POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Gestionale

Tesi Di Laurea Magistrale

Literature review of Big Data technology applications to the Supply Chain

Relatore:
Prof.ssa Anna Corinna Cagliano
Dott.ssa Mahsa Mahdavisharif

Candidato:
Chiara Acquasanta

Academic Year 2021/2022

Index

Figure Index

Introduction

In a society that has made the Web a necessity, the process of digitization is now considered a fundamental element and present in every company's foundation. Therefore, digitization is strictly necessary to be able to ensure that each company achieves an optimum level of competitiveness. For this reason, supply chains are increasingly attracted to this issue. For a supply chain to be considered digital, it isn't enough to convert documents and all the information contained in paper format to digital format. On the contrary, the digitization process requires considerable investment both in terms of cost and in terms of structural and managerial changes. Therefore, it is important for each company to be able to analyze its own starting condition, highlighting the critical issues on which action needs to be taken.

Having clarified the importance attached nowadays to the digitization process, this research work aims to study the application and its influences of a particular technology in the supply chain environment. This is a very recent technology called *Big Data*. In order to be able to delve into what activities and areas of the supply chain are impacted by Big Data, a literature review was performed using a particular research technique, namely *Snowballing*. In general, it was thus found that Big Data has a great influence on demand forecasting and customer relationship, and to achieve excellent results in the activities just mentioned, it is also important to share information among the various actors in the supply chain. Not coincidentally, Big Data plays a great role in the latter area of the supply chain. The following lines describe the structure that characterizes this thesis.

This research paper is divided into four chapters. Chapter 1, following an introduction of the concept of *Supply Chain* and *Industry 4.0*, proceeds to explore the enabling technologies of Industry 4.0 until it presents the concept of *Digital Supply Chain* and consequently the application of the star technology of this thesis, namely *Big Data*, in the supply chain environment. Thus, it is deduced how the first chapter aims to provide the right tools for the reader to be able to better understand the results that emerged.

Chapter 2, on the other hand, sets out and describes in detail the methodology through which the results were achieved. In this case, too, the chapter provides a brief introduction on *Systematic Literature Review* and *Snowballing* before discussing the research method in detail. Chapter 3 is completely focused on exposing the results of this research work and the critical analysis to which they were subjected.

In conclusion, in order to know what are the benefits, limitations and future research opportunities of this study, it is possible to read Chapter 4.

1. Applications of Industry 4.0 to the supply chain

The chapter aims to provide to the reader the main notions regarding the dominant topics of this research thesis in order to facilitate the understanding of the results exposed in Chapter 3. In fact, in addition to the introduction of the general concept of supply chain, the theme of the chapter also revolves around the concept of Digital Supply Chain, Industry 4.0 and its enabling technologies. Finally, there is an in-depth analysis of the focus of this study, namely Big Data and its influence on the supply chain.

1.1 Supply Chain

In order to understand the concept of a supply chain (SC) and what activities are included, it is necessary to know its definition. The supply chain can be defined as:

A system of organizations, people, activities, information, and resources involved in the process of transferring or providing of a product or service from supplier to customer [1].

This is a rather complex process, in which raw materials play a key role in the initial stages. Then there is the realization of the finished product and its warehouse management. The process ends with the supply of the finished product to the customer. Each step involves different professional figures, in fact it is possible to define them as *actors of the supply chain*.

From the supply chain concept derives the concept of *supply chain management* (SCM). Supply chain management is therefore the coordination of the various phases that compose the supply chain and aims to improve performance and the entire flow of resources and products. The three main flows of the supply chain are in fact the product flow, the information flow and the financial flow [2]. *Supply chain management* (SCM) coordinates and integrates these flows, both internally and between companies.

Many people equate supply chain with *logistics*, but logistics is really only one component of the supply chain.

Supply chain activities include procurement, product lifecycle management, supply chain planning (including inventory planning and maintenance of production lines and company assets), logistics (including transportation management), and order management.

Logistics is exclusively concerned with the management of goods, in particular their movement. *Supply chain management* includes many other activities, for example: marketing, supplier relations, procurement, inventory management and storage, production, purchasing management, delivery management, customer relationship management, etc.

The supply chain concept is quite dated. Supply chain management has continued to receive attention from managers, researchers and consultants since the early 1980s.

Many people identify the work *Supply Chain Management: Logistics catches up with strategy* published in 1982 by R. K. Oliver and M. Webber, as the first time that the term *supply chain management* was used [3]. It is possible to demonstrate the strong diffusion of the topic through the analysis of the trends of the main databases (EBSCO, Scopus, Google Scholar...) that show, starting from the end of the 1980s, an increase in the number of citations related to SCM. For many years, supply chain management was viewed as a linear function, managed by supply chain experts. Due to progressive changes, especially in technology, SCM systems have become increasingly sophisticated and complex. The Internet, the evolution of technological systems, and the advent of the demand-driven global economy have hindered the consideration of the supply chain as a simple linear function [4].

Nowadays, the supply chain is viewed as a complex set of networks. The key players in these networks are consumers, who, given the vastness of the marketplace, expect their orders to be processed within fairly short time periods. Thus, demand-driven supply chain management models, that can combine information, technologies and processes with high accuracy and speed, are needed. Although supply chain management has always been considered a very important issue for companies, today this importance seems to be tripled. The supply chain can be considered as an indicator of business success, in fact, only companies characterized by an effective SCM and able to adapt quickly to continuous technological changes, can be considered successful companies and able to survive the perennial volatility of today's world. An efficient and well-organized supply chain can certainly be considered a source of competitive advantage, as it allows to optimize available resources, avoid wastefulness and satisfy or anticipate customer requests.

In summary, there are basically three causes that drive researchers and managers to focus on *supply chain management* [3]:

- *Added value*, in recent times, is progressively increased and technological development has led to the need for an efficient supply chain for a company in order to survive;
- *Technological development,* that has improved a company's ability to control every stage of the supply chain, allowing each of them to be constantly updated;
- *Price control*, which is no longer a sufficient activity to guarantee the satisfaction of the market, only careful management of the supply chain can ensure all this.

Since its origin, supply chain management has been focused on the concept of efficiency. This priority has gradually become increasingly important over the years but, recently to support this priority, there is the role of the customer. Customer loyalty is based on the company's ability to satisfy their expectations and orders with precision and reasonable time. For this reason, companies need to analyze supply chains from the perspective of customers.

A new necessity that companies need to pay attention, for continuous evolution, is agility [4]. This is an aspect of the supply chain that is necessary for it to be considered efficient. For example, to date, consumers generally have greater purchasing power because there are different channels through which they can access the market (shops, websites, etc.). They also have higher expectations about the degree of customization of a product. An agile supply chain can address these necessities.

1.2 Industry 4.0

In order to respond to the priorities of the supply chain, for example the role of client, described in Section 1.1, it is important to seize the development opportunities offered by each technology. These technologies characterize Industry 4.0.

As it can be observed in Figure 1, the fourth industrial revolution was preceded by three other revolutionary changes, namely: the first change (Industry 1.0) happened between 1770 and 1870, with the introduction of the steam machine that allowed a mechanization of production. The second revolutionary change (Industry 2.0) occurred in the late 1800s and early 1900s with the discovery of electricity and the introduction of mass production. Then, the third revolutionary change (Industry 3.0) occurred towards the end of the 20th century with the computerization of production and organization, where the introduction of information and communication micro technologies accentuated the systematization of the production process [5].

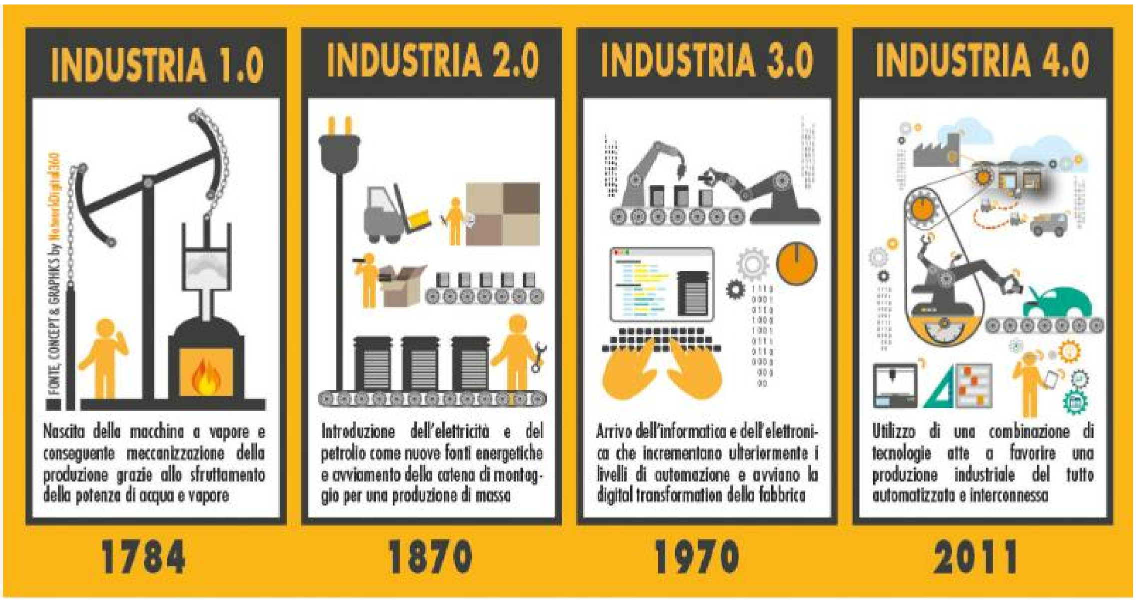

Figure 1 - A journey from the first to the fourth industrial revolution [5]

Industry 4.0 was born in 2011 in Germany. In detail, the term was first used at the Hanover Fair of automation and technology. The term was coined by a German working group with the goal of promoting an improvement of the manufacturing industry through technological innovation of the entire industrial production. In 2012, the group presented a plan for the implementation of Industry 4.0 to the German government [5]. Industry 4.0 is characterized by the introduction of new production technologies, which allow factories to have vertically and horizontally integrated production, flexible processes that enable individualized mass production, intelligent machines that exchange data with each other and control production and logistics processes independently. The technologies that appear in this latest version of industrialization are numerous, including artificial intelligence, machine learning, Internet of Things, automation and sensors. They are transforming the way in which companies produce, maintain and distribute new products and services. Industry 4.0 is based on the supply chain. Industry 4.0 has literally changed the way in which the supply chain is managed and analyzed. In other words, the use and application of technologies to the supply chain have undergone a revolution. It is possible to demonstrate this through a small example regarding the maintenance of production machines that make up the supply chain. In fact, while in the past technologies were used only in case of malfunction, today such malfunctions can be predicted [4]. It can be deduced that technologies help managers to deal with the risks of interruption to the SC and act quickly if necessary.

Today's *supply chain management* is about using technology to make the supply chain and the company smarter. In addition, technology support improves demand forecasting and consequently production planning. In this way, companies can, not only meet the requirements of customers, but they can also achieve financial goals through production cost containment. Intelligent supply chain management also benefits employees, who can take advantage of the automation of the most repetitive production processes, focusing their attention on the most complex steps and consequently increasing the quality of the final output.

1.3 Industry 4.0 and Enabling Technologies

Industry 4.0 represents a great opportunity of evolution for the supply chain, in fact the fourth industrial revolution allows to develop new business models able to generate smart and competitive products.
Industry 4.0 is divided into nine pillars, each represented by a technology, as shown in Figure 2 [6]:

- Advanced Manufacturing Solutions;
- Additive Manufacturing;
- Augmented Reality;
- Simulation;
- Horizontal/Vertical Integration;
- Internet of Things;
- Cloud Computing;
- Big Data Analytics;
- Cyber security.

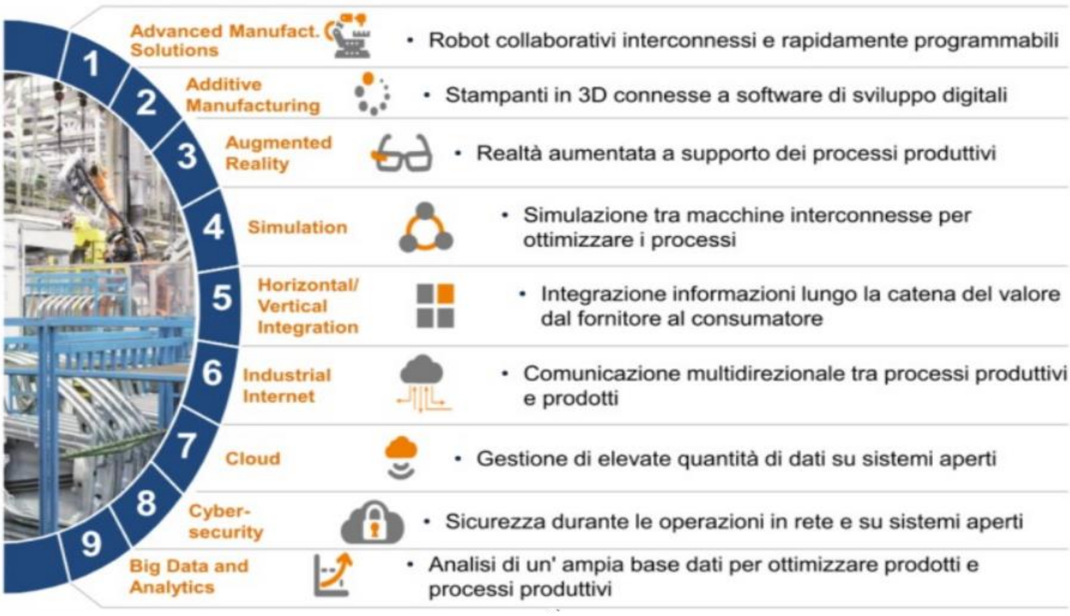

Figure 2 - Industry 4.0 enabling technologies [8]

In fact, although the Italian Ministry of Economic Development (MiSE) lists nine pillars, it is important to stress that this set of technologies is in continuous progress [7]. In fact, to this first list, it is possible to add two more fundamental pillars:

- Artificial Intelligence and Machine Learning;
- Blockchain.

These technologies are known as *key enabling technologies* (KET). According to the European Commission's definition, KETs are:

knowledge-intensive technologies associated with high levels of R&D, rapid innovation cycles, substantial investment expenditure and highly skilled jobs [8].

These are resources and tools that, thanks to the interconnection between them and with the Internet, allow companies to make their production processes more efficient and, above all, allows them to add value. Consequently, it is important to remember that added value is a source of competitive advantage. The positive impact of these technologies on the supply chain and its actors is evident.
In the following sections, each pillar of Industry 4.0 is briefly described.

1.3.1 Advanced Manufacturing Solutions

Advanced Manufacturing Solutions aren't really just one single technology, but on the contrary, they can be defined as advanced production systems, that is, interconnected and modular systems that provide flexibility and high performance. This first pillar includes automated handling systems and advanced robotics, which today enters the market with collaborative robots or cobots [8]. Cobots are designed to support people in their workspace, so they are not technologies that completely replace the human figure. In general, robots are connected with software and systems and increase the efficiency and agility of more complex production systems. They reduce the probability of error, time and cost.

1.3.2 Additive Manufacturing

Additive Manufacturing and 3D printers are often considered synonymous and can be defined as additive manufacturing systems that increase the efficiency of material use. They describe a set of technologies that can create three-dimensional objects from a design that takes shape in a very short time [8]. Additive Manufacturing allows, unlike traditional manufacturing which was based on the removal of material from a raw component, to add material through the technique of layering, or layer on layer. This technology allows to create complex and customized products, but the great advantage is represented by the possibility of reducing production time, less raw material used and lower storage costs. Raw materials may vary depending on the technology, while the common element in all manufacturing methods is the use of a computer with software. Additive Manufacturing and 3D printers are not recent, in fact, their origin dates back to 1984 and, due to technological evolution, they have undergone great changes [9]. Today, finished products are made with a greater variety of raw materials and, above all, the cost of the machines has decreased in comparison to the early years.

1.3.3 Augmented Reality

Vision systems with Augmented Reality (RA) better guide and support operators in their daily activities. Augmented reality integrates the physical world with different types of fictional information, for example sound, visual, textual, etc. [8]. The most widely used devices are smartphones and tablets, but there are many others. This technological pillar brings many benefits for logistics: it facilitates picking and makes order control more dynamic and accurate [7]. There are many other benefits, for example, RA devices indicate machinery failures, allow the transformation of written information into vocal messages, graphically represent work processes in order to make them more intuitive for the operator. Augmented reality facilitates remote collaboration by providing expert assistance in solving problems, replacing the need for people to be physically present.

1.3.4 Simulation

Simulations cannot be considered a real technology; on the contrary it constitutes a virtual representation of industrial production processes. In detail, Simulations offer the ability to carry out a process or service through digital tools. In this way, processes can be tested even before they are officially launched, allowing for a reduction in costs if physical machinery does not need to be installed [8].

Production processes can then be improved to increase the quality of the final product. The simulations use a large number of data and information in real time coming from the production or logistic centers, in fact in order for this technological pillar to be used it is necessary the collaboration of many other technologies (Internet of Things, Cloud Computing, Big Data). Thanks to this last feature, simulations allow companies to be able to reconfigure their strategies in a shorter period of time.

In order to allow effective simulations and to obtain updated data in real time, it is necessary to introduce the concept of *Digital Twin*. It is a logical-mathematical replica of a real process that composes, for example, the production process of the supply chain. In other words, Digital Twin is a mathematical model that translates into virtual terms what happens in a real process or system. In order to obtain an optimal mathematical model, it is essential to identify each component of the system; namely buffers, resources, transports, etc. By implementing the model in a code, it can be considered ready to use of benefits offered by simulations with Digital Twin [10].

