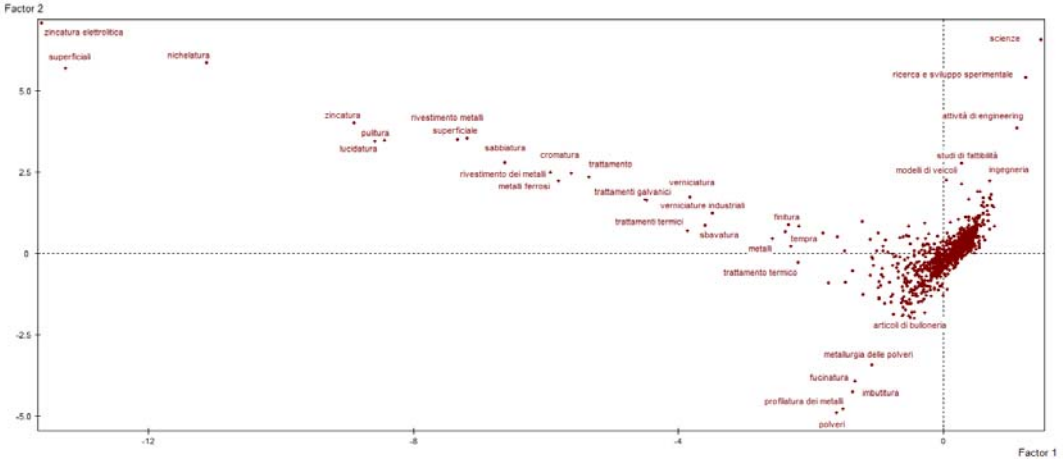
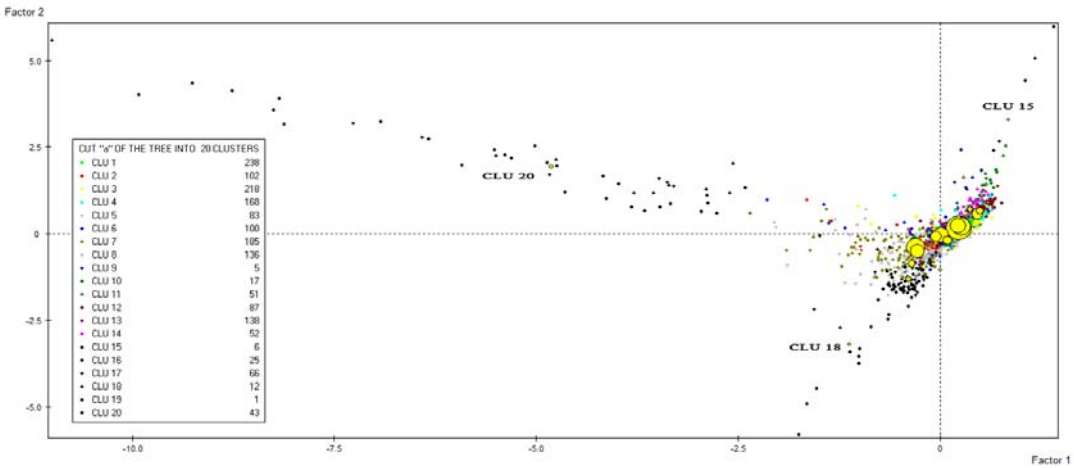


Figure 5 Distribution of documents by cluster, and cluster centroids



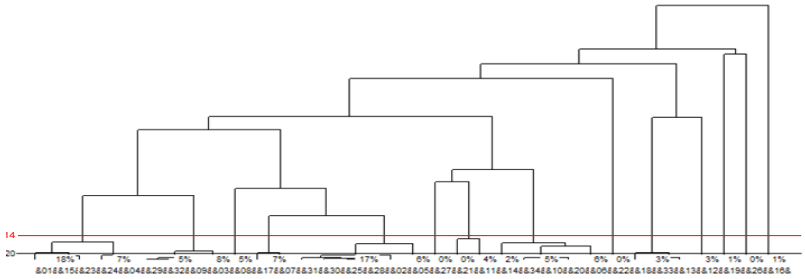
Legend: The 20 clusters are identified by different colours. Original IDs are listed in the table in the left side of the figure. The circles in yellow are the centroids of the clusters, (only centroids of clusters 15, 18 and 20 are marked with their id)

The vocabularies of the clusters, listed in Annex 4, are specifically used to label the groups of activities characterizing the companies in the automotive supply chain. Before interpreting that result, let us discuss the criteria to select the number of clusters. Figure 6 presents the dendrogram of the cluster analysis: it shows that 14 clusters may disentangle the main groups of specializations. We adopted a further split in 20 clusters that provided indications for more homogeneous bundles of specializations[15].

[15] The mixed method adopted in the cluster analysis of semantic data does not allow an unambiguous analysis of the agglomerative hierarchy: once a larger or a smaller number is set

Figure 6 Dendrogram (10 factors) cut-off at 20 a 14 clusters

Black horizontal line: cut-off 20 clusters; red horizontal line: cut-off at 14 clusters



The cluster analysis returns a list of the specific words of each cluster, forming each cluster dictionary. By browsing the list of the 20 dictionaries, we have ordered the clusters (and reassigned an id) according to six macro categories, listed in Table 6: Design; Parts, components, machining; Electrical equipment; Windows and car interiors; Spare parts; Miscellany.

Table 6 Number of companies and employees in 2015 of the 20 clusters of specialization

cluster id.	Aggregation	Cluters label	N. of companies	N. of employees
2+3-4	**Design**	Design, research and development	37	1002
5	**Parts, components, machining**	Parts for motor vehicles	205	15641
6		Transmission organs	104	10014
7		Machining,and surface treatments	213	21525
8		Molds and shapes	59	5639
9		Molding, and surface treatments	151	9377
10		Metal profiling	12	1134
11		Washer & hardware	27	1384
12		Metal hardware & taps	77	3820
13		Surface Treatments	86	1148
14	**Electrical equipment**	Electrical equipment, batteries; engi	106	22828
15		Electronics for refrigeration and hea	151	16966
16		Electrical electronic equipment and	59	5457
17+18	**Windows and**	windows and car interiors	126	13407
19	**Spare parts**	Vehicle spare parts	241	9966
20		Parts	266	40191
1	Miscellany	Miscellaneous business	106	9548
		Total	2026	189047

In implementing the results of the cluster analysis, some of the clusters have been aggregated. Three smaller clusters (id. 2, 3 and 4), made up of few documents, have been aggregated, in the category "design", grouping companies classified in "Industrial design", "Design studies" and "Research and experimentation". In addition, cluster 18 (glasses), with only one company, has been aggregated to cluster 17, having 125 companies with several specializations related to interior parts and glasses.

The category "miscellany", cluster 1, deserves our attention. The prevailing terms in those documents refer to generic business activities, even though a minor part of the description refers to specific processing or to production of parts and components. This is not a drawback of automatic classification. It can be

for the cut-off, the resulting clusters are reorganized to improve the similarity within the chosen number of clusters. On the better properties of mixed method in clustering of text data, see (Benzécri, 1992; Bolasco 2012; Bolasco, 2013).

explained by the nature of the corpus under analysis. If we had access to the full original description provided by the companies, the generic business activities would have been described in all the documents. In this way, they would not appear as specific terms. Since - in almost 5% of the documents – they were not omitted in the description verified by experts, they became the specific terms characterizing those companies. This consideration highlights the improvement of the results that could be obtained if full texts were available, a topic that we will discuss in Section 7.

With regard to the macro-categories, they encompass clusters of specializations that are relevant in the analysis of the supply chain. In particular, "Parts, components, machining" identifies, in distinct categories, companies producing systems and components that belongs to different technological domains: Parts for motor vehicles; Transmission organs: Machining and surface treatments; Molds and shapes; Molding, and surface treatments; Metal profiling; Washer & hardware; Metal hardware & taps; Surface Treatments. The latter overlaps in three clusters, and this could be explained by the fact that they could refer to different technologies of treatment (an issue that will be explored in Section 6).

With regard to the macro category "Electrical equipment", the three clusters appear technically separated: Electrical equipment, batteries, engines; Electronics for refrigeration and heating: Electrical, electronic equipment and software.

The macro category "Spare parts", the largest group of companies, encompasses Vehicle spare parts and Repair and spare parts, thus classifying two significant segments of companies operating in aftermarkets, as distinct from the producers of parts for the car makers.

Co-occurrence network analysis

Through the network analysis of co-occurrences of keywords in fragments of text, 33 modularity classes are identified, which in this case are semantic sub-networks. As can be seen in Figure 7, the main one of these sub-networks, with 232 related words, is linked to the word *produzione* (production). There are then a number of other major sub-networks, *fabbricazione, costruzione, lavorazioni in genere* (manufacturing, construction, machining in general), and over twenty further sub-networks with few related terms.

Figure 8 shows the co-occurrence graph where the word "production" is excluded. The main word of each sub-network is recalled in Table 7, in descending order of the number of nodes adjacent to the node that has the highest degree in the sub-network. The semantic network presents an effective means for a storytelling on the many specializations characterizing the automotive supply chain. Of no impact on a classification of companies, the graph highlights how the groups of specializations are linked across the description provided in the corpus, allowing a more detailed perspective than that offered by the semantic clustering of documents.

Figure 7 Co-occurrences graph

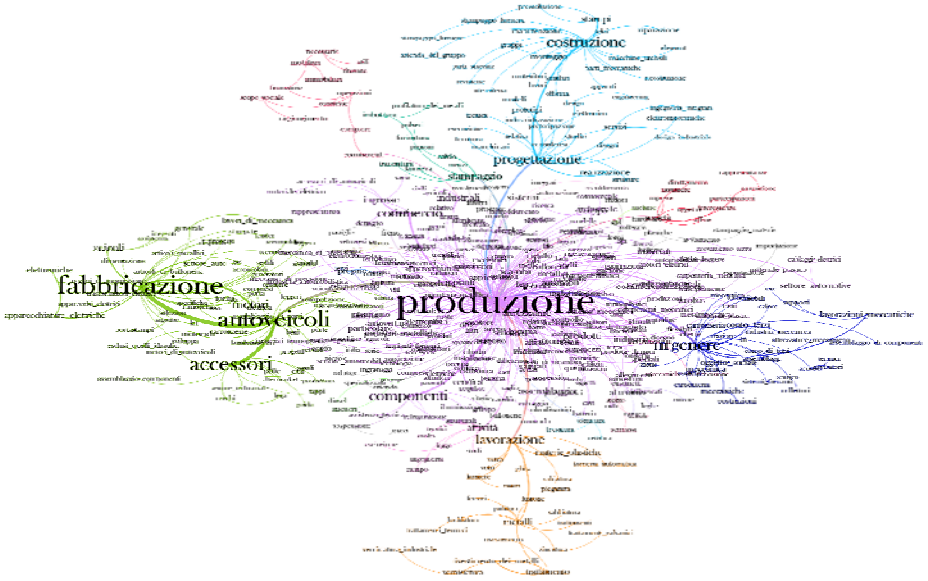
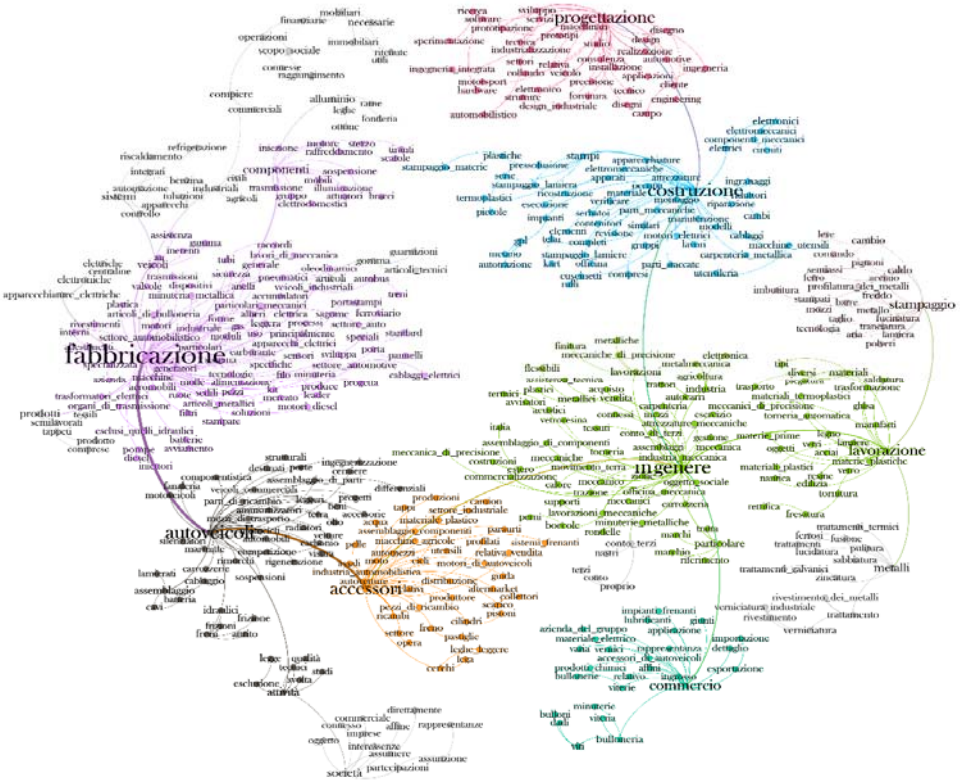


Table 7 Modularity classes: ID and main word of the class

	main word in the subnetwork		n. nodes		main word in the subnetwork		n. nodes
8	fabbricazione	*manufacture*	103	15	trattamenti superficiali	*surface treatments*	15
4	lavorazioni	*processing*	88	10	altre attività	*other activities*	12
11	costruzione attrezzature/ carpenteria	*construction equipment / carpentry*	56	6	sistemi	*systems*	9
16	autoveicoli -- parti	*cars - parts*	51	9	varie altre attività e …	*various other activities and …*	6
2	accessori	*accessories*	40	5	conto terzi/conto proprio	*Third party account / own account*	5
7	progettazione design sviluppo	*design design development*	38	12	fonderia e metalli	*foundry and metals*	5
13	commercio	*business*	26	0	apparecchiature elettriche	*electrical equipment*	4
1	stampaggio e lavorazione metalli	*molding and metal working*	23	14	gomma e guarnizioni	*Rrubber and gaskets*	4
3	altre attività	*other activities*	15				

Figure 8 Co-occurrence graph. Network analysis excluding the word "production"



5. Specializations within the automotive supply chain: Nace classification of semantic clusters and spatial categorization

At what level of detail in the Nace classification can the automotive supply chain be described? As matter of fact, in the Register of companies, description of company's activities and Nace classification are coherent information. But if we summarize the Nace codes for the 20 clusters, it is hardly to read a shortcut classification based on textual descriptions. Limiting the details at the 3-digit level, the matrix *Nace [at 3 digits] x 20 clusters* appears quite sparse, see Annex 4. To disentangle which Nace codes are characteristics of the 20 clusters, we can use the categorical classification elaborated from the cluster analysis. Annex 5 lists the main categorical variables (regions, provinces, class of employees and Nace) characterizing each cluster. With regard to Nace categories (at 2, 3 and 4 digits), Table 8 summarizes the clusters' classification. From this table we can clearly associate the groups of companies in the 20 clusters to a set of 2, 3 and 4 digits of the Nace classification: an aggregation created by the clustering and not defined ex ante.

Except for the macro category Design and its three clusters, with no specific clustering categorization in terms of Nace codes, we find an overall confirmation that cluster labels are coherent with the Nace codes (a result confirming that original descriptions provided by the Registry were coherent with Nace classification). The most significant outcome of our semantic analysis to single out companies' specializations is that we now know which aggregation of Nace codes characterizes the companies in the supply chain.

Table 8 Clusters by characterisation of Nace rev. 2

N. of companies	N. of employees			m.of rubber products	m.of plastics products	m.of basic metals	Forging, pressing, stampingand roll-fo	Treatment and coating of metals	Machining	m.of cutlery, tools and general hardw	m.of tools	m.of other fabricated metal products	m.of fasteners and screw machine pro	m.of electrical equipment	m.of machineryand equipment n.e.c.	m.of general – purpose machinery	m.of bearings, gears, gearing and driv	m.of motor vehicles, trailers and semi-	m.of parts and accessories for motor	m.of electrical and electronic equipme	m.of other parts and accessor ies for	Wholesale and retail trade and repair
				221	222	24	255	2561	###	257	##	259	##	27	28	281	##	29	##	##	##	45
8	190	**Design**	Design studies																			
22	683		Industrial design																			
7	129		Research and experimentation																			
205	15641	**Electrical equipment**	Electrical electronic equipment and sw																			
104	10014		Electrical equipment, batteries; engines																			
213	21525		Electronics for refrigeration and heating																			
59	5639	**Parts, components, machining**	Machining,and surface treatments																			
151	9377		Metal hardware & taps																			
12	1134		Metal profiling																			
27	1384		Molding, and surface treatments																			
77	3820		Molds and shapes																			
86	1148		Parts for motor vehicles																			
106	22828		Surface Treatments																			
151	16966		Transmission organs																			
59	5457		Washer & hardware																			
123	10995	**Spare parts**	Parts																			
2	2412		Vehicle spare parts																			
241	9966	**Windows and car interiors**	glasses																			
266	40191		windows and car interiors																			
106	9548	**Miscellany**	Miscellaneous business																			
2025	189047																					

In fact, from the list in Table 9 we observe cases in which 2 digits are specific of clusters, as for Manufacturing of basic metals (24); of electrical equipment (27) and in the case of Wholesale and retail trade and repair of motor vehicles and motorcycles (45). In other clusters, a 3 digit classification is present, as in the cases of Manufacturing of rubber products (221) and of plastics products (222); lastly, 4 digits are a specific characteristic of some clusters, as for Treatment and coating of metals (2561) and Machining (2562). It would not have been possible to single out such a selected list by simply aggregating data hierarchically at 4, 3 or 2 digits of the Nace classification.