1.3.5 Horizontal and Vertical Integration

A common keyword for all Industry 4.0 enabling technologies is definitely "interconnection". This term takes on greater importance in the case of Horizontal/Vertical Integration. In fact, Vertical or Horizontal Integration can be used only through interconnected technologies, which allow the creation of a real-time data sharing network. Vertical integration involves customers and suppliers, on the contrary, horizontal integration involves all companies in the same sector. The exchange of information and the digitalization of this communication between the above-mentioned parties facilitates the production process, reducing costs and increasing the added value of the final product. For this reason, vertical and horizontal integration don't really represent a technology, on the contrary they are a structural and organizational evolution of the supply chain on which digitization can have a strong impact. This fifth pillar of Industry 4.0 can be defined as a set of strategies used by companies that belong to the same sector or production process [11]. In particular, Industry 4.0 vertical integration provides a great competitive advantage to the company: it allows to detect and respond very quickly to continuous changes in the market, offering the possibility to capture new development opportunities. Thanks to this aspect, companies can advance faster within a market that seems to be increasingly competitive. In practice, vertical integration refers to the integration of technologies that facilitate communication between machines and equipment among factories. This leads to greater availability of production data and subsequent process improvement due to the ability to monitor parameters in real time. In addition, in order to respond to more complex manufacturing processes, horizontal integration involves Cyber Physical System (CPS) networks that increase the levels of automation, flexibility, and agility in the supply chain [11]. CPSs represent a set of systems of different nature that aim to achieve control over the physical process [12].

1.3.6 Internet of Things

The Internet of Things (IoT) represents a set of devices connected to the Internet. These devices are not only TVs, tablets or smartphones, but it is also necessary to think of all those that take part in the production process or even more simply in an office. For example, IoT enables the improvement of the new product development process through data derived from previous versions of related products or improvement of product quality control or workplace safety. Thanks to the use of IoT, it is possible to monitor the entire distribution process of the supply chain, thus promoting the movement of goods from the producer to the final customer under the established conditions [13]. Moreover, the introduction of IoT in an office ensures customized workplaces according to the work needs, this allows for easily adjustable lighting or heating systems hence reducing consumption and business costs [14].

In other words, the term *Internet of Things* refers to the application of technological components within physical objects, making them smart and able to communicate with each other through the use of the internet and a common language [8]. Communication between the customer, actors in the design and production phase, and suppliers is one of the benefits provided by IoT. Moreover, logistics is an area of the supply chain that benefits most from the use of such technology. In fact, it is useful for predictive maintenance, effective inventory management, improved lot traceability and safer operations in the product lifecycle.

This technological pillar presents a criticality. On the one hand, any device connected to the internet offers the advantage of being able to receive and use a large amount of data in real time, but on the other hand it represents a risk concerning the management of data. This data could be subject to cyber-attacks and it is important that each company uses appropriate solutions to mitigate the risk and ensure its protection. It's no coincidence that the goal of the Internet of Things is to empower the company and not compromise it.

1.3.7 Cloud Computing

The term cloud computing refers to the technology that makes it possible its operation, that is, an IT infrastructure, useful for sharing data and applications via the Internet. The continuous evolution of today's world drives companies towards business models capable of managing a huge amount of information [8]. Given the current state of the majority of supply chains and the fragmented sharing of information, cloud technologies offer a fair solution to these issues. There are cloud applications on the market that are useful for transportation management, equipment maintenance and repair. For example, there are systems that can automate the management of shipments, providing critical delivery and pickup data needed to reduce transportation costs. Through these solutions, more efficient suppliers can be identified. In conclusion, automating processes allows all parties involved to more easily access information and therefore take more "informed" decisions. The use of a cloud in a company optimizes logistics and encourages synergistic collaboration between the various human parts, makes data archives accessible at any moment, and finally reduces the costs connected to hardware and software [15].

1.3.8 Big Data Analytics

Traditional databases that were used to support businesses are no longer considered suitable for containing the large volume of data that needs to be tackled today. In this regard, it is possible to introduce the term Big Data, which identifies the large set of available data that companies have. In particular, the term Big Data Analytics represents the set of processes and technological systems that facilitate the analysis of data through which it is possible to make predictions. Big Data is therefore a key enabling technology (KET), which since the 1990s is no longer a simple support for business managers, but a real starting point for making more effective decisions. The future perspectives for each company using this technology are quite significant and positive, so it is necessary to have state-of-the-art architectures and adequately trained employees in order to take full advantage of the benefits [7]. Only through the acquisition of appropriate skills, companies can customize their path related to Big Data and obtain added value, as well as competitive advantage. The benefits guaranteed by the use of this enabling technology in the supply chain are discussed in the following subchapters.

1.3.9 Cyber security

The technologies described so far have shown great potential, especially in accessing, sharing and storing data. However, data is not only a source of competitive advantage, it also exposes companies to risk. In a world constantly connected to the internet, these risks increase not only in terms of volume but also in terms of entity. It is for these reasons, that cyber security can be considered the main pillar of Industry 4.0. The term cyber security identifies the set of technologies required to ensure that the company and the entire supply chain can protect the information relevant for their business [8]. In order for an adequate level of data defense to be used, each supply chain must be able to identify and assess risks. In addition, there are cyber security technology solutions that do not necessarily require the installation of software, so this is definitely a benefit to the supply chain, consisting of reduced IT costs [19].

1.3.10 Artificial Intelligence and Machine Learning

The term Artificial Intelligence (AI) indicates machines or systems that use human intelligence to perform and improve tasks. Artificial intelligence aims to enhance the capabilities of human resources without replacing them completely, for this reason this technology is useful in order to improve the production process and automating all those repetitive steps that previously required the expenditure of human energy. The use of AI in supply chains has been constantly increasing in recent years. The positive results brought by AI in SC are numerous and cover different aspects: extraction of important data from customers and suppliers, simplification of supply, demand and inventory management, accurate demand forecasting, optimization of logistics and warehouse management, automation of vehicles in distribution centers [16]. To justify these benefits there is machine learning, which is another aspect of artificial intelligence. It analyzes highly complex data and develops useful models for predicting future trends. The ability to manage data from different sources enables AI to effectively control the flow of information. AI thus uses a wide spectrum of knowledge, necessary for warehouse management and digitization of the vehicles.

Machine Learning (ML) is a subset of artificial intelligence that creates systems in order to improve performance depending on the data used. Often ML and AI are used as synonyms but, in reality, they do not indicate the same thing and do not have the same meaning. In detail, everything that indicates Machine Learning is part of artificial intelligence, the same can’t be said for AI. Machine Learning bases its learning mode on human intervention for data processing [17]. There are various examples of application of ML in the corporate world, such as computer voice recognition or conversion of speech into written text, digitization of customer service (online chatbots are replacing human resources in customer support through the proposal of automatic answers to frequently asked questions), promotion of algorithms capable of proposing complementary or alternative products to customers through the analysis of their previously made purchase data [17].

1.3.11 Blockchain

The blockchain or block structure is a digital ledger containing data that is shared securely between the various blocks that make up the structure. The data is protected through the use of cryptography, in fact, the blockchain can be considered a useful technology because it ensures the protection of the registry and avoids the presence of duplicate information. The data store can be shared but not modified. In case of modification all parties are informed and can trace the author, for this reason to modify the registry it is necessary to receive consent from each component. Depending on the type of blockchain, transactions between different blocks can be of different types, for example money, information or documents [18]. There are mainly three aspects of the supply chain affected by blockchain: decentralization, collaboration and traceability. As explained in the previous lines, the blockchain is a structure composed of multiple nodes, each having access to information. This decentralization of data means that the supply chain is able to deal with any disruptions due to the malfunction of a node. In addition, the new data storage methods associated with blockchain implementation allow top management to be able to modify the governance of existing collaborations with other entities, such as suppliers. Finally, this technological pillar makes traceable any information exchange that takes place between both internal and external actors in the supply chain. The increased traceability facilitates the identification of possible risks or problems, consequently allowing for more immediate resolution [18].

1.4 Digital Supply Chain

In addition to the supply chain priorities mentioned in Section 1.1, there is also agility, a necessary prerequisite for each company to gain competitive advantage. In order to acquire an agile supply chain, it is necessary to adopt the digital technologies that, with Industry 4.0, have revolutionized the market. In this regard, people talk about *Digital Supply Chain* (DSC). In fact, with the advent of the fourth industrial revolution, the supply chain has undergone a digitization process. The main applications of the Digital Supply Chain concern [20]:

- Inventory;
- Product logistics;
- Electronic tracking.

Digital Supply Chain therefore includes a large number of activities relating to the traditional supply chain and it is of fundamental importance for the optimization of resources and reduction of business wastefulness. Only with an excellent understanding of the analog supply chain, the new digital version of the supply chain can be effectively managed.

Given the rapid changes in technology, there are mainly two strategic approaches adopted by the top management of Digital Supply Chain [20], namely:

- *Lean* approach, a strategy whose main objective is to streamline processes and reduce wastefulness (use of materials, machines or labor that does not add value to the product);
- *Agile* approach, a strategy whose main objective is to give flexibility and vulnerability to the supply chain, ensuring the ability to adapt in a short time to the continuous changes that characterize the competitive market.

The role attributed to the many technological pillars that compose Industry 4.0 within the supply chain affects many aspects.

Nowadays, the complexity of the supply chain has increased and this can be attributed to several reasons, including the increasing number of actors interacting with the supply chain. The increase in interactions leads to the growth in the number of information, which automatically makes transparency necessary. It, in fact, makes the supply chain able to react in real time to external incentives and even anticipate them. Digitization promotes manufacturing transparency and makes the supply chain more efficient in meeting customer needs [21]. In reality, applying digitization to all supply chain activities is not so simple. This difficulty can be explained by the large number of facets that characterize supply chain processes. Not all manufacturing companies have already digitized their supply chains; some organizations have only undertaken this path at certain levels of the supply chain [21]. It is important that, all companies aiming to compete on a global scale, understand the need to anticipate this process and thus play in advance of their competitors.

Another positive impact conferred on the supply chain through digitization concerns the installation of new production machines [21]. In fact, having up-to-date machines makes the SC more efficient in the production phases; for example, making more products in less time. This translates into economic advantages and better response to the client. Technological advances in machines also bring advantages in the handling (unloading, packaging, picking up) of goods, avoiding the physical effort previously required by human resources.

Industry 4.0 has introduced new approaches to product distribution that can reduce lead times and avoid inventory management losses. Through advanced forecasting techniques, such as predictive data analytics, supply chains are more flexible and durable than planning times [21]. The Digital Supply Chain is able, through its capabilities, to focus its attentions also on the relationship with the customer, which today is the protagonist of the market. Real-time knowledge of information allows the supply chain to respond to customers' requests for product customization. In fact, in contrast to some years ago, when customers were willing to wait up to a week to receive a standard product, today this is no longer possible. With the inclusion of colossus like Amazon, customers are making their decisions depending on the degree of customization and speed with which they receive their product.

In this digital age, companies must shift their focus from cost reduction and function management toward how their supply chains can generate real value from innovative new technologies to meet new customer expectations, enable new processes, and make their organizations more integrated, transparent, and agile.

Due to the many benefits related to supply chain digitization, companies often rush into initiating this transformation without giving too much importance and attention to the initial analysis of their starting point. In fact, thinking that this process will necessarily lead to success, almost always results in not being the right approach to take [21]. All of the enabling technologies described above have different cost implications, different skill requirements, different technology requirements, and different implementation times. It is necessary to understand which of the technologies are most crucial and should receive the most attention. In order to achieve a solid business strategy, with positive results in the medium to long term, a customized transformation must be undertaken taking into consideration the needs of the company's business. It is essential to execute a real scanner that brings to light the initial criticalities and clarifies which are the final goals.

In substance, in this first phase, top management assumes the role of protagonist. In the majority of cases, companies have difficulties in managing data, which therefore represent the first frontier to be addressed in order to start a valid digitalization project. Many companies have difficulties in monitoring inventory, controlling production phases and consequently tracking the entire product life cycle [22]. The solution is to assign more importance to data. This requires a system capable of collecting, analyzing and storing all information, without neglecting any node of the supply chain. The difficulty in optimizing processes often comes from inconsistent data [22]. It is therefore important to leverage quality and, above all, meaningful data. It can be deduced that the effectiveness of the Digital Supply Chain depends on the updating and accessibility of data, which is often only accessible to a limited number of people. This aspect is the source of a major limitation, because the data could be useful to different departments in the supply chain; such as sales team, purchasing team, maintenance team, etc. As anticipated, it is necessary that not only a digital transformation of everything material takes place, but on the contrary, it is important that there is a change in the company culture [22].

1.5 Big Data and Supply Chain

In a contemporary business environment, where market forces drive economic performance, competitiveness in a company has become strictly necessary. Competitiveness enables a company to establish itself in new markets, achieve superior performance and overcome the competition. So far, the concept of Digital Supply Chain has been described, highlighting what are the benefits of Industry 4.0. In reality, these benefits mainly revolve around a particular enabling technology, namely *Big Data* (BD). This study aims to identify what are the impacts of Big Data on the supply chain and in particular what aspects of the supply chain are affected by *Big Data analytics* (BDA).

1.5.1 Big Data Analytics Definition

BDA, usually defined as the application of advanced analytical techniques, helps an organization analyze huge amounts of data [23]. In reality, there are many other definitions (Cetindamar *et al*., 2021):

BDA is defined as the process of using advanced technologies to examine BD, to reveal useful information, for making better decisions across business processes among functions or companies (Waller and Fawcett,2013).

BDA is defined as applying tools (e.g., database and data mining tools) and techniques (e.g., analytical methods) that a company can employ to analyze large-scale, complex data for various applications intended to augment firm performance in various dimensions.

BDA is defined as the application of multiple analytic methods that address the diversity of BD to provide actionable, descriptive, predictive, and prescriptive results.

BDA is defined as applying tools and processes to large and disperse datasets to obtain meaningful insights.

However, despite the existence of multiple definitions, each of them highlights the great benefits obtained through the use of BDAs, without ever mentioning the resources needed to transform BDs into managerial actions. It is for this reason that the acronym Big Data Analytics Capability (BDAC) is often mentioned when discussing Big Data. BDAC refers to a firm's capabilities to obtain information from BDs. In 72% of cases, scholars describe BDAC with different human capabilities, particularly technical and managerial skills, and technological and relational knowledge (Cetindamar *et al*., 2021).

Through BDA, it is possible to gain insights that are relevant to the company's business. This improved knowledge supports the company in exploring new markets, increasing its operational and economic efficiency, and facilitating new product development. In conclusion, Big Data Analytics can be considered an asset for organizations.

1.5.2 From 3V Model to 6V Model

Due to the continuous evolution of production machines, there is no specific number that can identify the exact size of the dataset, not surprisingly, Big Data is continuously growing.

In 2001, analyst Doug Laney defined this growth model as a three-dimensional model, known as the *3V Model*.

The 3V acronym identifies the three main factors - Volume, Velocity and Variety - that have characterized Big Data from the beginning. To date, BDs have been associated to 5V [24]:

- *Volume*, not coincidentally the name Big Data is associated with a huge amount of data and in order to determine its value, it is important to assess its size. This volume of data can' t be analyzed with traditional technologies and especially there is a strong growth in this regard. For this reason, researchers have difficulty in determining the threshold above which it is possible to speak of Big Data.
- *Velocity*, it is an aspect related to the speed with which data is accumulated. As with volume, in recent years there has been an increase in the speed with which data is acquired. Each technological device has sensors that can accumulate data in real time, so organizations not only face the difficulty of storing a high volume of data, but they must at the same time be able to analyze it in a rather limited amount of time.
- *Variety*, refers to the nature of structured, semi-structured and unstructured data. In fact, the sources from which they are derived are heterogeneous. Organizations manage data not only from internal sources, but also from external sources such as social media.
- *Veracity*, refers to the degree of inconsistency that sometimes characterizes data, which must be of high quality and accurate to ensure optimal analysis. Despite the increasing volume, speed and variety that characterizes them, Big Data cannot ignore an essential prerequisite: quality.

- *Variability*, in fact, means considering how quickly the structure or form of the data is changing. It is important to give the right attention to this aspect, since it is responsible for the interpretation of the data.

After the 5V just described, a sixth V was added. In this case the V represents the initial of the term *Value*. In recent times Big Data are considered as a source of wealth for organizations, in other words a source of value [24]. Often the terms *data*, *information* and *knowledge* are used interchangeably with each other, in reality they are not exactly the same thing. By definition, the term data is an encoded elementary description of information, an entity, a phenomenon, a transaction, an event or something else [25]. Only through the analysis of a datum, information can be obtained. Finally, knowledge is obtained when a person uses information to make decisions [24]. The value is therefore generated in the passage involving these three phases, which are based on a process of analysis and use of data.

1.5.3 Data categories and their analysis

As mentioned before, data can be classified into three categories:

- *Structured data*;
- *Semi-structured data*;
- *Unstructured data*.

As noted by their name, structured data follows a set of rules and can be of various types (number, date, address, etc.). In this case, the data can be represented in a spreadsheet program such as Excel and can be organized using rows and columns. The set of rows and columns form a table and in the presence of multiple tables it is possible to create links between them. Links between columns form relational databases. Databases are processed using SQL (Structured Query Language). Finally, structured data is easy enough to analyze due to its quantitative nature [26]. Next, there is the semi-structured data, which is not suitable for a relational database. They are usually represented with labels, graphs, or tree structures. Examples of semi-structured data are emails or files with HTML extension, used to transmit data between a server and a web application [27]. Unstructured data cannot be organized into rows or columns (e.g., images, audio, video). They do not have a predefined schema; in fact, they are more flexible. Despite the vastness of existing software, the analysis of unstructured data is quite complicated due to its qualitative nature. Since they make up most of the data owned by organizations, they represent a major obstacle. The commonly used database is NoSQL (not only SQL). It is a database with higher data processing capabilities [26].