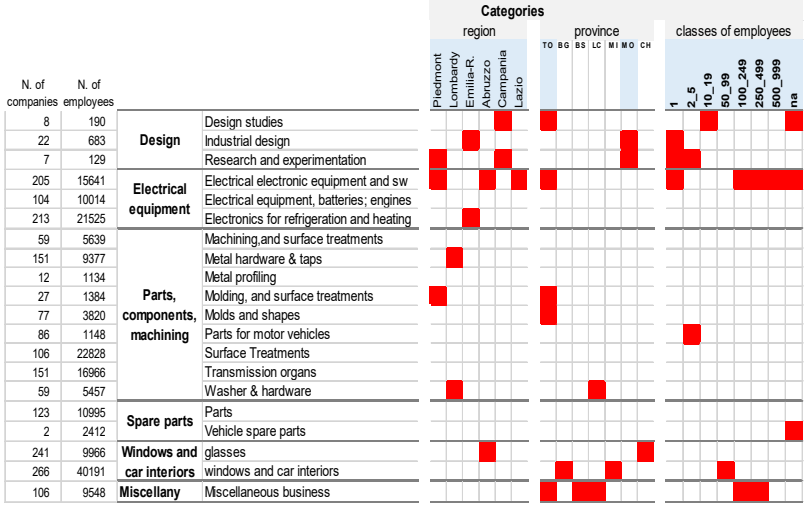
Table 9 List of 2, 3 and 4 digits Nace rev. 2 codes characterising the 20 clusters

	221	Manufacturing of rubber products
	222	Manufacturing of plastics products
24		Manufacturing of basic metals
	255	Forging, pressing, stamping and roll-forming of metal; powder metallurgy
		2561 Treatment and coating of metals
		2562 Machining
	257	Manufacturing of cutlery, tools and general hardware
		2573 Manufacturing of tools
	259	Manufacturing of other fabricated metal products
		2594 Manufacturing of fasteners and screw machine products
27		Manufacturing of electrical equipment
28		Manufacturing of machinery and equipment n.e.c.
	281	Manufacturing of general — purpose machinery
		2815 Manufacturing of bearings, gears, gearing and driving elements
29		Manufacturing of motor vehicles, trailers and semi-trailers
	293	Manufacturing of parts and accessories for motor vehicles
		2931 Manufacturing of electrical and electronic equipment for motor vehicles
		2932 Manufacturing of other parts and accessories for motor vehicles
45		Wholesale and retail trade and repair of motor vehicles and motorcycles

Although this is satisfactory, a more complete textual description of business activities could provide more targeted guidance on which Nace aggregations to adopt in the analysis.

As observed for the Nace classification, also spatial analysis benefits from the semantic clustering. Table 10 summarizes the results of the characterization of the 20 clusters in terms of location: regions and provinces.

Table 10 Clusters by characterisation of location: region and province



N. of companies	N. of employees		
8	190	**Design**	Design studies
22	683		Industrial design
7	129		Research and experimentation
205	15641	**Electrical equipment**	Electrical electronic equipment and sw
104	10014		Electrical equipment, batteries; engines
213	21525		Electronics for refrigeration and heating
59	5639	**Parts, components, machining**	Machining,and surface treatments
151	9377		Metal hardware & taps
12	1134		Metal profiling
27	1384		Molding, and surface treatments
77	3820		Molds and shapes
86	1148		Parts for motor vehicles
106	22828		Surface Treatments
151	16966		Transmission organs
59	5457		Washer & hardware
123	10995	**Spare parts**	Parts
2	2412		Vehicle spare parts
241	9966	**Windows and car interiors**	glasses
266	40191		windows and car interiors
106	9548	**Miscellany**	Miscellaneous business

6. Implementing clustering results

In order to discuss the implications of semantic clustering proposed in this paper, let us discuss some applications in two cases of use of the sampling: one instrumental to increase the robustness of the empirical survey, the other one to better identify "snakes and spiders" in the automotive supply chain.

Firstly, let us implement the results of the cluster analysis in comparing the distribution of companies in the automotive supply chain, in the population and in the respondent companies of the survey conducted by the Observatory in Spring 2017[16]. Table 11 presents data by cluster and region. As discussed in Section 4, there is a similarity of activities in three regions, Piedmont, Lombardy and Emilia-Romagna, in which almost 72% of the companies in the automotive supply chain are located. Almost 79% of the respondent companies are located in the three regions.

Table 11 Companies by cluster of specialization and region

Population and respondent companies from the Italian Automotive Observatory, survey 2017

			Companies in the supply chain									
			POPULATION					RESPONDENT Companies				
cluster id.	Aggregation	Cluters label	Piedmont	Lombardy	Emilia-Romagna	other regions	Total	Piedmont	Lombardy	Emilia-Romagna	other regions	Total
2+3-4	**Design**	Design, research and development	22	2	5	8	37	7		1	1	9
5	**Parts, components, machining**	Parts for motor vehicles	66	39	28	72	205	13	9	8	11	41
6		Transmission organs	39	33	15	17	104	11	3	2	4	20
7		Machining,and surface treatments	87	70	14	42	213	25	12	5	7	49
8		Molds and shapes	34	7	3	15	59	8		2	2	12
9		Molding, and surface treatments	85	26	7	33	151	18	5	1	5	29
10		Metal profiling	7	2		3	12	2	1		1	4
11		Washer & hardware	9	13	3	2	27	3	3	2	0	8
12		Metal hardware & taps	15	55	2	5	77	6	3	1	1	11
13		Surface Treatments	25	34	1	26	86	6	2		2	10
14	**Electrical equipment**	Electrical equipment, batteries; engines	23	26	13	44	106	3	1	4	3	11
15		Electronics for refrigeration and heating	51	28	28	44	151	15	8	4	6	33
16		Electrical electronic equipment and sw	24	11	4	20	59	5	2	1	4	12
17+18	**Windows and c**	windows and car interiors	52	32	10	31	125	13	6	1	8	28
19	**Spare parts**	Vehicle spare parts	59	49	37	96	241	17	5	8	11	41
20		Parts	77	72	32	85	266	25	13	6	13	57
1	Miscellany	Miscellaneous business & other	52	28	8	23	111	14	7	1	6	28
			727	**527**	**210**	**566**	**2030**	**191**	**80**	**47**	**85**	**403**

Using figures in Table 11 we applied the chi-square goodness-of-fit test to ascertain the consistency of the sampling distribution of each regional stratum and total over the specialization categories, with that of the corresponding population from which the sample of respondents was drawn, even if the various sampling sizes were low to have robust results. The observed level of significance (p-values) obtained from the observed values of the statistical test were higher than the ordinary 0.05 level of significance, involving the acceptance of the null hypothesis concerning the consistency of the sampling distributions with their corresponding population distributions, i.e. the respondent companies are distributed by specialization and region as is the population. Moreover, the proportion of sampling units in Piedmont (47.4%) was higher than that of the

[16] Annex 7 shows data on the number of companies and employees for the whole statistical population and for the respondent companies, by Nace 2 digit.

Piedmont population (35.8%), both with respect to the total, and this difference was statistically significant at the 0.001 level of significance (Z-test with p-value=0.0008). Since we have very small figures by stratum, no significance test may be applied.

Clustering could be adopted as a tool for statistical sampling, by stratifying the population according to the relevant specialization, the size of the company and the location. Although the overall sample of respondent companies is significant, clustering of specializations could be adopted during the interviewing phase as a tool for controlling the representativeness of the sample in all the strata and checking the robustness of the results, focusing recalling of companies to increase representativeness of all strata.

A second case of use is closely related to sampling. In particular, to enhance the technical and relational classification proposed by Anfia (Annex 1) as a significant step in the "snakes and spiders" analysis of the automotive supply chain, a statistical sampling would allow more robust considerations with regard to the various specializations in the chain. By assuming that the classification that emerged from semantic cluster analysis is robust, the respondent companies provide appropriate information on the "snake and spider" structure of the supply chain, or more specifically on its relational structure: an original result that we can derive from the semantic analysis, summarised in Table 12. We are now able to identify the different types of relational settings (from SIST/MOD to sub-contractors) characterizing the various specializations in the automotive supply chain, and we could also discuss them with regard to their special location.

Table 12 Classification of companies by specialization and relational position in the supply chain

cluster id.	Aggregatic	Cluters label	NON respondent	Respondent companies												Total	
				E&D		SIST/MOD		SPEC		SPEC Aftermarket		SPEC Motorsport		SUB supplier			
			n.	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%
2+3-4	**Design**	Design, research and development	28	6	19,4			1	0,8					2	1,4	9	2,2
5	**Parts,**	Parts for motor vehicles	164	6	19,4	4	10,8	11	9,1	9	15,3	1	10,0	10	6,9	41	10,2
6		Transmission organs	84					9	7,4	5	8,5			6	4,1	20	5,0
7		Machining,and surface treatments	164			4	10,8	15	12,4	2	3,4	1	10,0	27	18,6	49	12,2
8		Molds and shapes	47	8	25,8	1	2,7			1	1,7	1	10,0	1	0,7	12	3,0
9		Molding, and surface treatments	122					10	8,3					19	13,1	29	7,2
10		Metal profiling	8			1	2,7	2	1,7					1	0,7	4	1,0
11		Washer & hardware	19					2	1,7	3	5,1			3	2,1	8	2,0
12		Metal hardware & taps	66					4	3,3					7	4,8	11	2,7
13		Surface Treatments	76											10	6,9	10	2,5
14	**Electrical**	Electrical equipment, batteries; engines	95	1	3,2			3	2,5	2	3,4	1	10,0	4	2,8	11	2,7
15		Electronics for refrigeration and heating	118			5	13,5	12	9,9	3	5,1	1	10,0	12	8,3	33	8,2
16		Electrical electronic equipment and sw	47	2	6,5	1	2,7	3	2,5	1	1,7	2	20,0	3	2,1	12	3,0
17+18	**Windows**	windows and car interiors	96					18	14,9					11	7,6	29	7,2
19	**Spare**	Vehicle spare parts	200	3	9,7	4	10,8	11	9,1	12	20,3	2	20,0	9	6,2	41	10,2
20	**parts**	Parts	209	3	9,7	9	24,3	13	10,7	19	32,2	1	10,0	12	8,3	57	14,1
1	Miscellany	Miscellaneous business	83	2	6,5	8	21,6	7	5,8	2	3,4		0,0	8	5,5	27	6,7
		Total	1626	**31**	100,0	**37**	100,0	**121**	100,0	**59**	100,0	**10**	100,0	**145**	100,0	**403**	100,0

7. Lessons from text analysis and further research

Qualitative and quantitative automatic text analysis allow us to elaborate information for a classification of business specializations, useful to identify which aggregation of Nace codes characterize the companies in the automotive supply chain and which specializations and locations characterize its internal structure. Albeit not complete, information gathered from the Observatory 2017 of the automotive supply chain allows us to draw some conclusions about the

categorization emerging from semantic cluster analysis and to outline this categorization as a tool to improve the significance of the empirical survey, e.g. by focusing recalls on the strata requiring more instances.

The empirical analysis compares the classification that emerged from text analysis with Nace rev.2 classification at 2 and 3 digits, highlighting a significant clustering that otherwise would have been difficult to single out just by aggregating groups of digits. Moreover, the classification is implemented by comparing the population of companies in the automotive supply chain and the respondents to the on line survey, providing some indications on the quality of the survey, beyond the information acquired by analysing size and location of the companies.

In particular, the semantic classification of specialization coupled with the relational position of the companies in the automotive supply chain (information gathered by the empirical survey) is an original contribution provided by the proposed analysis, paving the way for a comparative analysis over the various editions of the Observatory and henceforth making possible the analysis of the dynamics of change in the specializations and relational structure over the past decades.

Further research will implement the methodology to the full archive of the Observatory of the Italian automotive supply chain, and possibly to the companion archive of companies created by CNR-Ceris to analyse the performance of companies in this supply chain. Another step will be explored by taking advantage of the open access to the Register of French companies, recently made available on active companies: this database will be explored to validate the methodology in identifying both the perimeter of the supply chain and its internal structure of specializations.

References

Baldwin, R., Venables, A.J., 2013. Spiders and snakes: Offshoring and agglomeration in the global economy. J. Int. Econ. 90, 245–254. doi:http://dx.doi.org/10.1016/j.jinteco.2013.02.005

Bastian, M., Heymann, S., Jacomy, M., others, 2009. Gephi: an open source software for exploring and manipulating networks. Icwsm 8, 361–362.

Benzécri, J.-P., 1992. Correspondence analysis handbook, Statistics, textbooks and monographs. Marcel Dekker, New York.

Bolasco, S., 2013. L'analisi automatica dei testi: fare ricerca con il text mining. Carocci, Roma.

Bolasco, S., 2010. TaLTaC 2.10: sviluppi, esperienze ed elementi essenziali di analisi automatica dei testi. LED, Milano.

Bolasco, S., Pavone, P., 2007. Automatic dictionary and rule-based systems for extracting information from text, in: Automatic Dictionary and Rule-Based Systems for Extracting Information from Text. p. 255.258.

Bonifati, G., 2013. Exaptation and emerging degeneracy in innovation processes. Econ. Innov. New Technol. 22, 1–21. doi:10.1080/10438599.2012.689674

De Backer, K., Miroudot, S., 2012. Mapping global value chains, in: Paper Prepared for the Final WIOD Conference: Presented at the Causes and Consequences of Globalization, Groningen, The Netherlands, April 24-26, 2012., Groningen.
Fally, T., 2011. On the Fragmentation of Production in the US. Work. Pap. Univ. Colo.-Boulder July.
Flament, C., 1981. L'Analyse de Similitude, une Technique pour les Recherches sur les Représentations Sociales. Cah. Psychol. Cogn. 375–395.
Flament, C., 1962. L'analyse de similitude. Cah. Cent. Rech. Opérationnelle 63–97.
Fortunato, S., Hric, D., 2016. Community detection in networks: A user guide. Physics Reports 659, 1–44. doi:10.1016/j.physrep.2016.09.002
Gereffi, G., Fernandez-Stark, K., 2016. Global value chain analysis: a primer. Duke CGGC.
Helper, S., Sako, M., 2010. Management innovation in supply chain: appreciating Chandler in the twenty-first century. Ind. Corp. Change 19, 399–429. doi:10.1093/icc/dtq012
Lane, D.A., 2010. Innovazione e distretti industriali, in: Russo, M. (Ed.), Processi Di Innovazione e Sviluppo Locale: Teorie e Politiche. pp. 57–68.
Marchand, P., Ratinaud, P., 2012. L'analyse de similitude appliquée aux corpus textuels: les primaires socialistes pour l'élection présidentielle française (septembre-octobre 2011). Actes 11eme Journ. Int. D'Analyse Stat. Données Textuelles JADT 2012, 687–699.
Memedovic, O., 2007. Presentation Prepared for the UNCTAD Expert Group Meeting on Enhancing the Participation of Small and Medium-Sized Enterprises into Global Value Chins, Held on 18–19 October 2007, in Geneva.
OECD, 2017. OECD Skills Outlook 2017. Organisation for Economic Co-operation and Development, Paris.
Pavone, P., 2010. Sintagmazione del testo: una scelta per disambiguare la terminologia e ridurre le variabili di un'analisi del contenuto di un corpus, in: Bolasco, S., Chiari, Isabella, Giuliano, Luca (Eds.), Statistical Analysis of Textual Data. Proceedings of 10th International Conference Journées d'Analyse Statistique Des Données Textuelles 9-11 June 2010 - Sapienza University of Rome. Led on Line, Roma, pp. 131–140.
Prometeia, UnionCamere, 2015. Il settore automotive nei principali paesi europei. Ricerca promossa dalla 10a Commissione Industria, Commercio, Turismo del Senato della Repubblica Roma, 28 luglio 2015.
Rinaldi, A., 2008. I distretti industriali italiani a specializzazione metalmeccanica dalle origini agli anni Novanta, in: Russo, M. (Ed.), L'industria Meccanica in Italia 1951-2001. Carocci-Press online, Roma, pp. 153–199.
Rosenberg, N., 1963. Technological Change in the Machine Tool Industry, 1840–1910. J. Econ. Hist. 23, 414–443. doi:10.1017/S0022050700109155
Russo, M., 2015. Distretti, piccole imprese e sapere diffuso nei sistemi produttivi della meccanica, in: Sciolla, L., Salvati, M. (Eds.), L' Italia e le sue regioni: l'età repubblicana. V. II: Territori. Istituto della Enciclopedia Italiana, Roma, pp. 147–160.
Russo, M., 2008. L'industria meccanica in Italia: analisi spaziale delle specializzazioni produttive 1951-2001. Carocci, Roma.
Sako, M., 2005. Governing automotive supplier parks: leveraging the benefits of outsourcing and colocation, in: DRUID Tenth Anniversary Summer Conference. pp. 27–29.
Sturgeon, T., Van Biesebroeck, J., Gereffi, G., 2008. Value chains, networks and clusters: reframing the global automotive industry. J. Econ. Geogr. 8, 297–321. doi:10.1093/jeg/lbn007
Sturgeon, T.J., Memedovic, O., Biesebroeck, J.V., Gereffi, G., 2009. Globalisation of the automotive industry: main features and trends. Int. J. Technol. Learn. Innov. Dev. 2, 7. doi:10.1504/IJTLID.2009.021954

Annex 1 Classification of companies specialized in production of intermediate goods and in engineering & design

SIST/MOD	Suppliers of complex systems and/or entire vehicle parts (eg power supply system, braking system, door module, dashboard). This includes some companies that do not have a production unit in Italy but sell their products to car manufacturers.
SPEC	Suppliers of specific components (eg engine starter battery, brake pads, brake disks, shift levers).
SPEC Strictly defined	Company that sells products in the OEM (Original Equipment Manufacturer) or in the spare parts market (Aftermarket). Clients can be car makers or other companies producing components.
SPEC Aftermarket	Companies that sell their products (lubricants, antifreeze liquids, snow chains, interior cleaning sprays, scratch-resistant wipers for bodywork, internal divisions, tool holders, fire extinguishers, etc.) essentially in the spare parts industry (Aftermarket). These are companies that have no relationship with car makers and sell their products to independent distribution operators (Regional Distributors, Purchasing Groups, Retailers).
SPEC Motorsport	Companies that produce high performance products almost exclusively for the Motorsport and Racing sectors (eg alloy wheels, sport seats, sports steering wheels, special seat belts, roll bars, suits, shoes and boots).
Telematics	Companies that produce products (hardware, eg black boxes) or software (satellite location, driving mode, etc.) for connections in the mobility/vehicle sector.
Processing	Companies that work on certain products (eg thermal treatments, painting, lamination, cutting, welding, sandblasting, etc.). Molding activities falling within the SPEC category are not considered under processing.
Subcontractors	Second/third level companies in the supply pyramid. They are companies that supply some components to SIST/MOD or SPEC; they do not usually have direct delivery reports with the car makers. They also carry out work on behalf of third parties. The certification is ISO 9001. Subcontractors are companies that produce non-high-performance fastening systems (eg screws, nuts, bolts, washers, hinges, etc.). Manufacturers of special fastening systems are to be classified as SPECs
Engineering & Design:	Companies engaged in the design, engineering, design/style phases of the realization of a vehicle. This category also includes companies that carry out design/engineering activities on some vehicle systems (eg powertrain, interior, etc.)