Basically, given the enormous amount and type of data that companies need to know how to manage in today's market, they can be considered officially immersed in the Big Data era. The increase in data availability has led to an increase in data processing activities. Obviously, Big Data Analytics activities help organizations to realize superior performance and to gain competitive advantage. In reality, it is not sufficient to own technologies that deal with such activities, but it is necessary to integrate BDA within the value chain (Bameland Bamel, 2020). In fact, the technologies can be easily imitated by competitors, who would not be able to duplicate the effectiveness with which a company uses its resources. In addition, BDA offers a great advantage: those who make decisions without consulting data, have only a partial vision of the market. In general, Big Data analysis is divided into two macro areas: qualitative and quantitative [28]. Quantitative analysis mainly uses numerical data and describes a phenomenon in detail, while qualitative analysis identifies in broad terms the methods and causes of development of the phenomenon. There are three types of data analysis: *descriptive* analysis, *predictive* analysis and *prescriptive* analysis.

Descriptive analysis, as its name suggests, describes and synthesizes raw data. It focuses on analyzing past events, from which companies can learn and identify solutions for possible future influences. Descriptive analysis includes simple reports, such as inventory, workflow. These are useful documents to describe the state of a company's operations [28]. However, this type of analysis has a limitation: it does not identify future trends, but simply allows raw data to be translated into useful information [29].

Predictive analytics focuses on understanding future events. In other words, it tells what will happen in the future. It provides intuitive information and requires a large volume of data to do this. It is obvious that what is predicted by this type of analysis does not

have a 100% probability of happening. The results of predictive analytics are intended to support the decision-making process of business managers [28].
Prescriptive analytics supports users in prescribing different actions to drive activities toward a solution. It aims to measure the effect of future decisions in a way that avoids unpleasant outcomes. It is a complex analysis and still not widely adopted by companies [28].

1.5.4 Big Data Application

Following this brief overview of each type of data analysis, it is possible to identify the areas of the supply chain where there are positive results. For example, thanks to data analytics, managers can learn in detail about customer preferences and tastes, thus aiming at product refinement and manufacturing improvement. In addition, Big Data provides a comprehensive overview of the market condition and consequently helps in the identification and evaluation of risks, which can be more easily mitigated in this way. BDA provides useful suggestions for strategies and decision-making processes to be faced. As explained in the previous paragraph, the customer is becoming increasingly important, so often feedback is requested on purchased products. The analysis of this information can reveal useful signals in the functioning of the product. So once again, Big Data enables customer monitoring and facilitates the refinement of the final output.
In summary, thanks to the 6Vs that characterize Big Data, it can positively influence the supply chain in the planning, procurement and development phase, execution phase, and product management phase [30].
The data of the supply chain together with statistical models help to obtain an accurate forecast of the demand and to contrast the phenomenon of the inventories in excess (phase of planning), guaranteeing in this way also economic benefits due to the reduction of the costs of stocking. The availability of the data supports the management in the evaluation and selection of the suppliers (phase of supplying), aiming to products more and more of high quality, gaining consequently the trust and the so-called *willingness to pay* of the customer. In other words, such aspect contributes to customer loyalty. The use of real-time data benefits failure prediction and maintenance planning

(execution phase), leading to a more effective and responsive production process. Multiple benefits are also found in the selection of optimal routes for product delivery, lowering transportation costs. In addition, a higher degree of information provides greater accuracy and security in the product return phase, for example, customer transaction tracking enables the archiving of feedback issued (product management phase). Integrating BDA into supply chains develops an organization's capabilities for information exchange, intra- and inter-company integration, and responsiveness to macro- and micro-company changes in supply chain management [30].

Another effect that can be focused on is the *Bullwhip Effect* (BWE). This effect, which was initially found by Forrester in 1958 through a case study in supply chain management, can be effectively alleviated after the application of digital technologies (DT). Thus, it is inferred that Big Data also alleviates BWE through the improved demand forecasting mentioned above (Ran *et al.*, 2020). It is also found that decision science models based on big data provide optimization for sustainable technologies. We often talk about *Green Data*, but before delving into this concept, it is important to understand what precedes it. In most cases, incomplete and unstructured data is collected, and even not correlated with each other. This represents a major obstacle that organizations face. Starting from this premise, it is possible to talk about Green Data, a term that identifies the need to perform a digital cleansing necessary to initiate digital sustainability. This is a method of data collection that generates valuable and useful information to launch sustainable strategies [31].

There are numerous goals that can be achieved through Big Data Analytics [31], for example:

- Reduction in energy consumption;
- Minimization of carbon dioxide emissions through the study of vehicle travel;
- Ecosystem health monitoring;
- Analysis of customer welfare useful for the improvement of services and quality offered;
- Reduction of water consumption within the production process.

Regarding sustainability, this concept is often expressed in terms of a triple bottom line - financial, social and environmental performance. So, when talking about sustainable supply chains, it should not consider exclusively environmental issues. If the company undertakes data-driven decisions, then this will certainly have consequences for its organizational and strategic capabilities, which in turn will impact sustainability initiatives. In fact, BDA affects several supply chain management issues, such as providing skills and capabilities to employees (Chalmeta *et al.,* 2021). In order to achieve a good level of BD utilization, it would be necessary to provide employees with an adequate level of knowledge regarding this new technology. Such an aspect would help organizations to fulfill social goals. To support a sustainable supply chain, the concept of Big Data and Predictive Analytics (BDPA) needs to be used. These are useful tools to improve business performance. Originally the acronym BDPA was closely related to the financial domain, in reality BDPA can have several impacts on the triple bottom line (Hazen *et al*., 2016). For example, BDPA could be used to study climate change or BDPA initiatives could bring useful financial benefits for sustainable supply chain management. In this regard, each organization could implement a datawarehouse to facilitate the collection of economic data, such as fuel consumption and its associated expenses, costs, revenues, etc. In this way, it would be possible to obtain the financial ratios necessary to improve the level of sustainability of the supply chain. For example, overspending on fuel could make top management think and stimulate them towards more cost-effective and environmental friendly transportation solutions.

Finally, considering the potential advantages that the supply chain could obtain from the use of Big Data and considering the continuous evolution to which today's market is subjected, the study proposes in the following chapters an analysis of the papers, obtained through a particular research technique called *Snowballing*. In order to highlight what are the current conditions of manufacturing companies regarding the use of Big Data and its impact on the supply chain, the main results emerging from the snowballing papers are reported.

2.Research method

This chapter aims to describe in detail each step addressed during the study. In particular, Section2.2 will outline the research methodology applied to the set of articles selected in the first phase of the literature analysis, which represented the starting point. Following this first step, the research methodology that has largely characterized the entire study was carried out and hence led to the results described in Chapter 3. This second part will be analyzed in Section 2.4.

2.1 Systematic Literature Review

This research thesis basically consists of two major steps:

- Searching papers with applications of all digital technologies to the supply chain;
- Searching papers with applications of Big Data to the supply chain.

It is important to specify that they were implemented in the order in which they are listed. In fact, the second step is completely based on the first. This section focuses on the first step, which was performed through a Systematic Literature Review (SLR).
SLR is recognized as a fundamental research methodology to achieve efficient analysis of a large number of different sources (Frehe *et al*., 2014). In fact, several researchers argue that by performing a synthesis of research in a systematic and transparent way, an SLR can be considered high-quality evidence. Therefore, on the one hand the review process increases methodological rigor for academics, and on the other hand it supports practitioners in developing a solid knowledge base through the accumulated knowledge from a series of studies (Meriton *et al*., 2021).
For such reasons, this methodology was considered useful in achieving the objectives of the present study.

Typically, a Systematic Literature Review consists of various steps [33]:

1. Definition of the team of work;
2. Development of the research protocol;
3. Data collection;
4. Database selection;
5. Selection of studies;
6. Assessment of bias;
7. Evaluation of unpublished studies;
8. Summary of results;
9. Review;
10. Update.

They represent the entire process that needs to be carried out in order to obtain the desired results. In reality, the real research activity is between step 3 and step 10, which can be summarized in four macro phases represented in figure 3.

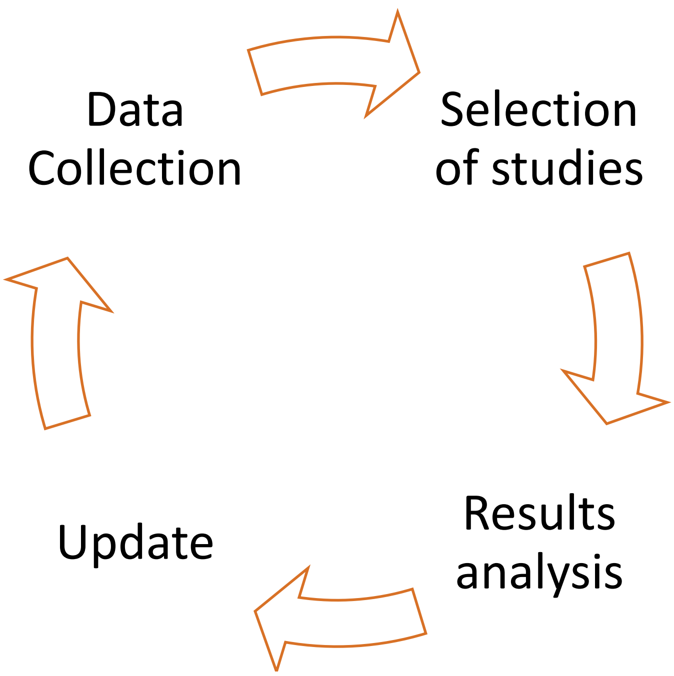

Figure 3 - Macro phases of Systematic Literature Review

A brief description of each step follows.

The process starts with the definition of the working team, in fact, the presence of a research group allows to compare different points of view, thus bringing out any criticalities of the working method. It is not by chance that this method requires a well-defined research protocol, able to clarify to each team member the research methods. Like all research methodologies, the Systematic Literature Review produces data. It is important to have the objectives of the study in mind in order to select articles that are relevant to the research area. It is possible to extrapolate these articles from a variety of databases, so the databases to be used during the research should be defined. It is essential that the articles collected are also of high quality to obtain a high-quality study. The researcher must carefully select the articles, for example, by reading the abstracts or identifying appropriate keywords. Regarding the sixth step, the term bias properly means "oblique" or "slanted" [34]. In statistical terms, it identifies the tendency to deviate from the average value. Thus, there are publications in which such bias cannot be eliminated, so this aspect definitely limits the results. In addition, these limitations may also be due to the presence of unpublished studies. At the end of these considerations made about the results, it is necessary to make a synthesis of them. The term *synthesis* doesn't simply refer to the elaboration of a summary of all the articles, on the contrary it is appropriate to perform a critical analysis of each of them so as to produce a not banal discussion, on the contrary, organic and well articulated. Finally, as in all research methodologies, it is appropriate to perform a review and update. This last phase is important in order to avoid the exclusion of articles relevant to the research and that have been published only recently.

2.2 Systematic Literature Review application

The discussion of the steps that typically characterize an SLR, performed in Section 2.1, is followed by a description of the application of those steps to the research process in question.

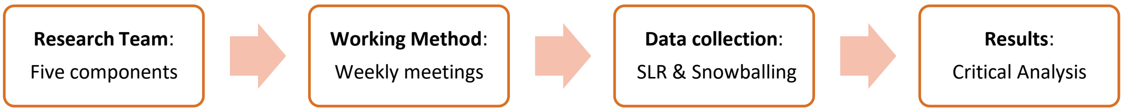

Figure 4 - Steps of SLR

As it can be noted in Figure 4, first of all, the research team was defined, composed by three students from Politecnico di Torino and teachers Anna Corinna Cagliano and Mahsa Mahdavisharif. Then, the objectives and the working methods were identified: each student was associated to specific Industry 4.0 technologies and through weekly meetings the team was constantly updated on the progress of the research. Regarding data collection, however, it was performed in two phases, listed at the beginning of Section 2.1. In this first phase, two databases were selected, namely Scopus and Google Scholar. Considering that the main objective of this first step was to identify articles having as topic the applications of Industry 4.0 digital technologies to the supply chain, the following keywords were used for the selection of articles:

- "*additive Manufacturing*" and "*supply chain*";
- "*Big data*" and "*supply chain*";
- "*Blockchain*" and "*supply chain*";
- "*Block chain*" and "*supply chain*"
- "*cloud computing*" and "*supply chain*";
- "*Digital supply chain*";
- "*Digital supply Chain*", "*lean*";
- "*digitalization*" and "*supply chain*";
- "*IoT*" and "*supply chain*".

A total of 338 papers were extrapolated. To them has been associated the name of *Basic Paper* (BP). In order to execute best the steps ranging from number 6 to number 10, the Basic Papers were classified through the definition of the following points characterizing each article:

- Authors;
- Title;
- Year of publication;
- Reading date;
- Journal;
- Impact Factor;
- Quartile (Q);
- Citations number;
- Database;
- Keywords used to find the paper;
- Information regarding the topic discussed.

Concerning the last point, that is, the information regarding the treated topic, details were reported regarding the topic area, the technology and methodology adopted, the data collection used by the author, the research questions (RQ), the research gaps, the main results, the limitations and also the future perspectives. Finally, the critical points derived from each article, a small summary and an evaluation regarding the impact of the topics discussed in the Basic Paper on the research objectives were also noted. This evaluation was assigned through a score contained in a range from 1 to 10.
This first phase of research provided a starting point for the second phase of research, which the object of the present thesis. In fact, the first phase was exclusively performed by the thesis supervisors Anna Corinna Cagliano and Mahsa Mahdavisharif. Moreover, since the second research phase is exclusively focused on Big Data technology, the 338 Basic Papers were reduced to 105. 105 is therefore the number of BPs concerning Big Data.

2.3 Snowballing

As anticipated in Chapter 1, the concept of Digital Supply Chain has become more emphasized in recent years, so there are still numerous literature gaps to be filled, scattered articles, and lack of a collection of publications from which the main critical issues emerge (Zekhnini *et al.*, 2021). A literature review helps researchers to analyze the literature that is considered interesting, to recognize the conceptual content, and to contribute to the development of publications. To achieve the objectives of the study, the review methodology is based on a particular research method, called *Snowballing.* As its name suggests, this technique is based on so-called *snowball sampling*. This method is part of the SLR research methodology described extensively in Section 2.1. It is a non-probabilistic method that consists of several phases [32]. The first phase sees the researcher dealing with a sample of previously selected articles, after which, this sample is destined to increase in size with the continuation of the research. For this reason, the sample is said to grow over time like a rolling snowball.

Often, this technique is used to undertake research in the social field, so it is a different application than the focus identified in this research. In fact, the commonality between these two different areas of application of snowballing is the reason that drives the researcher to opt for this method. In fact, this sampling technique is primarily used by social scientists who want to work with a population that is difficult to locate or identify [32]. In the case of this study, on the other hand, rather than talking about a population of individuals, it is a set of publications that are useful in filling literary gaps and describing results that are still not widespread due to only recent attention given to this technology and the impact or influence that Big Data can have on the supply chain. It is deduced that snowball sampling is an excellent technique to be able to process exploratory research.

This research method can be classified into two categories (Jalali and Wohlin,2012):

- *Backward* snowballing, that is the identification of papers through the list of references included in each starting article;
- *Forward* snowballing, identification of papers that cited the articles in the starting sample.

Thus, Snowballing represents a manual process based on a series of papers or publications to identify multiple studies on the basis of the references and citations contained in them.
If the researcher follows both snowballing methods, backward and forward, snowball sampling primarily consists of three steps that are iterated until satisfactory results are found. The three steps are as follows (Jalali and Wohlin,2012):

- Step 1: Start searches of major databases to obtain a set of source articles. This is the phase described in Section 2.2.
- Step 2: Go backward (backward snowballing), into the source articles identified in Step 1, to analyze the reference articles considered relevant. This step is iterated until new documents are identified.
- Step 3: Go forward (forward snowballing), to identify and analyze articles cited from the source set (step 1).

As noted, the time required to be able to perform this procedure is quite high.
Another potential disadvantage could be given by the finding of articles published by the same authors and this would cause a distortion of the results if there is a high number of repeated authors. In this case the search would be restricted to specific authors only (Jalali and Wohlin,2012).
Despite these first major limitations, the snowballing method also offers some advantages. Indeed, it could be considered as a complementary method for a systematic literature review. Moreover, unlike a traditional SLR, snowballing sampling does not explicitly require the use of a multiplicity of databases. The snowball approach is rather simple and easy to understand (Jalali and Wohlin,2012).

2.4 Snowballing application

In the present research, from a set consisting of 105 articles, mentioned in Section 2.2, seventy-six snowballing papers were found.

Also in this case, in order to identify and analyze the widest possible scope of the academic productions, two databases were used: Scopus and Google Scholar. All articles not viewable through the abovementioned databases were declared inaccessible. In fact, only those articles deemed relevant to the research objectives were analyzed using the Snowballing technique, namely the study aims to analyze and bring out critical findings related to the impact of Big Data on the supply chain.

To address the research purpose, the review methodology was based on the following content analysis approach. This study uses both methods, backward and forward, of snowball sampling. In both cases, various keywords found in the title or abstracts were used to select articles considered relevant to the research:

- "*Big Data*";
- "*Big data analytics*";
- "*Supply Chain*";
- "*Supply Chain 4.0*";
- "*Digital Supply Chain*";
- "*Logistics*";
- "*Supply chain management*";
- "*Digitalization of supply chain*";
- "*Data-driven*";
- "*Data analytics*"
- "*Industry 4.0*".