Source: ANFIA, Information on the Osservatorio componentistica automotive italiana 2017

Annex 2 Examples of types of text

ID Doc 2928 – Occorrenze Assolute 4
VARIABILI PRE-CATEGORIZZATE
PROVINCIA: TO
REGIONE: Piemonte
ATECO 4 CIFRE: 6201
DIP2015: 2
CLDIP: 2-5
[DescrAtt]
realizzazione di software
[AttVer]
infomobilita'

ID Doc 2825 – Occorrenze Assolute 271
VARIABILI PRE-CATEGORIZZATE
PROVINCIA: MI
REGIONE: Lombardia
ATECO 4 CIFRE: 2561
DIP2015: 41
CLDIP: 20-49
[DescrAtt]
la società ha per oggetto l' incisione di ogni tipo di stampi metallici e placche con sistema fotolitico od altro sistema, la cromatura, ramatura ed altre forme di deposito elettrolitico incluse la relativa apparecchiatura per tali procedimenti e la manifattura ed incisione con ogni metodo, di cilindri e placche per l' impressione di ogni tipo di materiale atto a ricevere un finissaggio decorativo come carta, plastica, gomma, metalli, fibre, vetro, legno, di cilindri e placche per la stampa di ogni tipo di materiale atto a ricevere una stampa di finitrua estetica come sopra, di cilindri applicatori di ogni tipo di liquido, come inchiostri adesivi, lacche, cere. nonché la manifattura di macchine per incisione fotolitica, montatura, messa a punto e prova di tutti i tipi di stampa a rilievo e flessografica, incisione dei rulli, placche, cilindri da incisione, incisione di un finissaggio decorativo su plastica e altri materiali. la società ha inoltre per oggetto la manifattura di abiti protettivi per sports acquatici, di rilievi flessibili e rulli di rulli di gomma o sintetici per stampa e incisioni su ogni materiale ed ogni operazione di meccanica e tecnica necessaria per l' esecuzione di tutti i procedimenti sopraelencati, nonché l' importazione, l' esportazione, il commercio, la riparazione, la manutenzione, l' applicazione di pompe centrifughe, motori elettrici, molini, dissipatori, ed altre macchine o congegni. essa potrà compiere tutte le operazioni commerciali, industriali, finanziarie, mobiliari e immobiliari, che saranno ritenute dal Consiglio d'amministrazione necessarie od utili per il conseguimento dell' oggetto sociale. potrà anche assumere, sia direttamente sia indirettamente, interessenze o partecipazioni in altre società, od imprese aventi oggetto analogo od affine o connesso.
[AttVer]
trattamento e rivestimento dei metalli

ID Doc 2627 – Occorrenze Assolute 32
VARIABILI PRE-CATEGORIZZATE DA FONTE ESTERNA
PROVINCIA: BA
REGIONE: Puglia
ATECO 4 CIFRE: 2211
DIP2015: 729
CLDIP: 500-999
[DescrAtt]
produzione di copertoni, pneumatici e parti accessorie, nonché di materiali di riparazione e ricostruzione degli stessi.
[AttVer]
produzione di copertoni, pneumatici e parti accessorie, nonché di materiali di riparazione e ricostruzione degli stessi.
.

Annex 3 Categorical variables and most characteristic terms of factors 1 and 2, by active frequency

file 07 Tab ACS f1f2.xlsx

Variable label	Coordinate	Weight		Variable label	Coordinate	Weight
Lombardia	-0,32	851,000		Molise	-1,31	13,000
Piemonte	-0,31	931,000		1000_più	-0,83	131,000
20_49	-0,30	835,000		Umbria	-0,83	103,000
50_99	-0,28	705,000		Liguria	-0,73	93,000
10_19	-0,26	744,000		Friuli Venezia Giulia	-0,60	164,000
Emilia-Romagna	-0,18	628,000		250_499	-0,56	399,000
100_249	-0,17	639,000		Trentino-Alto Adige	-0,40	120,000
MIDDLE AREA				*MIDDLE AREA*		
Sardegna	0,85	18,000		Marche	0,25	136,000
Basilicata	0,93	107,000		Campania	0,33	348,000
Trentino-Alto Adige	0,95	120,000		ND	0,40	567,000
Sicilia	1,07	78,000		Sicilia	0,41	78,000
Calabria	1,46	22,000		0_1	0,57	312,000
Molise	1,53	13,000		Lazio	0,62	208,000
Valle Aosta	1,61	36,000		Calabria	0,71	22,000

Annex 4 Clusters: specific terms (in Italian) and number of documents

Labels of clusters are translated in Table 6. Number of documents refer to the active cases (more than 5 occurrences of keywords) used for clustering, plus the additional cases, assigned to clusters accordingly (see footnote 15). Words in bold are highlighted to help in defining the label of the cluster

MISCELLANY Cluster 1 ex2 – 102+4 documents
acquisto, macchinari, attività, **rappresentanze**, esercizio, assunzione, oggetto sociale, **permuta**, immobili, estero, **vendita**, zione, qualsiasi, locazione, **brevetti**, tecnici, società, impianti industriali, **commissionaria**, industriale, minuto, gruppo, attrezzature, **utilizzazione**, **concessione**, terzi, **consulenza**, impianti, carrozzeria, prodotti industriali, dati, tecnico, esclusione, connessi, beni, commerciale, commercializzazione, opere, beni immobili, estere, costruzioni, materiali plastici.

DESIGN INDUSTRIALE Cluster 2 ex 9 – 5+3 documents
design industriale, studi di fattibilità, attività di design di moda, modelli di veicoli, servizi di consulenza, engineering, nuovi, settore auto, carrozzerie.

PROGETTAZIONE & DESIGN Cluster 3 ex10 – 17+5 documents
studi, attività di engineering, tecnica, design, attività, software, analisi, simulazione, ingegneria integrata, studio, progettazione, prototipi, sperimentazione, servizi, prototipazione, ingegneria.

RICERCA E SPERIMENTAZIONE Cluster 4 ex15 – 6+1 documents
ingegneria, ricerca e sviluppo sperimentale, scienze, poliuretano, sperimentazione, ricerca, sostituzione, campo.

*PARTI PER AUTOVEICOLI Cluster 5 ex4 – **168+37** documents*
parti, autoveicoli, **accessori**, fabbricazione, motori, rimorchi, carrozzerie, cerchi, **sedili**, **ruote**, **prototipi di autoveicoli**, **assemblaggio componenti**, lega, ufficio, navi, motocicli, auto, **frizioni**.

ORGANI DI TRASMISSIONE Cluster 6 ex5 – 83+21 documents
alberi, sterzo, ingranaggi, cambio, bracci, semiassi, giunti cardanici, scatole, sospensione, alberi di trasmissione, volani, leve, supporti, mozzi, mozzi ruota, alberi motore, bielle, pignoni, differenziali, lavori di meccanica, trasmissioni, organi di trasmissione, flange, esclusi quelli idraulici, aeromobili, freno, scatole cambio, riduttori, giunti, perni, velocità, barre, generale, crociere, coppie coniche, caldo, assali, boccole, lavorazioni meccaniche, tiranti, cambi di velocità, collettori, motore, motoriduttori, ruota, comandi, cardaniche, trasmissione, alta precisione, comando, pulegge, officina meccanica, guaine, flessibili, cambi, motocicli, pompe.

LAVORAZIONI, ma anche trattamenti superficiali Cluster 7 ex7 – 185+28 documents
acciaio, alluminio, fusione, molle, lavorazioni meccaniche, getti, fonderia, anelli, minuterie metalliche, precisione, metalli, lavorazione, leghe, tornitura, lavorazioni, metallici, ghisa, metallo, caldo, torneria, fresatura, fili, ottone, conto terzi, ferrosi, raccordi, rulli, trattamento termico, genere, particolari, lavori di meccanica, profilati, semilavorati, sfere, trattamenti superficiali, pressofusi, azienda, officina meccanica, meccanica di precisione, ferro, prodotti siderurgici, pneumatiche, stampati, pressofusione, metalliche, cuscinetti, meccanica, finiti, tubi flessibili, metallurgica, impieghi, filo, tubi, spirale, complessivi, metallico, acciai, particolari meccanici, rame, meccanici di precisione, prodotti.

STAMPI E FORME Cluster 8 ex11 – 51+8 documents
forme, **portastampi**, **sagome**, **stampi**, **modelli**, **prototipi**, servizi, macchine, engineering, progettazione, ingegneria integrata, design industriale, stile, progetti, veicolo, disegni, piccole, resina, tecnici, soluzioni, virtuale, sviluppo, **legno**, cliente, polistirolo, settore auto, completa, campi, design.

STAMPAGGIO, ma anche trattamenti superficiali Cluster 9 ex8 – 136+15 documents
lamiere, stampaggio, stampi, lamiera, lavorazione, costruzione, freddo, stampaggio lamiera, materie plastiche, tranciatura, grandi, plastica, plastiche, stampaggio materie, particolari, articoli tecnici, termoplastici, attrezzature, caldo, plastici, metalli, alimentare, ferro, gomma, piccole, stampaggio metalli, materie, medie, carpenteria, trattamenti termici, progettazione, termoindurenti, dimensioni, serie, settori, saldatura, acciaio, lavori di carpenteria metallica, metallo, iniezione, articoli, attrezzature meccaniche, stampaggio lamiere, laminati, deformazione, stampaggio di materie plastiche, termoplastiche, verniciature, manufatti.

PROFILATURA METALLI Cluster10 ex18 – 12+1 documents
profilatura dei metalli, imbutitura, fucinatura, meta, polveri, stampaggio, metallurgia delle polveri, taglio.

RONDELLE & FERRAMENTA Cluster 11 ex16 – 25+2 documents
mobiliari, finanziarie, immobiliari, partecipazioni, utili, operazioni, necessarie, ritenute, interessenze, aventi oggetto, analogo, imprese, commerciali, conseguimento, società, scopo sociale, affine, raggiungimento, connesso, oggetto sociale, garanzie, operazione, prevalente, immobiliare, finanziaria, terzi, legge, interesse, **rondelle**, quelle, favore, assunzione, industriali, connessa, commerciale, esclusione, compresa, **ferramenta**, **motorizzazione**, esercizio, **dadi**, connesse.