In addition, the selection wasn't only based on keyword research, but also on the description of the article content present in the abstract of each of them. A further criterion for the exclusion of articles concerns the year of publication, in fact all articles published in a period prior (2010) to the emergence of Industry 4.0, don't represent a

source of interest. The same is true for all publications that are results of conferences or congresses. Finally, as it is easy to presume, all articles that didn't have Big Data, digitization or supply chain management as their main topic were also discarded.
The selected articles, or snowballing papers, were defined in the same way as previously described in Section 2.2. In fact, also in this case, information regarding the following topics was reported for each snowballing paper:

- Authors;
- Title;
- Year of publication;
- Reading date;
- Journal;
- Impact Factor;
- Quartile (Q);
- Citations number;
- Database;
- Keywords used to find the paper;
- Information regarding the topic discussed.

Following this first stage of selection, articles considered relevant were analyzed from a content perspective. To facilitate the tracking of this analysis and more easily obtain quantitative and qualitative results, the articles were categorized into tables, depending on their content.
In fact, this classification has been carried out according to the type of supply chain activities argued, the supply chain areas described, the research methodology followed in each article, the theories used by the authors, and according to the type of technologies discussed simultaneously in the same article . The dimensions of analysis were defined during the first phase of the SLR, that is the one previously conducted by the thesis supervisors Anna Corinna Cagliano and Mahsa Mahdavisharif.
In particular, it can be shown which categories characterize the main macro areas.

The table regarding supply chain activities presents the following categories:

- Transportation;
- Inventory Management;
- Manufacturing;
- Procurement;
- Warehousing;
- Demand forecasting, marketing and pricing;
- Customer relationship;
- Overall supply chain/logistic.

The categories that constitute supply chain areas, on the other hand, are:

- Information sharing;
- Revenue/cost sharing;
- Risk;
- Performance assessment;
- Sustainability;
- Financial;
- Scheduling;
- Business model.

The possible research methodologies used in each snowballing paper are as follows:

- Optimization/mathematical modeling;
- Simulation;
- System Dynamics/ Agent Based modeling;
- Survey/case study;
- Literature Review;
- Conceptual/Framework;
- Hypothesis test;

- Model/Architecture development;
- Development of Tool/Platform/Computer system;
- MCDM methods.

Possible technologies combined with Big Data are:

- Augmented Reality;
- Cloud Computing;
- Robotics;
- Sensor Technologies;
- Internet of Things;
- Self Driving Vehicles;
- Unmanned Aerial Vehicles;
- 3D Printing;
- Blockchain;
- Artificial Intelligence and Machine learning.

Finally, as explained above, the theories used in the snowballing papers have also been traced.

The results derived from this executed research method are illustrated and discussed in the Chapters 3.

3.Results and Discussions

This chapter highlights the main results resulting from the application of the research methodology described in Chapter 2. In particular, it doesn't aim exclusively to list the main points that emerged, but, on the contrary, the results are analyzed from multiple perspectives. Thus, it is a true critical analysis, highlighting which main aspects of the supply chain are affected by Big Data technology. In addition, the critical analysis is amply supported by a thorough discussion of the results. This gives the reader ample opportunity to think on the causes and consequences of the application of Big Data in the supply chain environment.

3.1 Snowballing Results

As anticipated in Section 2.4, through careful categorization of articles considered important from the point of view of results, it was possible to subject these articles to critical analysis. The evaluation criteria used can be classified into two distinct categories:

- Quantitative evaluation criteria;
- Qualitative evaluation criteria.

The first criteria presuppose an extrapolation of quantitative data concerning, for example, the number of citations, the number of articles in a given category, etc. Thanks to this first extrapolation of quantitative data, it is possible to study and analyze the possible relationships or correlations present among the various data. This second scenario is referred to precisely as qualitative evaluation criteria.

All articles that presented in their results an interesting relationship present between the analyzed technology, Big Data, and the Supply Chain, were defined as *Snowballing Papers*, as already discussed in Section 2.4. Before going into detail, from the general results regarding these articles, it can be noted that at the end of the literature review, addressed by the two search modes, that is, backward snowballing and forward snowballing, seventy-six snowballing papers were found.

With more precision, it can be seen, as shown in Figure 5, that only twenty percent of snowballing papers were found through backward snowballing, while the remaining eighty percent came from the searching mode called forward snowballing.

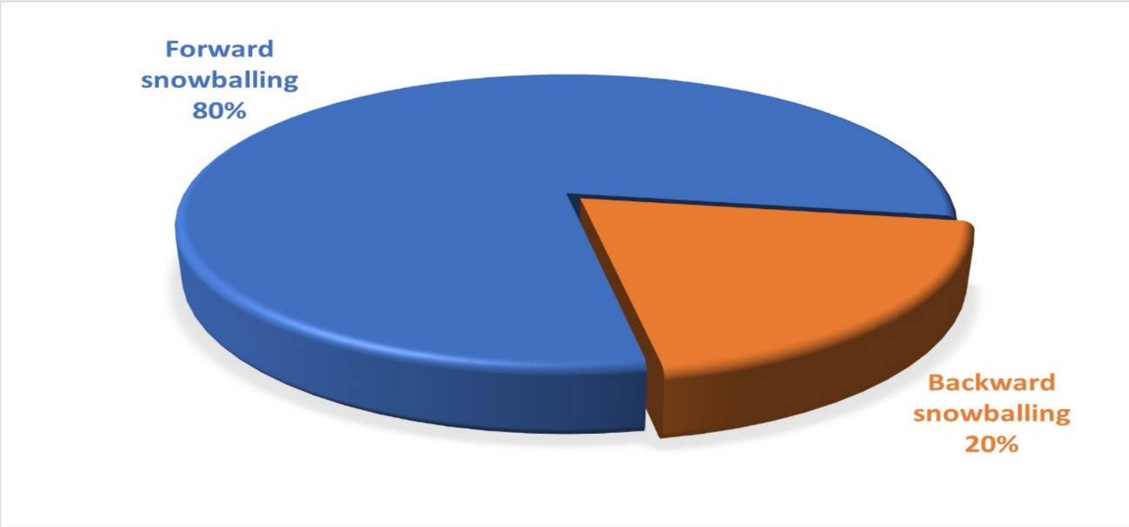

Figure 5 - Percentages of snowballing papers found according to the two searching modes

A further point of analysis is provided by considering the number of starting papers that were the source of origin for the new snowballing papers.

Indeed, as shown in Figure 6, fifty-three percent of the one hundred and five starting basic papers provided snowballing papers of interest for the purpose of this research, while the remaining forty-seven percent provided no papers of potential interest. These percentages suggest a new research trend toward this new topic. Through the observation of these initial results, therefore, it is inferred that Big Data and its application in the supply chain environment can be considered a topic of interest.

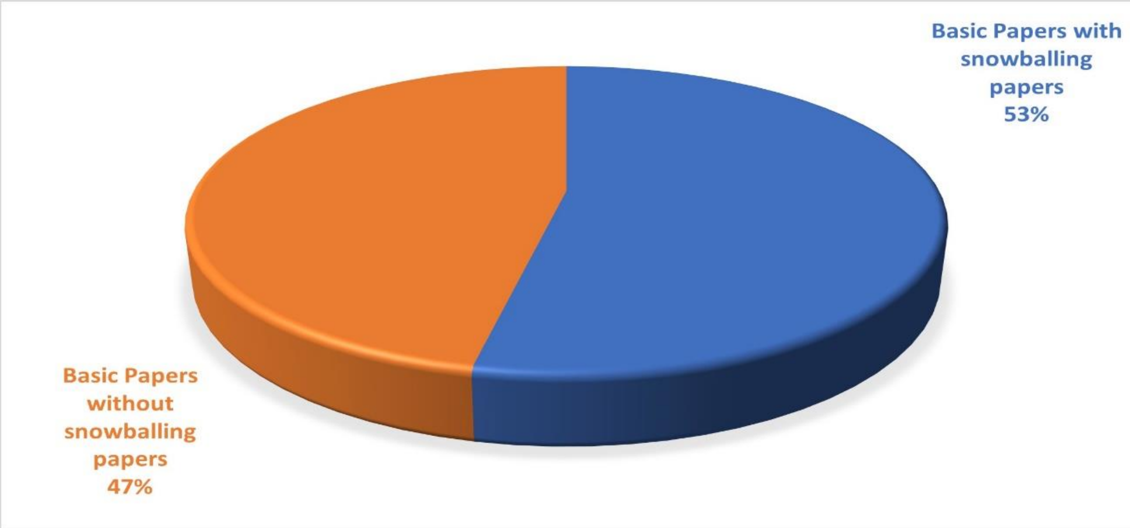

Figure 6 - Percentage of basic paper with or without snowballing papers

As described in Section 2.4, snowballing papers were classified according to the point of view analyzed, namely, according to the type of supply chain activity, according to supply chain areas, according to the research methodology followed in each paper, and according to the type of technologies discussed simultaneously in the same paper or theories used by the authors. This procedure then allowed the extrapolation of quantitative data, which was useful in answering the main objectives of this bibliometric analysis.

The main question, which this research study attempts to answer, is:

- What effects does the use of Big Data produce in the supply chain environment?

In reality, in order to answer this question, the critical analysis addressed in this research thesis attempted to capture many other aspects, answering, for example, the following questions:

- What are the main supply chain areas affected by the use of Big Data?
- What are the main supply chain activities influenced by the use of Big Data?
- Which research methods have led to greater results?
- Does Big Data need the support of other technologies in order to produce positive supply chain results?

These research questions are addressed in the following sections.

3.1.1 Bibliometric indexes and three analysis perspectives

It is possible to answer the research questions listed in Section 3.1 and many other questions through the analysis of the various bibliometric indices, shown in Figure 7.

PROSPETTO RIASSUNTIVO DEI PRINCIPALI INDICATORI BIBLIOMETRICI IN USO					
FONTE	**PRODUTTORE**	**INDICATORI**			
		(a livello di articolo)	***(a livello di rivista)***	***(a livello di autore)***	***(a livello di nazione)***
Web of Science	Clarivate	*- Citation index ("Times Cited")*	*- Impact Factor (IF) - Immediacy Index - Cited Half-Life - Eigenfactor - Article influence score*	- H-Index	
Scopus	Elsevier	*- Citation index ("Cited by") - Altmetrics*	*- Citescore - H-Index [rivista] - SCImago Journal Rank (SJR) - Source Normalized Impact per Paper (SNIP)*	- H-Index	- Scimago Country Rank (SJR)
Google Scholar	Google	*- Citation index ("Cited by")*	*- H5-Index [rivista]*	- H-Index	
iCite	NIH	*- Relative citation ratio (RCR)*			
Altmetrics	Altmetric	*- Altmetric attention score*			

Figure 7 - Summary of key bibliometric indices [35]

Bibliometric indicators are used within the scientific community to be able to precisely measure and analyze scientific publications and their impact [36]. As shown in Figure 7, bibliometric indicators measure the impact of individual authors, journals or articles. For this reason, the critical analysis regarding snowballing papers was performed considering these three different perspectives. In reality, as can be deduced from the sections into which this research work is divided, only two perspectives were used in order to approach the critical analysis, namely article and journal. This choice is mainly related to the absence of relevant results in the author's perspective. In addition, they differ depending on the reference database. These indices are a great support for obtaining quality results, as they not only lead to quantitative assessments, such as the number of citations received by an article, but also to qualitative assessments, such as the Impact Factor associated with a journal.

3.2 First analysis perspective: the article

This section addresses the first of the three perspectives, announced in Section 3.1.1, namely the article. The critical analysis therefore focuses primarily on all aspects closely related to the seventy-six snowballing papers, analyzing their reasons for rejection, the supply chain areas and activities argued, the methodologies and theoretical lenses used, the year of publication, and the combined use of the various technologies found in each paper.

3.2.1 Critical analysis of rejected papers

Before focusing exclusively on the articles that were analyzed, it is appropriate to examine the articles that were rejected so that any limitations arising from the existing literature on the study of Big Data related to the supply chain can be highlighted. A total of 15709 articles were rejected, and the reasons for this rejection, previously discussed in Section 2.4, are shown in Figure 8.
Considering that each article may have been rejected for more than one reason among those shown in the figure, it can be seen that in each case the predominant factor among the various exclusion criteria is due to the lack of focus in the articles on Big Data or the lack of connection between Big Data and supply chain. Next, most of the 15709 articles were rejected due to the lack of treatment regarding supply chain management. In addition to these first two main reasons, there are two other criteria that led to the exclusion of other articles. In detail, fourteen percent of the articles were rejected because they were published in a period (2010) prior to the emergence of Industry 4.0, and eleven percent of the articles were defined as duplicate, i.e., they were discarded because they were already present among the basic papers and therefore would not have led to new results but only caused research bias. Only a small number of the studies in the literature, i.e., two percent, don't concern applications related to the

manufacturing sector; in particular, a large proportion of these articles in fact concern the food and humanitarian sectors. Finally, the remaining articles were rejected because they were inaccessible or the result of conferences. Thus, it can be deduced that although the application of Big Data in supply chain is a trending topic, the literature still has some limitations in this regard.

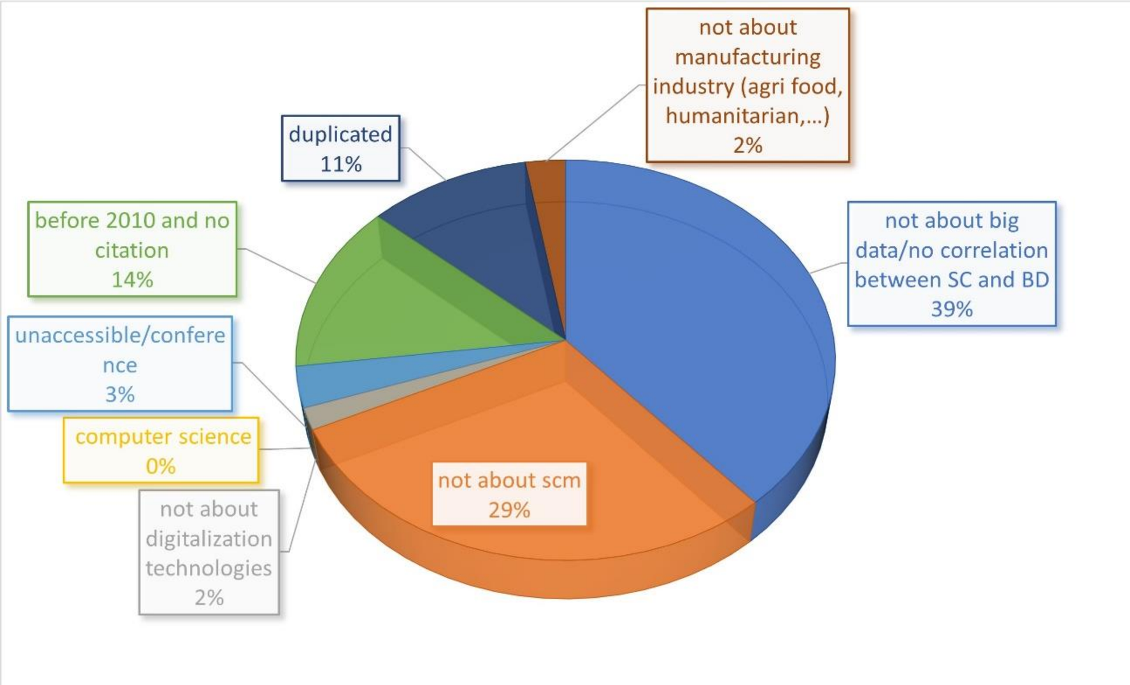

Figure 8 - Number of rejected snowballing papers

3.2.2 Critical analysis of three categories: SC activities, SC areas, methodologies

Further insight, regarding the perspective of the article, can be provided by analyzing the number of snowballing papers present in each of the activities or supply chain areas argued, or in each methodology used. Not surprisingly, as explained in Section 2.4, the snowballing papers were classified into three tables each corresponding to a category, namely: supply chain activities, supply chain areas, and methodologies.

Regarding the supply chain activities addressed in the various snowballing papers, as noted in Figure 9, in general, most of the papers addressed the application of Big Data on the supply chain without focusing on processes in particular. Instead, going into more specifics, there are two main activities that are affected by the use of Big Data. They are respectively named: *Customer relationship* and *Demand forecasting, marketing and pricing*. These first two activities are followed by *Procurement*, which has only six fewer items than the other two activities mentioned above. It is also important to note that *Manufacturing* is the only supply chain activity that didn't receive attention in any of the snowballing papers analyzed.

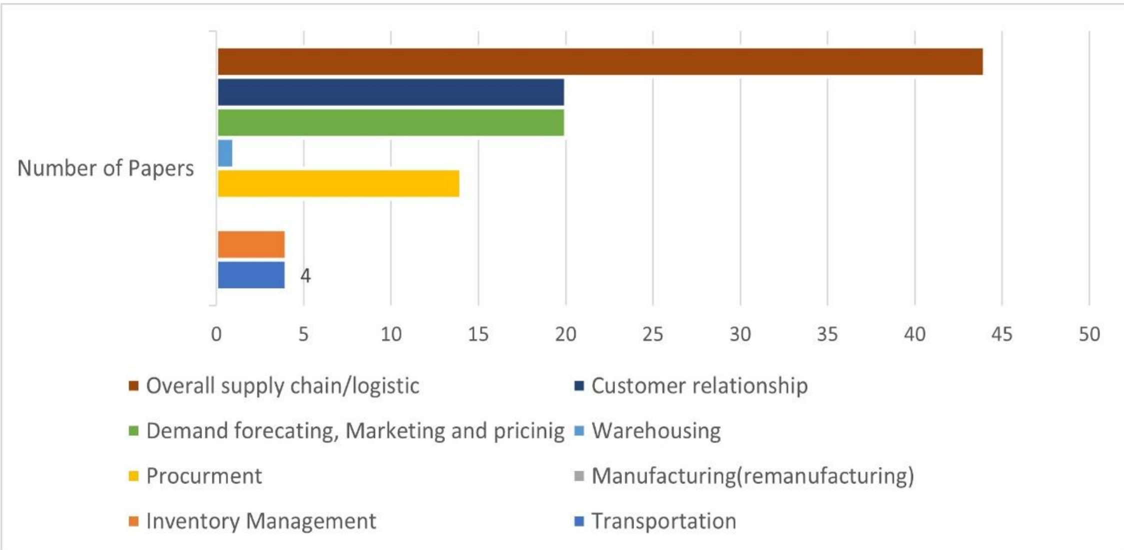

Figure 9 - Number of snowballing papers related to supply chain activities

Regarding the supply chain areas, it can be seen from Figure 10 that the areas with the highest number of snowballing papers are *Perfomance assessment* and *Information* *sharing.* The areas that next have a smaller number of snowballing papers than the previous ones, but still relevant, are *Sustainabilty* and *Risk*. Finally, the only area that has no snowballing papers instead is *Scheduling.*

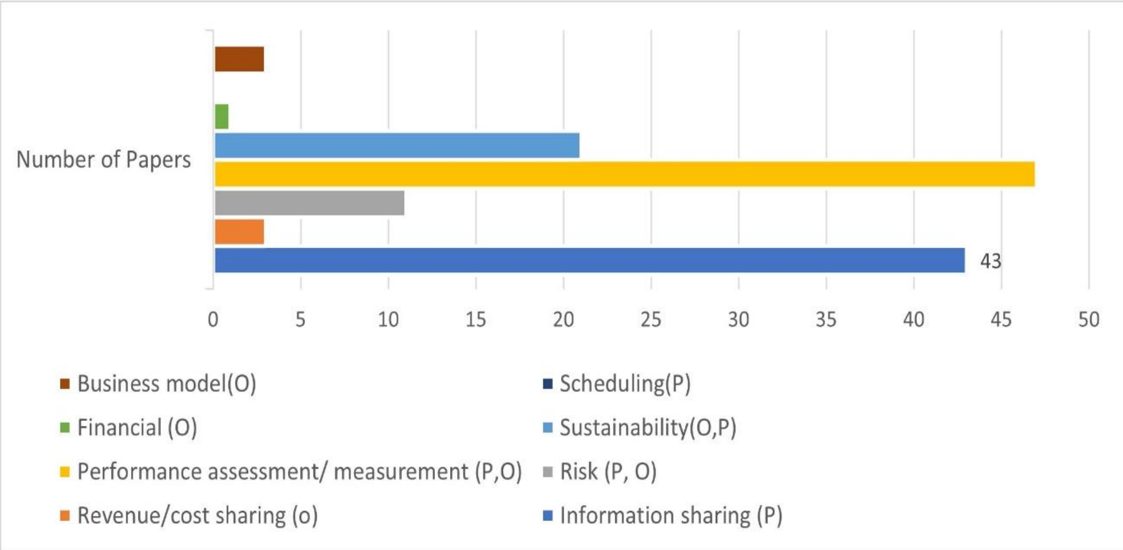

Figure 10 - Number of snowballing papers related to supply chain areas

From this first step, it is clear in what macro activity or macro area, Big Data can influence supply chains. In addition to providing a general benefit to the entire supply chain, with the constant increase in the volume and complexity of data that characterizes supply chains, it is easy to infer how among the main activities to which Big Data can lend support emerges demand forecasting, customer relationship and the procurement process. Data computerization, ensured through the use of Big Data, facilitates data sharing, making every actor in the supply chain participate. Such sharing enables the study of consumer behavior and choices. In addition, all this facilitates the observation of demand trends, ensuring its good forecasting. Big Data analysis supports, for example, the evaluation of various suppliers and consequently improves the procurement phase. Overall, these aspects ensure that the supply chain is well managed, leading to its good performance.

Regarding the activity called *Customer relationship*, the importance of which was discussed extensively in Chapter 1, there should be no doubt about the result that emerged from Figure 9. In fact, it is now certain that the customer is a priority for the entire supply chain and is also considered an external driver that drives companies to adopt digital technologies (Yang *et al*., 2021). Due to the continuous growth in demand for digitized products and processes, companies are forced to adapt to this need in order to meet customer requirements. The latter aspect explains the reason for that a large part of the items present in the *Demand forecasting, marketing and pricing* activity is also present in *Customer relationship*. Infact, these two activities are closely related. Big Data procures necessary information for companies regarding demand processes, so a supply chain based on this new technology can more accurately forecast demand, reduce shipment delays, and ensure effective supply chain operation in organizational terms (Thekkoote, 2021). A further result that simply confirms expectations concerns the large number of snowballing papers in the *Information sharing* area. Indeed, it is easy to imagine how the Big Data that characterizes today's supply chains is a source of a large amount of information and how sharing this information can reduce information asymmetry and consequently also the transaction cost supported by companies. In contrast, an unexpected result is the presence of only one item within the activity called *Warehousing.* With the advent of technology, the volume of data and computing devices within this activity has certainly increased, just think for example of the barcodes associated with the various products or robots that support warehousing activities. In fact, because of such a high volume of data, one would expect a large literature base, but in reality, this didn’t emerge.

Regarding the methodologies used by the authors in the snowballing papers, the results can be seen in Figure 11. It can be seen that the most frequently used methodologies are *Survey/case studies*. This is followed by *Conceptual/framework*, *Literature reviews* and *Hypothesis tests*. With far fewer snowballing papers than the methodologies just mentioned, there are O*ptimization/mathematical model, MCDM methods* and *Simulation.* Focusing on the most widely used methodologies, it is possible to describe some of their limitations. For example, *Survey/case studies* might lead to results that are not always valid because they are closely related to the respondents. The same goes for *Conceptual/framework* or *Hypothesis tests*, which might turn out to be related exclusively to the situation or data analyzed. Finally, Literature review could be limiting in terms of temporal validity, as over time the literature could expand and lead to new findings and evidence.

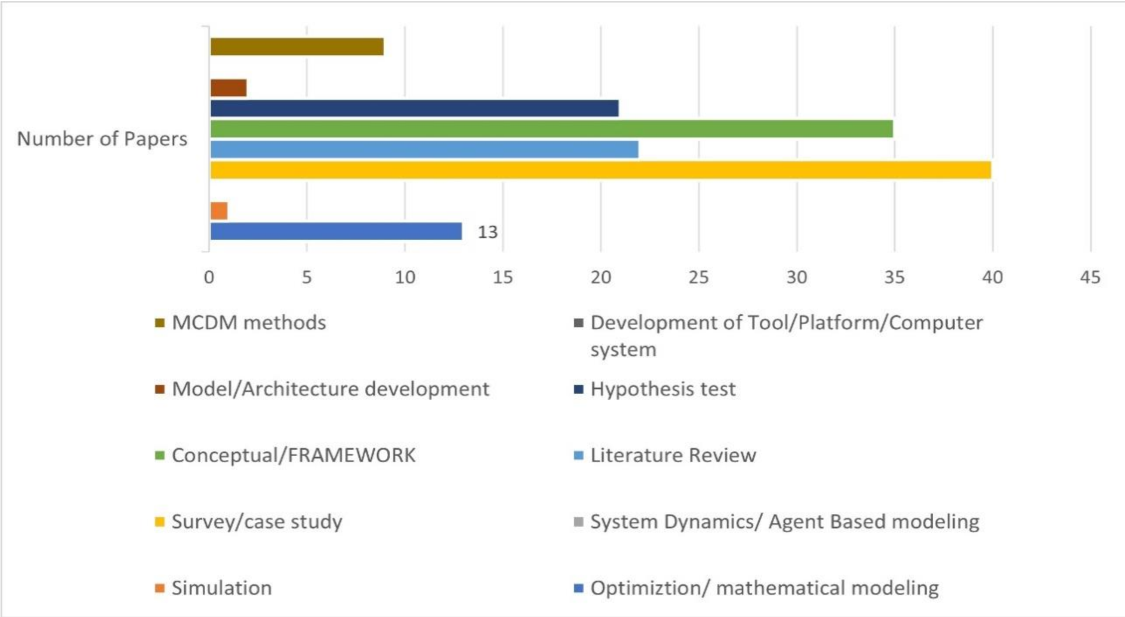

Figure 11 - Number of snowballing papers related to methodologies

3.2.3 The comparison of the three categories

In order to achieve a robust and well-articulated critical analysis, in addition to analyzing the results from the separate study of the three categories, as done in Section 3.2.2, an elaboration of the results regarding the combination of the three categories was also addressed. The goal of this comparison is to find any relationships that exist between:

- Supply chain activities and supply chain areas;
- Supply chain activities and methodology;
- Supply chain areas and methodology.

Figure 12 shows a greater concentration of supply chain areas at three activities: *Overall supply chain, Customer relationship and Demand forecasting, marketing and pricing*. An additional activity on which supply chain areas are predominantly concentrated is *Procurement.* These initial results confirm what is anticipated in Section 3.2.2. In fact, through this graph it is possible to understand not only which activities are impacted by Big Data, but it also provides insight into which areas of the supply chain materialize this influence. Furthermore, while *Transportation* and *Inventory Management* show a small concentration of supply chain areas, *Manufacturing* does not report any snowballing papers regardless of the supply chain area considered. This lack is due to the total absence of snowballing papers in this activity as shown in Figure 9. The four most relevant activities share the following characteristic: two areas prevail in each of them, namely *Information sharing* and *Performance assessment*. This second result also confirms what emerged from the analysis addressed in Section 3.2.2. This aspect can be further confirmed through the content analysis of the snowballing papers.

Each activity in the supply chain causes the movement of goods, starting from raw materials to finished products supplied to customers. This also necessarily involves a transfer of information associated with the goods, which is why it is important to provide for the integration of such information (Jayender and Kundu, 2021). Information is very useful for the collaboration of the various partners in the supply chain, and proper integration of this information enables early prediction of consumer preferences. Nowadays, such preferences change very rapidly, so *Information sharing* also ensures the achievement of customer satisfaction. What has been discussed so far explains why a majority of the snowballing papers that present a correlation between *Demand forecasting, marketing and pricing* and *Information sharing*, consequently show a connection between the area just mentioned and *Customer relationship*. The aspects addressed so far and influenced by Big Data, have a consequence on supply chain performance. In fact, it mainly depends on supply chain management.

Looking at Figure 12 in more detail, it shows that the supply chain areas analyzed so far are followed by *Sustainability*. This area prevails mainly in Overall supply chain. In fact, this topic is not necessarily related to one supply chain activity, but can cover different fields in the supply chain. Finally, the area that has an albeit small relevance is *Risk*. It is related exclusively to the four most important activities examined earlier. In this regard, reconnecting with the importance of information sharing, it is possible to point out that if this information is not carefully integrated and organized, decisions may involve risks. Therefore, Big data analysis can help remove risks (Jayender and Kundu, 2021).

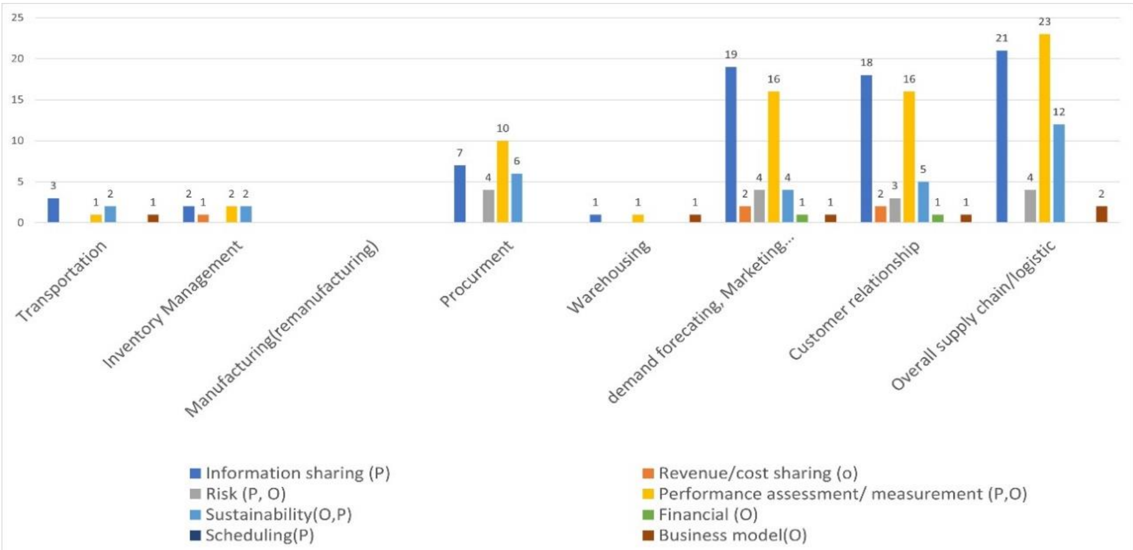

Figure 12 - Number of snowballing papers obtained from the intersection of activities and supply chain areas

Further quantitative and qualitative results can be obtained through analysis of the relationship present between supply chain activities and methodology, shown in Figure 13. The methodologies that are characterized by more supply chain activities are: *Conceptual/framework, Survey/case studies* and *Literature Review*. It represents a confirmation in relation to what is argued in Section 3.2.2. These three methodologies, besides being widely used in the activity called Overall supply chain, are also widely adopted with *Demand forecasting, marketing and pricing*, and *Customer relationship*. In addition, all three methodologies register a small relevance in *Procurement*.

The methodologies that, in contrast, were used less frequently are *Hypothesis Testing, Optimization/Mathematical modeling* and *MCDC methods*. In all three, no supply chain activity stands out. Interestingly, none of the methodologies in the Figure 13 legend have a connection to all supply chain activities. *System Dynamics/Agent Based modeling* and *Development of Tool/Platform/Computer system* have no connection with supply chain activities given the lack of articles with these methodologies. In conclusion, it is noted that the only article using *Simulation* methodology reports a link *to Demand forecasting, marketing and pricing*, and *Customer relationship*; while the article using *Model/Architecture development* is about supply chain in general.

Figure 13 - Number of snowballing papers obtained from the intersection of supply chain activities and methodologies

Are there any relationships between the methodologies usually used by the authors and the supply chain areas? In order to answer this question, Figure 14 can be observed.

Also in this case, the methodologies that show a higher concentration of supply chain areas are: *Conceptual/framework, Survey/case studies, and Literature Review*. Once again, the areas where these methodologies were most frequently used are *Performance assessment* and *Information sharing*. Following these first two prevalences is the *Sustainability* area. The methodologies that, instead, were used less frequently are *Hypothesis Test*, *Optimization/Mathematical modeling* and *MCDC methods*. In all three, the area called *Performance* stands out in particular. In addition, none of the methodologies in the legend in Figure 14 has a link to all areas of the supply chain. As explained in the previous lines, *System Dynamics/Agent Based modeling* and *Development of Tool/Platform/Computer system* have no connection with the supply chain areas given the lack of articles with these methodologies. In conclusion, it can be seen that the only article using the *Simulation* methodology reports a connection to *Information sharing* and *Revenue/cost sharing*; while the articles using *Model/Architecture development* are about *Information sharing* and *Risk.*

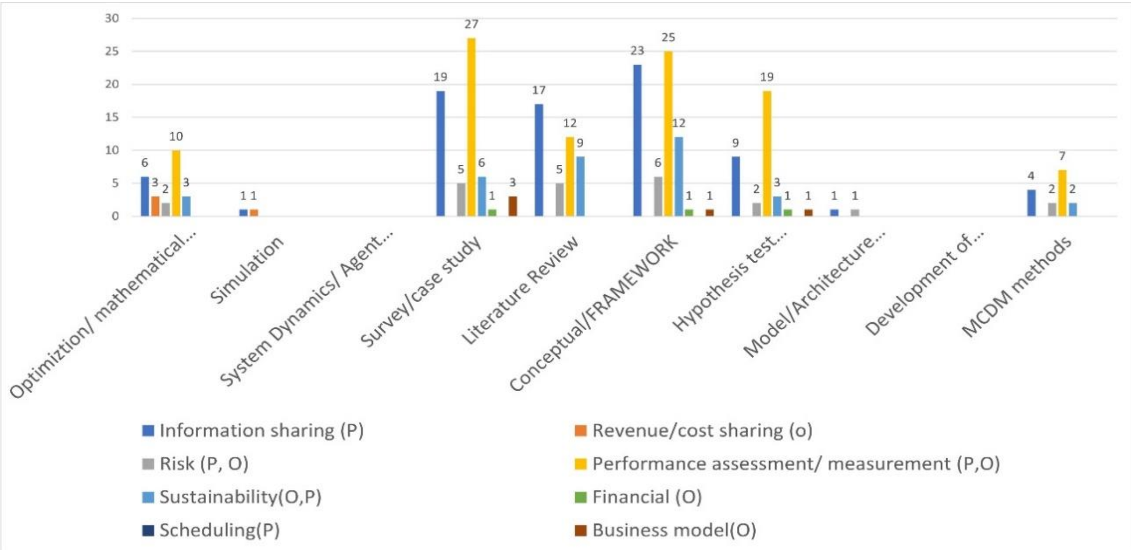

Figure 14 - Number of snowballing papers obtained from the intersection of supply chain areas and methodologies

3.2.4 The theoretical perspective and its critical analysis

Another aspect traced through the categorization of snowballing papers concerns theoretical lenses. In total, there are twenty-seven theoretical lenses identified, and in addition to these there is the so-called *No theoretical lens*, i.e., this category refers to all those papers that didn't present any theoretical lens.