MINUTERIA METALLICA & RUBINETTERIA Cluster 12 ex17 – 66+11 documents
minuteria metallica, bulloneria, viteria, viti, viterie, dadi, bulloni, articoli di bulloneria, minuterie, bullonerie, tornite, freddo, speciali, speciale, minuterie metalliche, torniti, stampata, rivetti, standard, articoli metallici, stampate, filo, produzione, boccole, filettatura, commercio, cliente, tornita, metallica, minuteria, elementi di fissaggio, disegno, affini, perni, nastro, rubinetteria, officine meccaniche, ottone, mobile, estrusione, prodotti derivati, acciaio inox.

TRATTAMENTI SUPERFICIALI Cluster 13 ex20 – 43+43 documents
trattamento, rivestimento dei metalli, metalli, sabbiatura, zincatura, pulitura, lucidatura, trattamenti, superficiale, metalli ferrosi, nichelatura, finitura, sbavatura, cromatura, lavorazioni, trattamenti galvanici, zincatura elettrolitica, trattamenti termici, conto terzi, verniciature industriali, verniciature, rivestimento, verniciatura industriale, opera, ferrosi, verniciatura.

APPARECCHIATURE ELETTRICHE, BATTERIE; MOTORI Cluster 14 ex 12 – 87+19 documents
apparecchiature elettriche, elettroniche, motori, veicoli, trasformatori elettrici, fabbricazione, generatori, batterie, avviamento, avvisatori, cablaggi elettrici, accumulatori elettrici, acustici, fornitura, accumulatori, termostati, apparecchiature elettroniche, elettrici, cablaggi, batterie di pile, corrente, cablaggio, componenti elettronici, elettronici, integrati, industrializzazione, dispositivi, centraline, motori elettrici, schede.

ELETTRONICA PER REFRIGERAZIONE E RISCALDAMENTO Cluster 15 ex13 – 138+14 documents
elettronici, elettromeccanici, sistemi, elettromeccaniche, elettrici, apparecchi, riparazione, apparecchiature, impianti, elettriche, manutenzione, condizionatori, sensori, installazione, refrigerazione, cablaggi, meccanici, indicatori, autotrazione, relativi, riscaldamento, commercio di componenti, interruttori, oleodinamiche, processi, industrializzazione, climatizzazione, luci, conversione, cavi, componenti, componentistica, costruzione, gpl, controllo, moduli, pressione, elettronica, scambiatori, assiali, tappi, carburante, oleodinamici, completi, attrezzature, attuatori, temperatura, strumenti, qualità, componenti meccanici, lampade, applicazioni.

APPARECHHIATURE ELETTRICHE ELETTRONICHE E SW Cluster 16 ex 14 – 52+7 documents
software, sistemi elettronici, hardware, sviluppo, informatica, servizi, telecomunicazioni, ricerca, mobilità, automotive, ingegneristica, apparati elettronici, progettazione, automazione, sperimentazione, strumentazione, ambito, assistenza, sistemi, campo, apparecchi elettrici, edizione, elettronico, applicazioni, ambientale, informatici, schede elettroniche, controllo, elettronica, innovativi, sicurezza, quadri, prestazione, utilizzo, elettronici, integrati, diagnosi, tecnologico, consulenza, centraline, moduli, ingegneria, innovativo, innovative, simulazione, apparecchiature, connesso, unità, performance, prestazioni.

INTERNI AUTO E VETRI Cluster 17 ex6 – 100+23 documents
resine, **interni**, prodotti chimici, **rivestimenti**, **materie plastiche**, edilizia, materiali, articoli, **pelle**, vernici, guarnizioni, gomma, fibre, vetro, espansi, **tessuto**, tappeti, tessili, tessuti, pannelli, prodotti, mescole, **cristalli**, abrasivi, isolamento acustico, filati, colori, esterni, porta, industria, adesivi, sigillanti, materiale, **interni auto**, **arredamento**, smalti, materiale plastico, applicazione, lavorazione, **pelli**, minerali, ceramica, preparazione, plastica, legno, trasformazione, taglio, sintetici, **resine sintetiche**, **selleria**, caravan, lubrificanti, interne, **poliuretano**, tradizionali, termoplastico, termoplastiche, rivestimento, fibra, nautica, fluidi, materie prime, **vetri**, **componentistica auto**, polimeri.

VETRI Cluster 18 ex19 – 1 document
vetri, isolanti.

RICAMBI per autoveicoli Cluster 19 ex3 – 218+23 documents
autoveicoli, accessori, parti, **ricambi**, auto, motocicli, **ingrosso**, fabbricazione, motoveicoli, **frizioni**, aeromobili, **commercio**, **marmitte**, **cerchi**, veicoli industriali, freno, treni, rimorchi, autocarri, autobus, lega, **componenti auto**, macchine agricole, **freni**, produzione, **pastiglie**, **corsa**, **epoca**, autovetture, leghe leggere, silenziatori, assali, cicli, automezzi, competizione, trattori, sospensioni, organi di trasmissione, esclusi quelli idraulici, motorsport, attrito.

RICAMBI Cluster 20 ex1 – 238+28 documents
ricambi, **iniettori**, **diesel**, **veicoli**, **pompe**, auto, **motori diesel**, commercio, originali, sistemi di trasmissione, autocarri, ingrosso, **impianti frenanti**, l, sistemi, componenti, apparati, leggeri, uso, moto, commercializzazione, telai, autotelai, natanti, terra, idraulici, **metano**, **motore**, **ricambi auto**, **parti di ricambio**, autovetture, **riduttori**, **riparazione**, **dettaglio**, **olio**, **accessori** di autoveicoli, **pezzi di ricambio**, esportazione, conto di terzi, tutti, ricostruzione, macchine movimento, aerei, pesanti, **anelli di tenuta**, mezzi, mezzi di trasporto, autobus, industriali, destinati, trattori, marchio, **lubrificanti**.

Annex 5 Number of companies by Nace (3 digits) and semantic cluster

		Design	Parts, components, machining									Electrical equipment			Wind ca	Spare parts		Miscellan	Total
		Design, research and develop	Parts for motor vehicles	Transmission organs	Machining and surface treat	Mold sand shapes	Molding and surface treat	Metal profiling	Washer & hardware	Metal hardware & tools	Surface Treatments	Electrical equipment, batteries	Electronics for refrigeration	Electrical electronic equipment	windows and car interiors	Vehicle spare parts	Parts	Miscellaneous business	
		2+3-4	5	6	7	8	9	10	11	12	13	14	15	16	17+18	19	20	1	
13	INDUSTRIE TESSILI																		
	130														1				1
	131														2				2
	132														1				1
	133														1				1
	139														4	1			5
14	CONFEZIONE DI ARTICOLI DI ABBIGLIAMENTO; CONFEZIONE DI ARTICOLI IN PELLE E PELLICCIA																		
	141														1				1
15	FABBRICAZIONE DI ARTICOLI IN PELLE E SIMILI																		
	151		1												3				4
16	INDUSTRIA DEL LEGNO E DEI PR	3																2	5
	162					3												2	5
18	STAMPA E RIPRODUZIONE DI SUPPORTI REGISTRATI														1				1
	181														1				1
19	FABBRICAZIONE DI COKE E PRODOTTI DERIVANTI DALLA RAFFINAZIONE DEL P	2																1	3
	192														2			1	3
20	FABBRICAZIONE DI PRODOTTI CHIMICI																		
	201														3				3
	203				1										6				7
	205			1											3	1	1	1	7
22	FABBRICAZIONE DI ARTICOLI IN GOMMA E MATERIE PLASTICHE																		
	221			2	8	1	11		1			1			14	2	13	4	57
	222		1		5	2	28		1			4			21	3	1	6	72
23	FABBRICAZIONE DI ALTRI PRODOTTI DELLA LAVORAZIONE DI MINERALI NON METALLIFERI																		
	231														5				5
	234														1		1		2
	239											1			4		1		6
24	METALLURGIA																		
	241				1		1									1			3
	242				5							1					2		8
	243			1	2		1												4
	244				5				1										6
	245			2	22		1											1	26
25	FABBRICAZIONE DI PRODOTTI IN METALLO (ESCLUSI MACCHINARI E ATTREZZATURE)																		
	250			2	2														4
	251		1	1	3		3						1		2	1	1	2	15
	252												1					1	2
	253											1							1
	255		1	7	17		34	11	2				1		2	1	1	1	78
	256	2		26	53	3	25	1	7	5	84		9	1	6	3	17	18	260
	257				6	13	11						2		1	1	2	4	40
	259		1	5	32	1	4		5	63			4			3	3	1	122
26	FABBRICAZIONE DI COMPUTER E PRODOTTI DI ELETTRONICA E OTTICA; APPARECCHI ELETTROMEDICALI, AP																		
	261				1							2	2		1				6
	262													3					3
	263												1	4					5
	264																1		1
	265					1							1	2					4
27	FABBRICAZIONE DI APPARECCHIATURE ELETTRICHE ED APPARECCHIATURE PER USO DOMESTICO NON ELET																		
	270												2						2
	271				3		1					13	6	1			1		25
	272											8						1	9
	273											2	4	1					7
	274												1	1	1		1	1	5
	279									1		7	12	3			2	4	29
28	FABBRICAZIONE DI MACCHINARI ED APPARECCHIATURE NCA																		
	28																1		1
	280			1															1
	281		1	15	7				1				8	2		6	15	4	59
	282				5	1	3						6	1	2	3	5	2	28
	283			1												1			2
	284			1			2									1			4
	289			1	1		1						2		1		2		8
29	FABBRICAZIONE DI AUTOVEICOLI, RIMORCHI E SEMIRIMORCHI																		
	291	1	4				1		1				2	1		3	11	1	25
	292		10			1	2					1			1	4	3	1	23
	293	3	178	30	27	12	18		5	1	1	66	66	16	21	182	150	39	815

		Design	Parts, components, machining									Electrical equipment			Wind ca	Spare parts		Miscellan	Total
		Design, research and develop	Parts for motor vehicles	Transmission organs	Machining and surface treat	Mold sand shapes	Molding and surface treat	Metal profiling	Washer & hardware	Metal hardware & tools	Surface Treatments	Electrical equipment, batteries	Electronics for refrigeration	Electrical electronic equipment	windows and car interiors	Vehicle spare parts	Parts	Miscellaneous business	
		2+3-4	5	6	7	8	9	10	11	12	13	14	15	16	17+18	19	20	1	
30	FABBRICAZIONE DI ALTRI MEZZI DI TRASPORTO																		
	302															1			1
	303															1			1
	309			1	1	1	1	1						1	1	2	2	1	12
31	FABBRICAZIONE DI MOBILI																		
	310			2											3	1			6
32	ALTRE INDUSTRIE MANIFATTURIERE																		
	323														1				1
	329														1				1
33	RIPARAZIONE, MANUTENZIONE ED INSTALLAZIONE DI MACCHINE ED APPARECCHIATURE																		
	332						1								3	3		2	9
43	LAVORI DI COSTRUZIONE SPECIALIZZATI														1				1
	432														1				1
45	COMMERCIO ALL'INGROSSO E AL DETTAGLIO E RIPARAZIONE DI AUTOVEICOLI E MOTOCICLI																		
	451																2	1	3
	452			1									1		1		1	2	6
	453		2	1					1				1		1		9	12	27
46	COMMERCIO ALL'INGROSSO (ESCLUSO QUELLO DI AUTOVEICOLI E DI MOTOCICLI)																		
	460															1			1
	461																2		2
	464														2		1		3
	465															1			1
	466				1					1			1				1	3	7
	467				3					3								2	8
	469															1			1
47	COMMERCIO AL DETTAGLIO (ESCLUSO QUELLO DI AUTOVEICOLI E DI MOTOCICLI)																		
	475															1			1
	479																1		1
52	MAGAZZINAGGIO E ATTIVITÀ DI SUPPORTO AI TRASPORTI																		
	522																	1	1
56	ATTIVITÀ DEI SERVIZI DI RISTORAZIONE																		
	561			1															1
62	PRODUZIONE DI SOFTWARE, CONSULENZA INFORMATICA E ATTIVITÀ CONNESSE																		
	620		2				1							11					14
63	ATTIVITÀ DEI SERVIZI D'INFORMAZIONE E ALTRI SERVIZI INFORMATICI																		
	631						1												1
64	ATTIVITÀ DI SERVIZI FINANZIARI (ESCLUSE LE ASSICURAZIONI E I FONDI PENSIONE)																		
	642																1	1	2
	649				1														1
68	ATTIVITÀ IMMOBILIARI																		
	682													1					1
69	ATTIVITÀ LEGALI E CONTABILITÀ																		
	692					1													1
70	ATTIVITÀ DI DIREZIONE AZIENDALE E DI CONSULENZA GESTIONALE																		
	701				1					1					1	1		1	5
	702				1										1		1		3
71	ATTIVITÀ DEGLI STUDI DI ARCHITETTURA E D'INGEGNERIA; COLLAUDI ED ANALISI TECNICHE																		
	711		13			7	1								1	3	1	1	1	2	30
	712																		1	1
72	RICERCA SCIENTIFICA E SVILUPPO																		
	721		6												1	1			1	9
73	PUBBLICITÀ E RICERCHE DI MERCATO																		
	732														1					1
74	ALTRE ATTIVITÀ PROFESSIONALI, SCIENTIFICHE E TECNICHE																		
	741		7			8												1	2	18
	749		2		1										1	1	1	1	1	8
82	ATTIVITÀ DI SUPPORTO PER LE FUNZIONI D'UFFICIO E ALTRI SERVIZI DI SUPPORTO ALLE IMPRESE																		
	829															1		1	2
na			1	1	1	2	1	2		1	3		2	1			3	1		19
na			1	1	1	2	1	2		1	3		2	1			3	1		19
Totale			37	205	104	213	59	151	12	27	77	86	106	151	59	125	241	266	106	2026

Annex 6 Characterisation of clusters by category

Aggregation	Cluters label	id cluster	Variable label	Caracteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
Design	Industrial design	2	**Group: Cluster ex_ 9 / 20 (Count: 29 - Percentage: 0.16)**							
	Industrial design	2	v_PROVINCIA	C2=MO	65,52	3,48	2,92	9,42	0,000	650
	Industrial design	2	v_Regione	C3=Emilia-Romagna	65,52	10,70	0,95	6,92	0,000	1995
	Industrial design	2	v_ClDip	C7=0_1	27,59	3,43	1,25	4,44	0,000	639
	Design studies	3	**Group: Cluster ex_ex_10 / 20 (Count: 123 - Percentage: 0.66)**							
	Design studies	3	v_PROVINCIA	C2=TO	62,60	29,80	1,39	7,45	0,000	5558
	Design studies	3	v_ClDip	C7=10_19	32,52	15,56	1,38	4,60	0,000	2902
	Design studies	3	v_Regione	C3=Campania	13,01	3,90	2,20	4,07	0,000	727
	Design studies	3	v_ClDip	C7=ND	19,51	9,79	1,31	3,17	0,001	1826
	Research and experimentation	4	**Group: Cluster ex_15 / 20 (Count: 49 - Percentage: 0.26)**							
	Research and experimentation	4	v_ClDip	C7=0_1	28,57	3,43	2,19	6,08	0,000	639
	Research and experimentation	4	v_Regione	C3=Campania	28,57	3,90	1,93	5,81	0,000	727
	Research and experimentation	4	v_ClDip	C7=2_5	32,65	7,69	1,12	4,93	0,000	1434
	Research and experimentation	4	v_PROVINCIA	C2=MO	18,37	3,48	1,38	3,93	0,000	650
	Research and experimentation	4	v_Regione	C3=Piemonte	53,06	35,54	0,39	2,37	0,009	6629
Parts, components, machining	Parts for motor vehicles	5	**Group: Cluster ex_ 4 / 20 (Count: 1149 - Percentage: 6.16)**							
	Parts for motor vehicles	5	v_ATECO 2 cifre	C6=29	94,34	41,64	13,95	39,41	0,000	7768
	Parts for motor vehicles	5	v_ATECO 3 cifre	C5=293	87,55	39,46	13,67	34,96	0,000	7360
	Parts for motor vehicles	5	v_ATECO 4 cifre	C4=2932	79,55	32,32	15,16	34,06	0,000	6029
	Parts for motor vehicles	5	v_ClDip	C7=2_5	13,40	7,69	10,74	6,86	0,000	1434
	Transmission organs	6	**Group: Cluster ex_ 5 / 20 (Count: 850 - Percentage: 4.56)**							
	Transmission organs	6	v_ATECO 4 cifre	C4=2815	15,29	2,36	29,48	17,65	0,000	441
	Transmission organs	6	v_ATECO 3 cifre	C5=281	18,00	3,56	23,04	17,17	0,000	664
	Transmission organs	6	v_ATECO 2 cifre	C6=28	21,29	5,78	16,79	15,78	0,000	1078
	Transmission organs	6	v_ATECO 4 cifre	C4=2562	18,00	8,10	10,13	9,50	0,000	1510
	Machining,and surface treatments	7	**Group: Cluster ex_ 7 / 20 (Count: 2087 - Percentage: 11.19)**							
	Machining,and surface treatments	7	v_ATECO 2 cifre	C6=24	16,58	2,63	70,61	31,67	0,000	490
	Machining,and surface treatments	7	v_ATECO 2 cifre	C6=25	53,57	26,26	22,82	28,29	0,000	4899
	Molds and shapes	8	**Group: Cluster ex_11 / 20 (Count: 518 - Percentage: 2.78)**							
	Molds and shapes	8	v_ATECO 4 cifre	C4=2573	27,80	2,26	34,12	23,30	0,000	422
	Molds and shapes	8	v_ATECO 3 cifre	C5=257	27,80	2,66	29,03	22,14	0,000	496
	Molds and shapes	8	v_PROVINCIA	C2=TO	54,63	29,80	5,09	11,89	0,000	5558
	Molding, and surface treatments	9	**Group: Cluster ex_ 8 / 20 (Count: 1445 - Percentage: 7.75)**							
	Molding, and surface treatments	9	v_ATECO 3 cifre	C5=255	22,49	4,80	36,31	25,20	0,000	895
	Molding, and surface treatments	9	v_ATECO 2 cifre	C6=25	55,64	26,26	16,41	24,69	0,000	4899
	Molding, and surface treatments	9	v_ATECO 4 cifre	C4=2550	20,83	4,38	36,84	24,35	0,000	817
	Molding, and surface treatments	9	v_ATECO 2 cifre	C6=22	20,62	6,33	25,25	19,18	0,000	1180
	Molding, and surface treatments	9	v_ATECO 3 cifre	C5=222	14,53	3,59	31,34	18,28	0,000	670
	Molding, and surface treatments	9	v_PROVINCIA	C2=TO	51,49	29,80	13,39	17,96	0,000	5558
	Molding, and surface treatments	9	v_Regione	C3=Piemonte	55,71	35,54	12,14	16,26	0,000	6629
	Molding, and surface treatments	9	v_ATECO 4 cifre	C4=2573	10,17	2,26	34,83	16,11	0,000	422
	Metal profiling	10	**Group: Cluster ex_18 / 20 (Count: 137 - Percentage: 0.73)**							
	Metal profiling	10	v_ATECO 3 cifre	C5=255	100,00	4,80	15,31	29,05	0,000	895
	Metal profiling	10	v_ATECO 4 cifre	C4=2550	85,40	4,38	14,32	25,14	0,000	817
	Metal profiling	10	v_ATECO 2 cifre	C6=25	100,00	26,26	2,80	19,02	0,000	4899
	Washer & hardware	11	**Group: Cluster ex_16 / 20 (Count: 653 - Percentage: 3.50)**							
	Washer & hardware	11	v_ATECO 2 cifre	C6=25	68,15	26,26	9,08	22,79	0,000	4899
	Washer & hardware	11	v_PROVINCIA	C2=LC	21,13	2,69	27,49	19,60	0,000	502
	Washer & hardware	11	v_ATECO 3 cifre	C5=256	40,43	11,50	12,31	19,35	0,000	2145
	Washer & hardware	11	v_Regione	C3=Lombardia	58,19	26,54	7,68	17,32	0,000	4950
	Washer & hardware	11	v_ATECO 4 cifre	C4=2562	28,64	8,10	12,38	15,79	0,000	1510
	Metal hardware & taps	12	**Group: Cluster ex_17 / 20 (Count: 581 - Percentage: 3.11)**							
	Metal hardware & taps	12	v_ATECO 3 cifre	C5=259	80,38	5,79	43,24	48,13	0,000	1080
	Metal hardware & taps	12	v_ATECO 4 cifre	C4=2594	59,72	2,71	68,58	45,15	0,000	506
	Metal hardware & taps	12	v_ATECO 2 cifre	C6=25	88,98	26,26	10,55	32,46	0,000	4899
	Metal hardware & taps	12	v_Regione	C3=Lombardia	69,54	26,54	8,16	21,97	0,000	4950
	Surface Treatments	13	**Group: Cluster ex_20 / 20 (Count: 271 - Percentage: 1.45)**							
	Surface Treatments	13	v_ATECO 4 cifre	C4=2561	90,04	3,30	39,67	39,93	0,000	615
	Surface Treatments	13	v_ATECO 3 cifre	C5=256	93,73	11,50	11,84	31,60	0,000	2145