Most of the twenty-seven identified theoretical lenses have only one snowballing papers. For this reason, in order to be able to obtain more detailed results, only those theories used more than once were considered, thus reducing the twenty-seven theoretical lenses to only nine theoretical lenses. Premising that some snowballing papers have multiple theoretical lenses, each theory is characterized by a certain "relevance" expressed in percentage terms depending on the number of papers in which it was adopted.

As shown in Figure 15, just over fifty percent of the authors didn't use theoretical lenses in order to argue the impact of Big Data on the supply chain.

Focusing instead on snowballing papers using theoretical lenses, it is noted that the authors' most favored theory is *Resource Based View* (RBV) with a relevance of fourteen percent. Following with a difference of only three percentage points there is *Dynamic Capability*. Among the least relevant are *Information Processing Theory* (IPT), *Game Theory, Technology, Organization* and *Environment Framework* (TOE), and *Lean strategy*. Finally, the theoretical lenses characterized by an absolute low percentage are *Stakeholder Theory, Absorptive Capacity*, and *Resource Dependence Theory*. In fact, the latter three theoretical lenses have a relevance of only two percent.

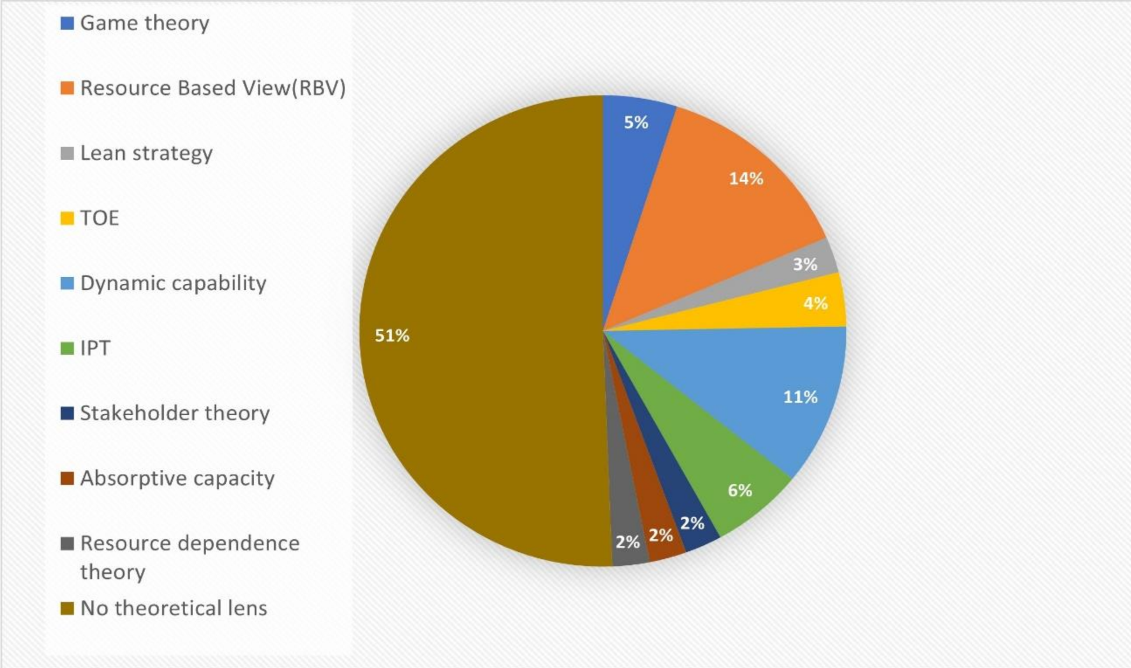

Figure 15 - Percentage of snowballing papers related to theories used more once

To confirm what has already emerged in Figure 15, there is the percentage of the number of authors who didn't use theoretical lenses equal to just over fifty percent.

In detail, focusing on the authors who represent the minority, that is, the authors who used more than one theory, it is interesting to note that on average they adopted about 3 theories each.

In order to make the critical analysis of the theoretical lenses further interesting, it is possible to go in search of any links present between the theories and supply chain activities. However, this represents only one starting point around which considerations can be developed.

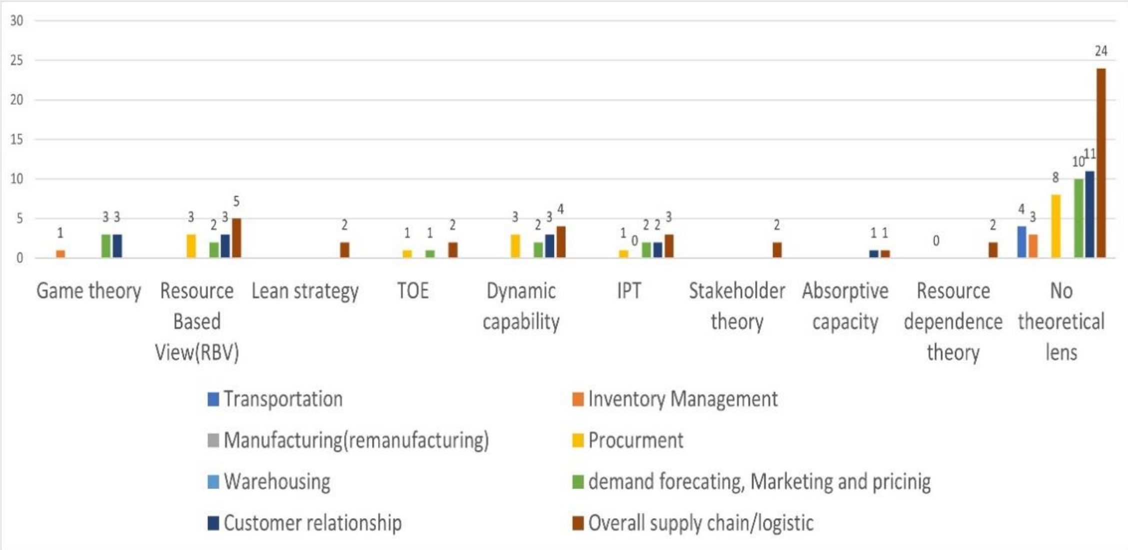

Figure 16 - Number of snowballing papers obtained from the intersection of supply chain activities and theories

As can be seen from Figure 16, the activity known as *Overall supply chain* is present in almost all the theories analyzed, this is because perhaps the lack of focus of the corresponding articles on a specific aspect of the supply chain makes them more compatible with the different theoretical lenses.

In addition, the theoretical lenses with a greater variety of supply chain activities are those characterized by greater relevance: *Resource Based View, Dynamic Capability*, and *Information Processing Theory*. To be precise, this last theory is characterized by lower relevance than the first two, but it is still significant compared to the remaining theories. The supply chain activities included in all three theories just mentioned are *Demand forecasting, marketing and pricing, Customer relationship* and *Procurement*. Regarding the *No theoretical lens* category, on the other hand, there is a strong predominance of the activity called *Overall supply chain*, and this is followed by the activities already mentioned with the three most relevant theories.

To better understand what incentivized the authors to favor a particular theory, it is necessary to search for any connections present between the theoretical lenses and the various areas of the supply chain. These results can be analyzed through Figure 17. It not only again confirms the three theoretical lenses considered relevant already in Figure 16, namely *Resource Based View, Dynamic Capability* and *Information Processing Theory*, but also highlights the presence of an additional theory with wide variety of supply chain areas. This is the theory called *Game theory*. Unlike *Game theory*, in the case of the three aforementioned theoretical lenses it is possible to identify a predominance of the *Performance assessment* area, and following it there is *Information sharing.*

Consistently with the *Resource Based View*, Big Data can be considered a resource of the company as it is unique and difficult to replicate. To date, a supply chain has a large volume of data, which is useful for driving its performance, but in order to achieve supply chain effectiveness and efficiency, it is necessary to access data from different functional areas and different partners, both internal and external to the organization. Therefore, information must be shared in order to maximize profits, improve product quality, cost, speed, timeliness and flexibility (Chavez *et al*., 2017). In addition, data analytics capabilities improve an organization's procurement and management competitiveness. In detail, internal data analytics capabilities positively impact performance, while external data analytics capabilities positively impact digital procurement capabilities (Hallikas *et al*., 2021). The latter aspect also explains the presence of the *Procurement* activity in the three most relevant theories examined in Figure 16.

Looking closely at Figure 17, it is evident that the supply chain area *called Revenue/cost sharing* is contained exclusively in *Game Theory*. The presence of this area may be motivated by the additional costs each firm faces in order to obtain timely consumer information and more accurately forecast demand. For this reason, in order to balance the effects due to the additional costs, many decision makers adopted the cost-sharing model, which involve role-playing among different actors in the supply chain (Liu Pan, 2021). Overall, what has been discussed so far also motivates the supply chain activities mentioned at Figure 16.

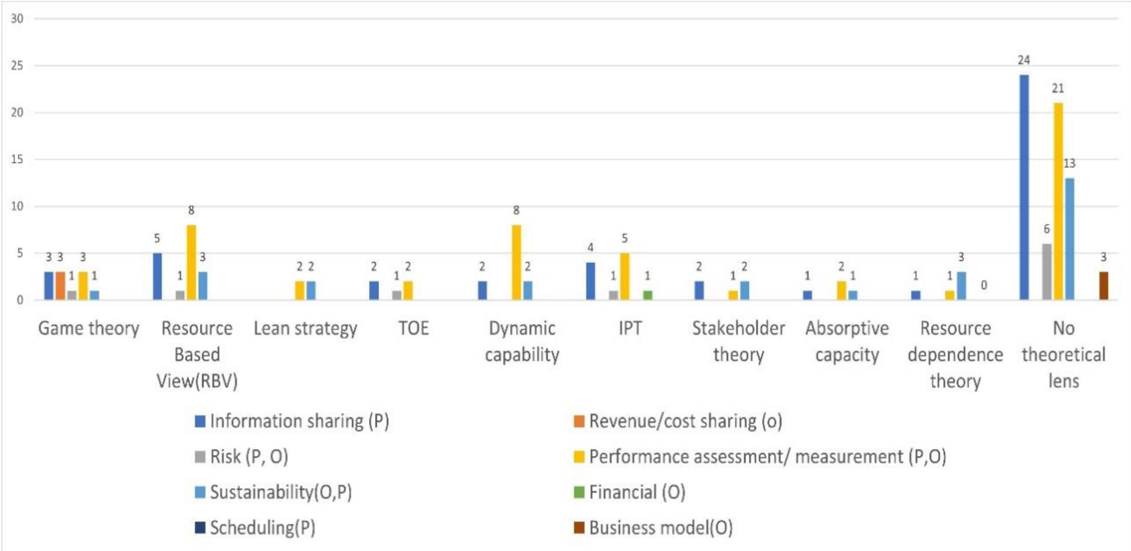

Figure 17 - Number of snowballing papers obtained from the intersection of supply chain areas and theories

One area, which isn't mentioned and appears overall in most of the theories in Figure 17 although to a small measure, is *Sustainability*. In fact, sustainable development has become a strategic goal. It has not only environmental but also economic and reputational implications for the company. All this requires a new approach, so-called Sustainable Supply Chain Management (SSCM), which includes *Lean Strategy* (Chalmeta *et al*., 2021). Equally little relevant among the theoretical lenses is the area of *Risk*. There are numerous risks, caused by external or internal sources, that supply chains must be able to address. One important characteristic that enables the supply chain to do this is resilience. The *Resource Based View* theory provides a relevant starting point on which to base strategic decision making for a resilient supply chain (Bag *et al*.,2021). Considering the key criteria of the *Resource Based View theory*, it can be concluded that Big Data Analytics (BDA) tools represent very valuable resources: they ensure rather efficient procurement and bid management processes, to the point of making the supply chain ready to face unpredictable risks. This last aspect clarifies the direct link present between *Resource Based View* and *Risk i*n Figure 17.

Finally, regarding the *No theoretical lens* category, the areas of supply chain already discussed extensively in the previous lines prevail in it, namely: *Information sharing, Performance assessment* and *Sustainability*.

In conclusion, the study regarding the number of snowballing papers resulting from the intersection of theories used more than once and methodologies can be observed in

Figure 18. Again, what emerged in Section 3.2.2 can be confirmed, in fact, it is precisely the most adopted methodologies that have a greater variety of theories. These are *Survey/case studies, Conceptual/framework* and *Literature Review* methodologies. While all three show a predominance of the *No theoretical lens* and *Resource Based View*, only two of these show a peak, albeit a small one, in correspondence with the *Dynamic Capability theory*. In essence, beyond the considerations just made, nothing more can be added given the failure to recognize a connection between the methodologies and theoretical lenses adopted by the authors.

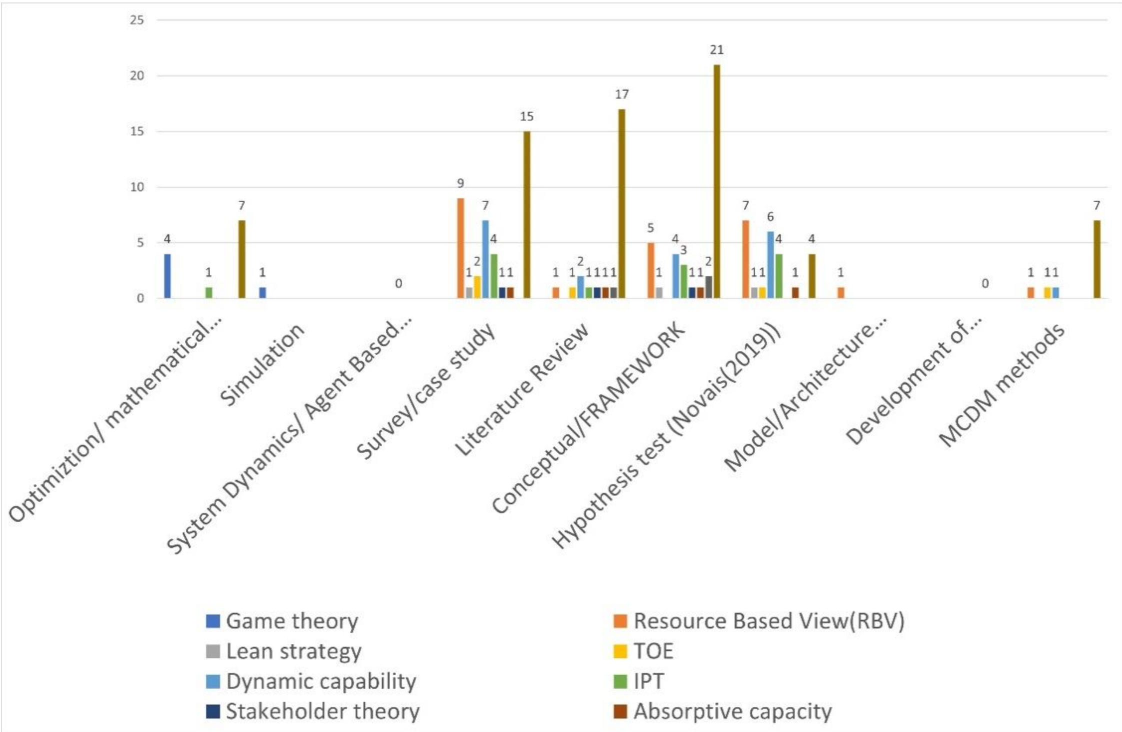

Figure 18 - Number of snowballing papers obtained from the intersection of methodologies and theories

3.2.5 Critical analysis of articles' years of publication

The seventy-six snowballing papers, subjected by this research thesis to critical analysis, were published in a time interval between 2013 and 2021. This new aspect represents a potential starting point from which to derive new observations.

From Figure 19 it is possible to see a not entirely increasing trend over the aforementioned time interval. There is a peak at the last year of publication.

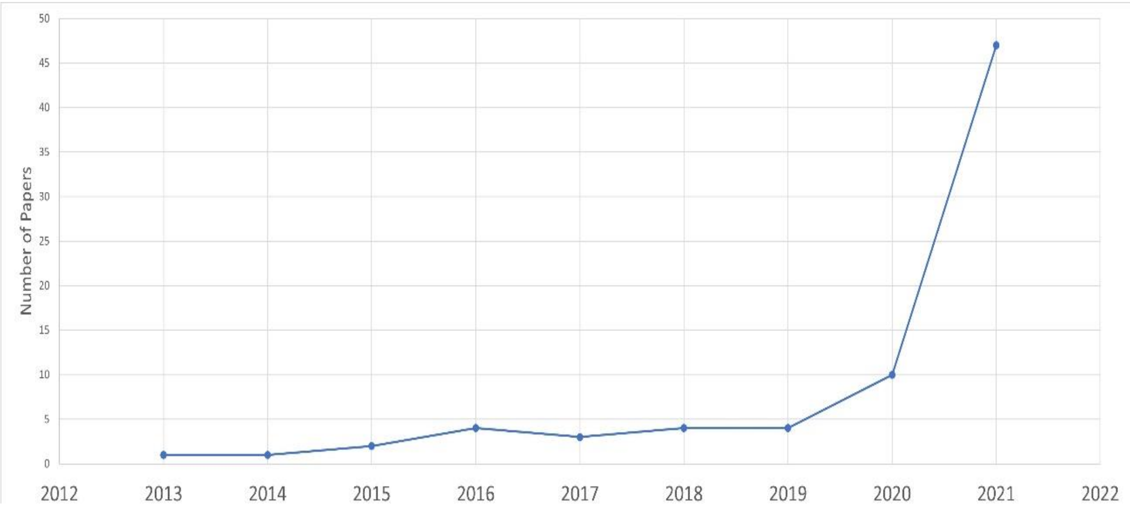

Figure 19 - Number of snowballing papers related to their years of publications

In detail, looking at the line that connects the number of snowballing papers published over the years, an increasing trend can be seen only from 2013 to 2016 and onward to 2019. Thus, it can be inferred that in recent years the literature manifests an increased focus on the study of the influence of Big Data in the supply chain environment, and consequently the term Big Data can be seen as a recent and significant trend for practitioners and academics (Frehe *et al*., 2014).