Aggregation	Cluters label	id cluster	Variable label	Caracteristic categories	% of category in group	% of category in set	% of group in category	Test-value	Probability	Weight
Electrical equipment	Electrical equipment, batteries; engine	14	**Group: Cluster ex_12 / 20 (Count: 713 - Percentage: 3.82)**							
	Electrical equipment, batteries; engine	14	v_ATECO 4 cifre	C4=2931	65,08	5,57	44,66	44,78	0,000	1039
	Electrical equipment, batteries; engine	14	v_ATECO 2 cifre	C6=27	24,54	3,73	25,18	20,81	0,000	695
	Electrical equipment, batteries; engine	14	v_ATECO 3 cifre	C5=293	68,86	39,46	6,67	16,15	0,000	7360
	Electronics for refrigeration and heati	15	**Group: Cluster ex_13 / 20 (Count: 1658 - Percentage: 8.89)**							
	Electronics for refrigeration and heati	15	v_ATECO 4 cifre	C4=2931	21,23	5,57	33,88	23,45	0,000	1039
	Electronics for refrigeration and heati	15	v_ATECO 2 cifre	C6=27	14,11	3,73	33,67	18,72	0,000	695
	Electronics for refrigeration and heati	15	v_Regione	C3=Emilia-Romagna	20,51	10,70	17,04	12,33	0,000	1995
	Electrical electronic equipment and sv	16	**Group: Cluster ex_14 / 20 (Count: 601 - Percentage: 3.22)**							
	Electrical electronic equipment and sv	16	v_ATECO 2 cifre	C6=27	11,65	3,73	10,07	8,43	0,000	695
	Electrical electronic equipment and sv	16	v_ATECO 4 cifre	C4=2931	14,64	5,57	8,47	8,30	0,000	1039
	Electrical electronic equipment and sv	16	v_PROVINCIA	C2=TO	43,09	29,80	4,66	6,98	0,000	5558
	Electrical electronic equipment and sv	16	v_Regione	C3=Piemonte	48,59	35,54	4,40	6,62	0,000	6629
	Electrical electronic equipment and sv	16	v_ClDip	C7=0_1	7,99	3,43	7,51	5,33	0,000	639
	Electrical electronic equipment and sv	16	v_ClDip	C7=250_499	9,48	4,90	6,24	4,68	0,000	914
	Electrical electronic equipment and sv	16	v_Regione	C3=Lazio	5,32	2,10	8,16	4,66	0,000	392
	Electrical electronic equipment and sv	16	v_Regione	C3=Abruzzo	6,32	3,24	6,29	3,81	0,000	604
	Electrical electronic equipment and sv	16	v_ClDip	C7=ND	13,98	9,79	4,60	3,28	0,001	1826
	Electrical electronic equipment and sv	16	v_ClDip	C7=100_249	14,48	10,38	4,49	3,14	0,001	1936
	Electrical electronic equipment and sv	16	v_ClDip	C7=500_999	4,99	2,98	5,41	2,64	0,004	555
Windows and car interiors	windows and car interiors	17	**Group: Cluster ex_ 6 / 20 (Count: 1085 - Percentage: 5.82)**							
	windows and car interiors	17	v_ATECO 2 cifre	C6=22	22,03	6,33	20,25	17,72	0,000	1180
	windows and car interiors	17	v_ATECO 3 cifre	C5=222	13,36	3,59	21,64	14,09	0,000	670
	windows and car interiors	17	v_ATECO 3 cifre	C5=221	8,66	2,73	18,43	10,02	0,000	510
	windows and car interiors	17	v_ClDip	C7=50_99	20,83	12,91	9,38	7,46	0,000	2409
	windows and car interiors	17	v_PROVINCIA	C2=MI	15,02	8,48	10,30	7,27	0,000	1582
	windows and car interiors	17	v_PROVINCIA	C2=BG	7,74	3,32	13,55	7,20	0,000	620
	glasses	18	**Group: Cluster ex_19 / 20 (Count: 23 - Percentage: 0.12)**							
	glasses	18	v_PROVINCIA	C2=CH	100,00	2,09	5,90	13,13	0,000	390
	glasses	18	v_Regione	C3=Abruzzo	100,00	3,24	3,81	12,32	0,000	604
Spare parts	Vehicle spare parts	19	**Group: Cluster ex_ 3 / 20 (Count: 2204 - Percentage: 11.82)**							
	Vehicle spare parts	19	v_ATECO 4 cifre	C4=2932	71,73	32,32	26,22	40,59	0,000	6029
	Vehicle spare parts	19	v_ATECO 2 cifre	C6=29	77,22	41,64	21,91	36,21	0,000	7768
	Vehicle spare parts	19	v_ATECO 3 cifre	C5=293	74,23	39,46	22,23	35,32	0,000	7360
	Vehicle spare parts	19	v_ClDip	C7=ND	16,24	9,79	19,61	10,11	0,000	1826
	Parts	20	**Group: Cluster ex_ 1 / 20 (Count: 2807 - Percentage: 15.05)**							
	Parts	20	v_ATECO 2 cifre	C6=29	63,73	41,64	23,03	25,57	0,000	7768
	Parts	20	v_ATECO 4 cifre	C4=2932	53,12	32,32	24,73	24,79	0,000	6029
	Parts	20	v_ATECO 3 cifre	C5=293	58,75	39,46	22,40	22,40	0,000	7360
	Parts	20	v_ATECO 2 cifre	C6=45	6,20	2,26	41,33	13,15	0,000	421
Miscellany	Miscellaneous business	1	**Group: Cluster ex_ 2 / 20 (Count: 1670 - Percentage: 8.95)**							
	Miscellaneous business	1	v_ATECO 3 cifre	C5=256	20,18	11,50	15,71	10,75	0,000	2145
	Miscellaneous business	1	v_ClDip	C7=250_499	10,60	4,90	19,37	9,99	0,000	914
	Miscellaneous business	1	v_ATECO 2 cifre	C6=22	11,44	6,33	16,19	8,19	0,000	1180
	Miscellaneous business	1	v_ATECO 4 cifre	C4=2561	6,95	3,30	18,86	7,73	0,000	615
	Miscellaneous business	1	v_ATECO 4 cifre	C4=2562	13,23	8,10	14,64	7,48	0,000	1510
	Miscellaneous business	1	v_PROVINCIA	C2=LC	5,81	2,69	19,32	7,24	0,000	502
	Miscellaneous business	1	v_ClDip	C7=100_249	15,63	10,38	13,48	6,94	0,000	1936
	Miscellaneous business	1	v_ATECO 3 cifre	C5=222	6,89	3,59	17,16	6,80	0,000	670
	Miscellaneous business	1	v_PROVINCIA	C2=TO	36,83	29,80	11,07	6,44	0,000	5558
	Miscellaneous business	1	v_PROVINCIA	C2=BS	8,98	5,33	15,08	6,39	0,000	995
	Miscellaneous business	1	v_ATECO 2 cifre	C6=27	6,47	3,73	15,54	5,65	0,000	695

Annex 7 Number of companies and employees in the population and in the respondent companies, by Nace code, 2 digits

Source: Anfia satabase, 2017