As anticipated in Section 3.1.1, in order to be able to develop an overall assessment of snowballing papers from an article perspective, it is possible to take advantage of a bibliometric index called *Citation Index*. As its name implies, it provides the number of citations related to an article, and this information is useful to know its level of notoriety. Following this brief introduction, a further point of view to be analyzed concerns precisely the *Citation Index*. The information suggested by this index can be analyzed

through a comparison with the years of publication belonging to each article. Is there a correlation present between these two types of information? Is it possible to extrapolate qualitative information from such quantitative data?
The number of citations is, as observed from Figure 20, between 0 and 683, and about twenty-four percent of the articles have no citations. Included in this newly expressed percentage are articles published exclusively in 2021, with the exception of one published in 2020. In conclusion, it can be deduced from the results obtained that the number of citations seems to be negatively affected by the year of publication. In fact, depending on how recent an article may be considered, its number of citations varies. An article that is older in time shows a higher number of citations. This can also be easily guessed since with the passage of time each snowballing paper increases its probability of being cited.

3.2.6 Critical analysis on the combined use of technologies

Following the analysis of the years of publication regarding the various snowballing papers, an ulterior aspect to which attention can be drawn concerns the combined use of multiple technologies. As repeatedly stated, the focus of this thesis is on supply chain digitization and places its concentration on one particular technology, namely Big Data. This is why all the snowballing papers analyzed feature content regarding Big Data. Despite this, more than fifty percent of the articles used or mentioned multiple technologies. In Figure 20 it is possible to observe in detail the percentages regarding the number of snowballing papers characterized by a combination of technologies and the number of snowballing papers characterized by the absence of a combination of technologies.

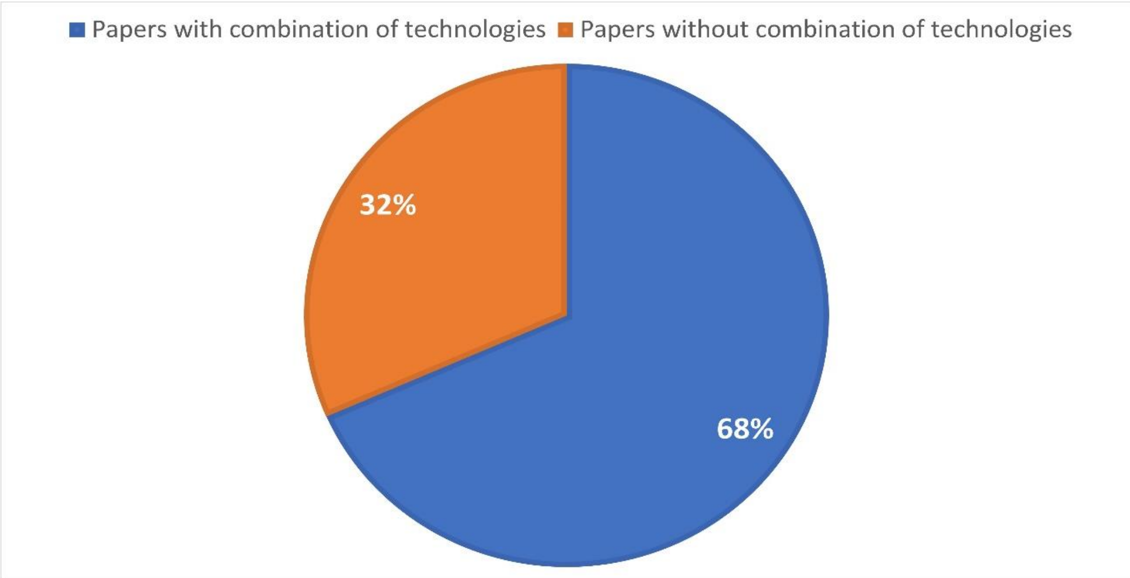

Figure 20 - Percentage of snowballing papers with or without combination of technologies

The objective of this type of critical analysis focusing on the combined use of multiple technologies is to know the existence of any links present between the various technologies in the case of their application in the supply chain environment. Such an objective in fact would also allow to understand what are the causes that pushed the authors towards the combined use of multiple technologies.

Before a conclusion can be reached and thus satisfy this goal, an attempt is made to outline a general framework that can clarify which connections or combinations were the most popular.
Figure 21 focuses exclusively on the combination of technologies that occurred between Big Data and the remaining technologies.

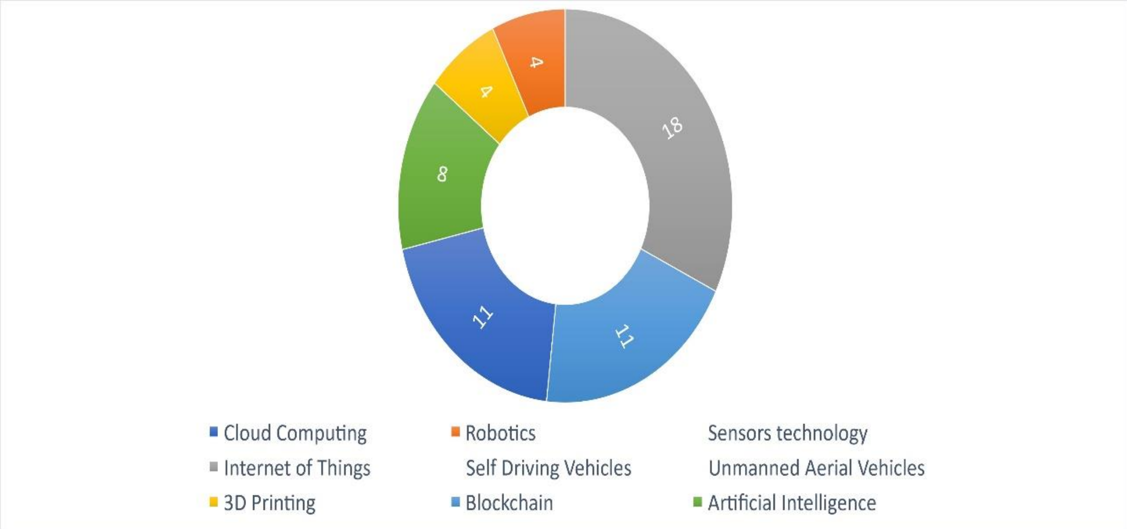

Figure 21 - Snowballing papers with combination of Big Data and other technologies

It is observed that Big Data has been mainly combined with *Internet of Things* (IoT). Next, an equal number of snowballing papers present the connection between Big Data and *Cloud Computing* (CC) and Big Data and *Blockchain*. Finally, *Artificial Intelligence* (AI), *Robotics*, and *3D Printing* follow. The only technologies that don't present combinations with Big Data are *Sensor technology* and *Self Driving Vehicles*.
So, what links Big Data to the three most relevant technologies that emerge from Figure 21? The need to raise the degree of digitization of the supply chain is increasing more and more. In addition, consumers, thanks to greater supply and more technological channels of access to the market, appear to be increasingly aware and attentive, for example, to the timing of delivery. As a result, it is important that supply chains know how to digitize effectively and efficiently, and to do so, Big Data should be combined with other technologies. The findings from Figure 21 confirm what was learned through the content analysis of the snowballing papers: the *Internet of Things, Big Data Analytics*, and *Blockchain* technology are the most important enablers for supply chain digitization (Agrawal and Narain, 2021). From a management perspective, *Internet of*

Things can increase supply chain transparency and consequently ensure clarity and timeliness for the flow of information. Moreover, this technology can help identify a problem before it occurs or facilitate updated decision making through real-time data collection. *Blockchain* technology, instead, as anticipated in Section 1.3.11, represents a turning point for the supply chain because it is useful for the exchange of digital information. Such information, thanks to *Blockchain*, would be constantly protected. On the other hand, the fear of losing confidential information represents one of the most feared risks for any organization. In conclusion, it is inferred that organizations must first possess structures to capture data (*Internet of Things*), which must be able to be analyzed (*Big Data Analytics*) and shared. Finally, speaking of sharing, it is important to ensure a high level of data protection (*Blockchain*).

Equally important is *Cloud Computing* technology, introduced in Section 1.3.7. Indeed, in order for information obtained from Big Data to be communicated and shared within the supply chain, the presence of this additional technology should be provided. Cloud computing thus facilitates collaboration, reduces software installation costs, and optimizes inventory chain management (Agrawal and Narain, 2021).

Thus, this information is useful in answering the goal anticipated in the previous lines. In conclusion, in order to get an overall framework, which therefore also includes the possible combinations present among all other technologies outside of Big Data, it is possible to look at Figure 22.

The thickness of the line connecting two technologies in Figure 22 is related to the number of snowballing papers that presented the considered technology combination. In other words, a greater number of snowballing papers with the combination of technologies in question corresponds to a greater thickness. The graph again confirms what has been said in the previous lines. In fact, it is precisely the most important enabling factors that count more connections with the other technologies.

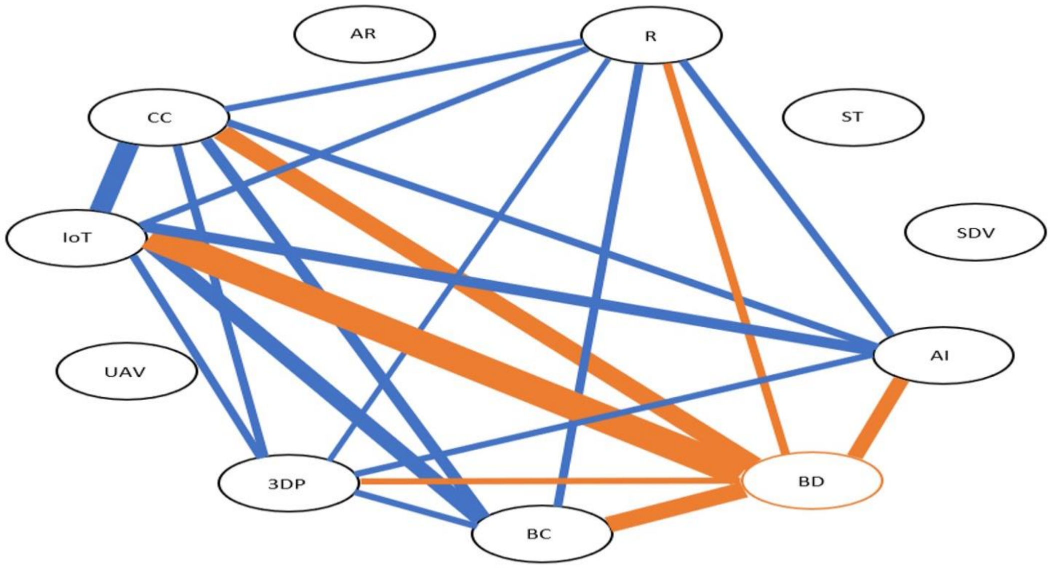

Figure 22 - Combination of all technologies

3.3 Second analysis perspective: the journal

This section addresses the second of the three perspectives announced in Section 3.1.1, namely, the journal. In fact, new bibliometric indices are introduced in the following sections, from which quantitative data useful for qualitative information processing can be derived. Thus, these are indices closely related to the paper. In this regard, the chapter focuses not only on the analysis of the indices but also proposes a comparison of them, an observation of them from the point of view of the three categories.

3.3.1 Journal distribution

Before delving into the analysis of the results regarding the quality of the journals in which the snowballing papers were published, it is appropriate to provide some general information regarding the journals and the number of snowballing papers associated with them. Premising on the fact that several snowballing papers may have been published in the same paper, it was found that the number of different papers is forty-nine. Thus, it is confirmed that one journal doesn't necessarily correspond to one snowballing paper.

Figure 23 shows that the most popular journal is *International Journal of Production Research*. The snowballing papers associated with this first journal were found by forward snowballing except for one. In addition, the journal mainly includes the most relevant supply chain activities: *Customer relationship*, *Demand forecasting, marketing and pricing,* and *Procurement*. The same applies to the most relevant supply chain areas, i.e., the journal includes snowballing papers mainly concerning *Information sharing* and *Performance assessment*. The remaining supply chain areas argued include *Risk* and *Sustainability.*

This first journal is followed with three fewer articles *by International Journal of Production Economics*. In this case the articles were found using the backward snowballing technique, with the exception of one. As happened with the previous journal, the articles reconfirmed the most relevant supply chain areas and activities. The only exception is the presence of the *Inventory management* activity and the *Financia*l area, argued in two separate articles, respectively. The remaining journals count between one and three snowballing papers.

The results obtained confirm the high relevance argued so far regarding the most popular supply chain areas and activities.

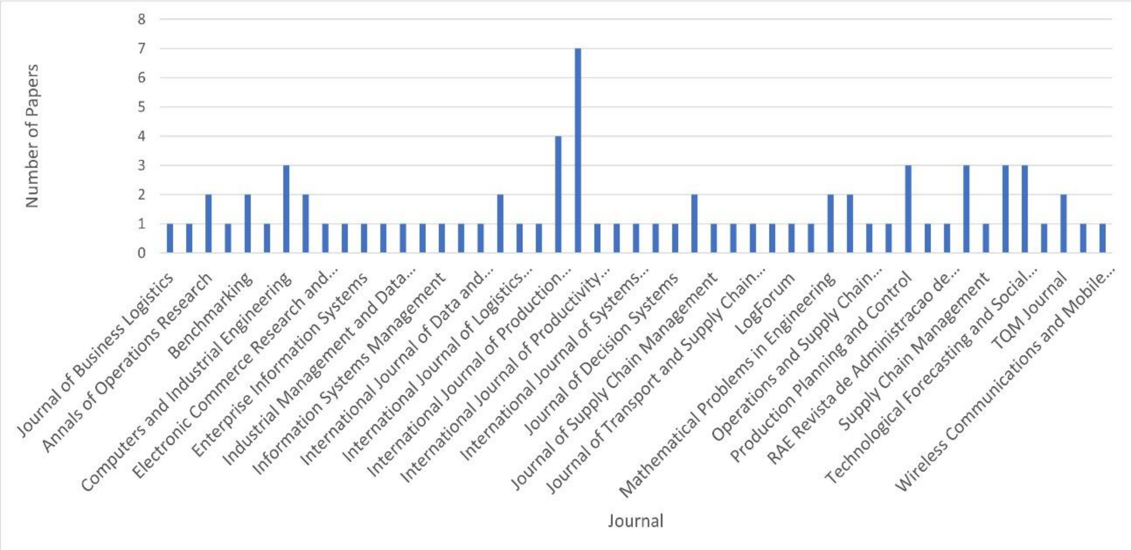

Figure 23 - Number of snowballing papers related to the journal

3.3.2 Journal quality and its critical analysis: 5-Impact Factor

As announced in the first lines of this chapter, it is possible to conduct critical analysis through the use of bibliometric indices that provide quantitative information, which in turn, are sources of qualitative considerations. In order to be able to deepen the critical analysis from the perspective of the journal, it is necessary to introduce a new index, called the *5-Impact Factor*. This information was obtained through the official website of each journal, specifically in the section called *Indexing & Ranking*. *5-Impact Factor* measures the average number of citations received, in the reference year under consideration, by articles published by a scientific journal in the previous five years [37]. *5-Impact Factor* expresses the impact or popularity of a publication on the reference scientific community. It represents the unit of measurement that "weighs" the value of scientific publications. However, in order to understand what prompted the author of this thesis toward the use of this index when evaluating journals, it is necessary to do a little background about scientific publications and their importance.

When an author completes his research paper, he sends it to a scientific journal. Several steps follow at this point. In general, a majority of journals, in order to decide whether or not to publish the research, perform a *peer-review*. Several experts from around the world are asked to evaluate the paper, and only if it is successful can the publisher proceed with publication [38]. The *Impact Factor* represents the "goodness" of a journal and that is why the higher its value, the more positively the journal is considered. So not all journals have equal importance, and to measure this importance, it is strictly necessary to know this index.

Having extensively defined this index and its importance in assessing the reliability of the results of this research thesis, it is possible to define the general situation regarding journals associated with snowballing papers.

It is premised that just under fifty percent of journals don’t present this information. It is important to clarify that only *Impact Factors* derived from *Scopus* were considered valid and no other sources, such as *Clarivate*.

From the twenty-seven journals presenting the *5-Impact Factor*, an average of 4.7 is obtained for this index. This is certainly a very good result since a *5-Impact Factor* equal to 1 indicates that on average there was one citation in the reference year for each article published by the journal in the previous five years.

As can be seen from the graph in Figure 24, *Journal of Supply Chain Management* is the journal that represents the highest *5-Impact Factor* at 8.647.

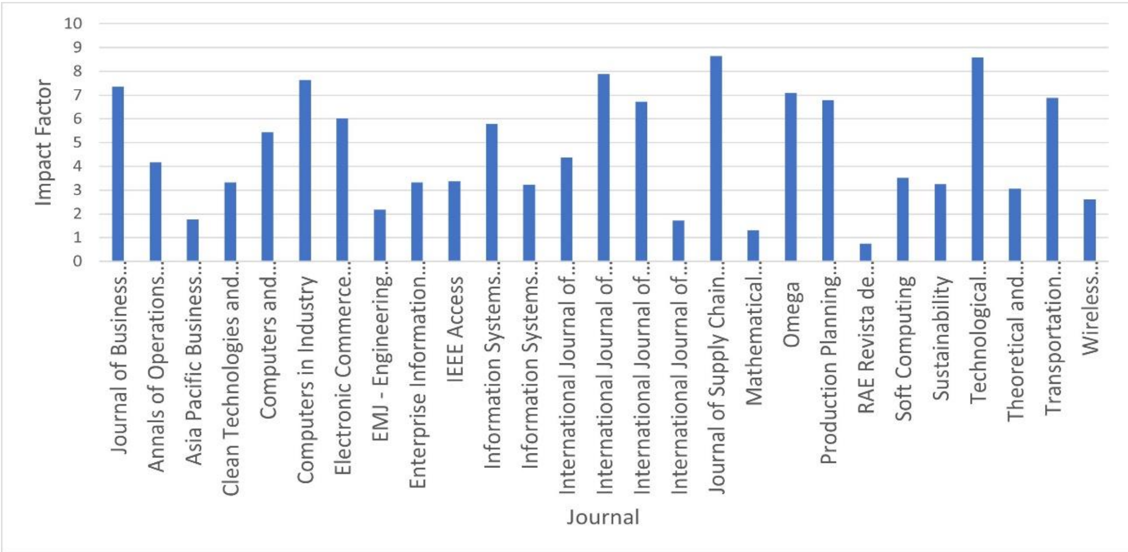

Figure 24 - 5-Impact Factor of journals

The journals having an index rather near the value just reported are: *International International Journal of Production Economics* and *Technological Forecasting and Social Change*. The latter two include the journal mentioned in Section 3.3.1. Next, journals report successively lower values, until they reach a value of 0.755.

However, there are other insights that the impact factor offers. They are useful for the analysis of journals and consequently for the analysis of snowballing papers. In fact, there is an additional reason that recognizes the support of impact factor in evaluating the journals in which articles are published. It is believed that journals with a higher impact factor are more selective in choosing papers for publication [39]. Therefore, through assessing the degree of "hospitality" of each journal, it is possible to qualitatively rank snowballing papers from the perspective of their content.

Not surprisingly, *Journal of Supply Chain Management*, which has the highest *5-Impact Factor*, is associated with an article that argues the most relevant supply chain area, namely *Information sharing*. Regarding supply chain activities, it doesn't identify one in particular, rather it talks about *Overall supply chain*. In general, the fact that such articles, with high impact factor, discuss the most relevant supply chain areas and activities is highly significant to the results of this research. However, the presence of a rather high percentage of journals without *5-Impact Factor* might be somewhat limiting for the considerations elaborated in this section.

3.3.3 Journal quality and its critical analysis: Q-index

Alongside the Impact Factor, there is another bibliometric index useful in evaluating the journal and the article contained in it. This is the *Scimago Journal Ranking* (SJR), an index considered much more complex than the Impact Factor. In fact, *the Scimago Journal Ranking* is also calculated on the basis of citations received by a journal, however it "weighs" citations according to the prestige of the citing journal [40].

In order to look up the *Scimago Journal Ranking*, simply enter the journal title on the Scimago *Journal & Country Rank* platform, from which it is possible to view the quartile of the category rank in which a journal ranks.

The quartiles are denoted by Q1, Q2, Q3, Q4, respectively. Quartile information could not be obtained for only about five percent of the cases, as shown in Figure 25, that is, five snowballing papers out of a total of seventy-six.

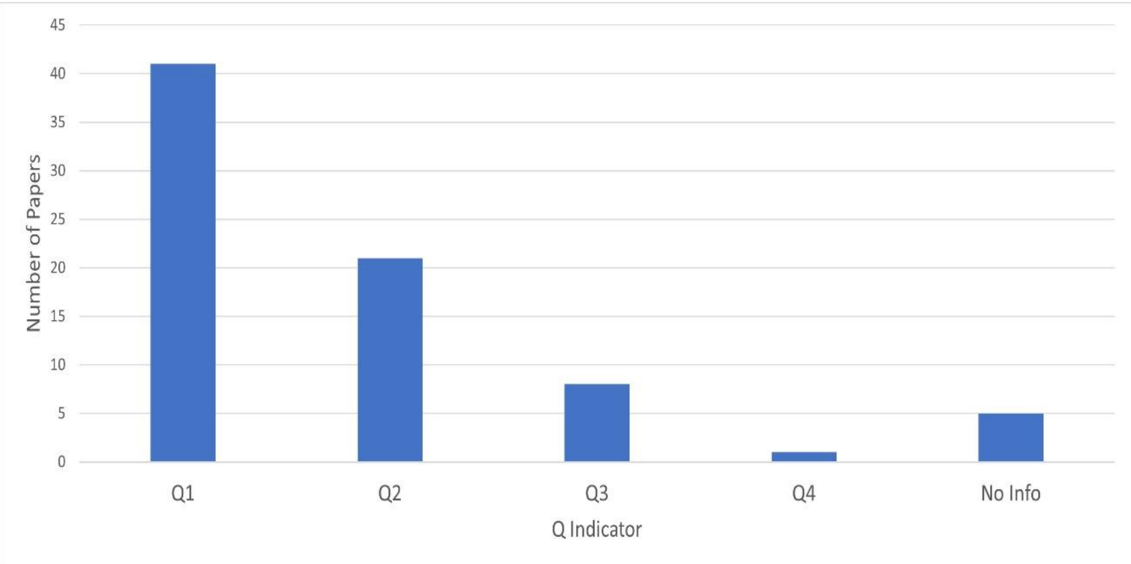

Figure 25 - Number of snowballing papers related to quartiles

The remaining snowballing papers are included in journals with quartiles that are in most cases relevant. In fact, the most "reliable" journals belong to quartile Q1. Thus, it is easy to deduce that the degree of reliability of journals decreases with Q2 and so on. A high quartile in Scopus is certainly a good result for the journal. In that case, it would be a journal that can determine research trends and whose authors can be considered experts in their field [41]. Moving on to the information shown in Figure 26, it is evident, as anticipated, that just over fifty percent of the papers report a Q1 quartile. In contrast, almost thirty percent report a Q2 quartile, and ten percent of the snowballing papers report a Q3 quartile. Only one snowballing paper presents a Q4 quartile.

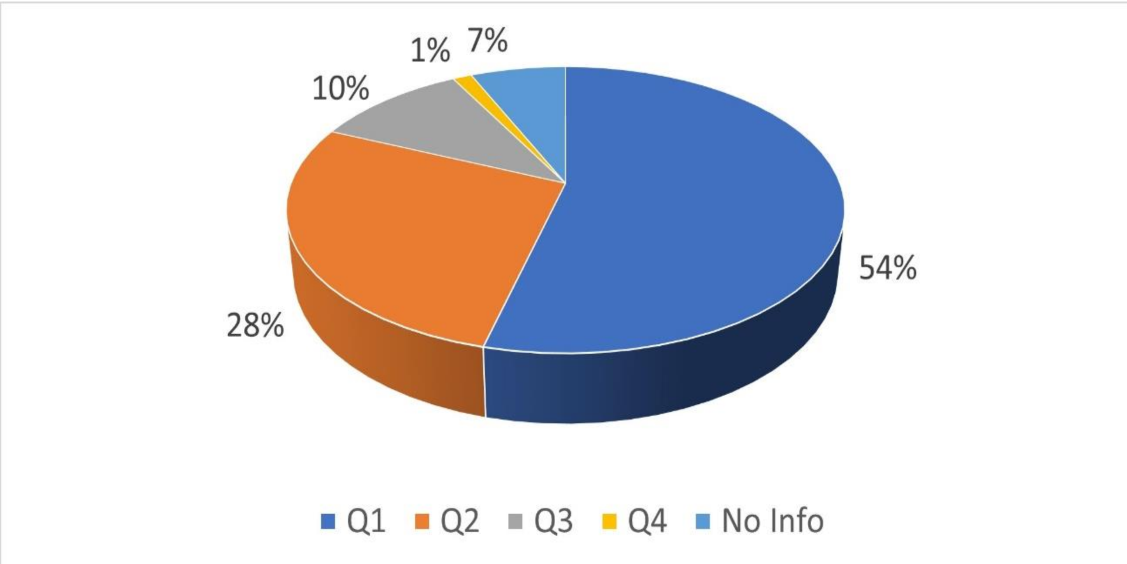

Figure 26 - Percentage of snowballing papers related to quartiles

Given the results obtained, it is concluded that since a large proportion of the snowballing papers were published by a journal with quartile Q1 and Q2, the critical analysis conducted in this research paper is based on significant outcomes. These results can therefore be considered as a potential springboard for future researchers despite the presence, albeit small, of snowballing papers that were not classified from a quartile perspective.

In summary, the Q-Index ,discussed in this Section, and the 5-Impact Factor, discussed in Section 3.3.2, define the quality of the journals in which the snowballing papers were published. Apparently, these two indices would seem to be connected through a direct link. In other words, a high Impact Factor value corresponds to a relevant quartile. In reality, what makes the reliability of these considerations limited is the rather high percentage of papers that don't have information on the Impact Factor. The same can't be said about quartile information. In fact, the percentage of missing data, shown in Figure 26, is quite small, so even if there were such information, the results wouldn't be completely skewed.

3.3.4 Q-index and three categories

The Q-Index, more properly called quartile, can be analyzed from an additional perspective, namely the three categories. This insight would identify a possible link present between journal quality and areas, supply chain activities, or methodologies. For example, do journals with higher quartile Q1 prefer the most relevant supply chain areas or activities? What topics are mainly addressed by journals with higher quality?
Starting with the supply chain activities and looking at Figure 27, it can be seen that no particular new correlations can be revealed. In fact, from the bar graph it can be seen exclusively that the highest quartiles, i.e., Q1 and Q2, argue mainly for the most popular activities and previously discussed in Section 3.2.2.

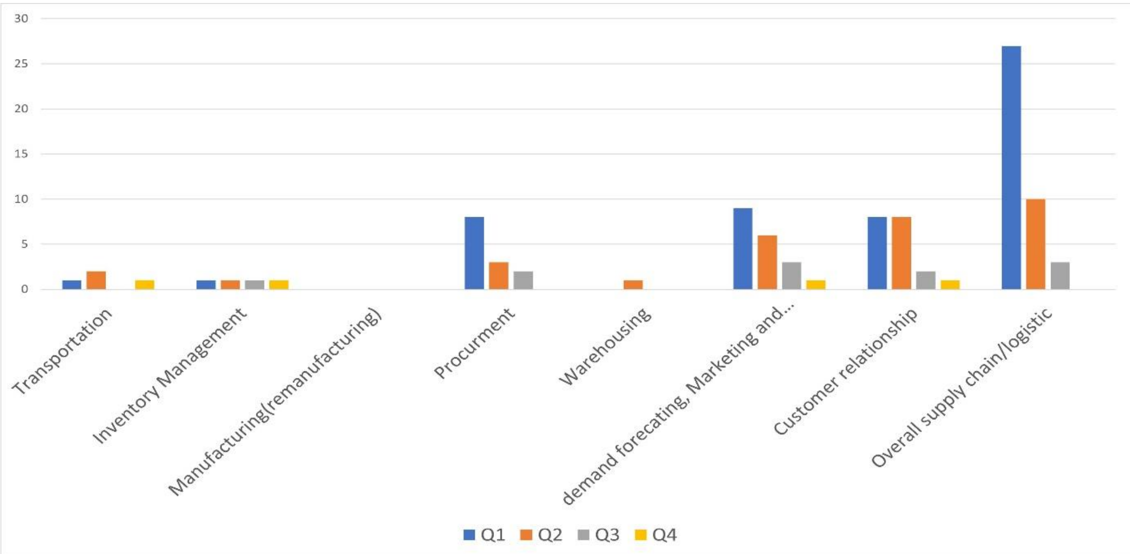

Figure 27 - Number of snowballing papers grouped by quartile and related to supply chain activities

Recalling that, as anticipated in Section 3.3.3, a journal with high quartile is able to determine research trends, the fact that these most relevant activities are argued in snowballing papers included in journals with quartile Q1 and Q2 means that the activities also represent a potential focus on which academics will further concentrate in the future. The result therefore is not only significant in terms of reliability, but also in terms of future research prospects. The same results, even if with reduced numbers of snowballing papers, occur at quartile Q3. Quartile Q4, on the other hand, because it has only one snowballing paper is not relevant for critical analysis. Finally, the absence of columns in the graph corresponding to *Manufacturing* activity is noted, and this is intuitively due to the absence of snowballing papers in that category.
Turning to the supply chain areas, it can be seen from Figure 28 that again the Q1 and Q2 quartiles mainly argue the most popular supply chain areas.

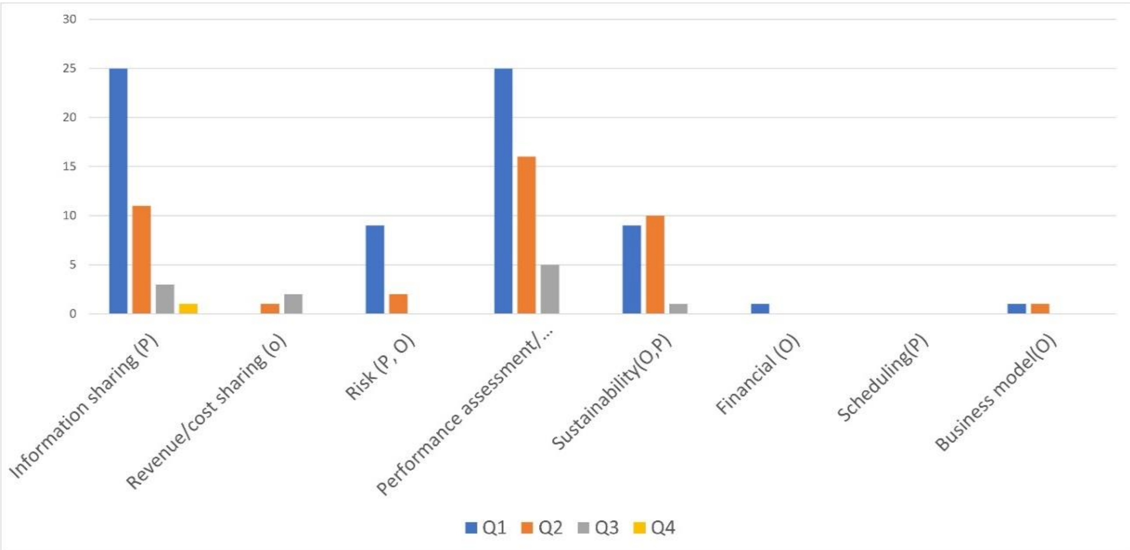

Figure 28 - Number of snowballing papers grouped by quartile and related to supply chain areas

For this reason, what was discussed in the previous lines with supply chain activities can also be considered valid for the *Performance assessment* and *Information sharing* areas. Indeed, this result not only confirms the relevance of the aforementioned areas but also announces a potential focus on which future research can concentrate. Moreover, the areas are also confirmed to be dominant in the case of quartile Q3. Quartile Q4 isn't considered relevant for qualitative considerations because it has only one snowballing paper. The *Scheduling* area has no columns in the graph given the absence of snowballing papers in that category.

The third and final category, against which the quartiles of papers and the snowballing papers contained in them are analyzed, is methodology. It is possible to observe its details in Figure 29.

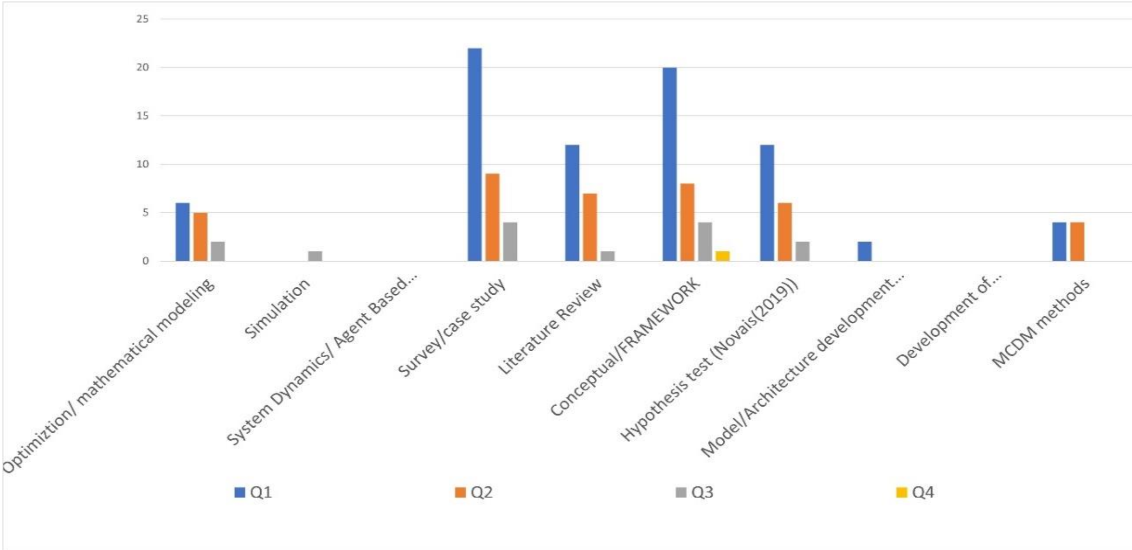

Figure 29 - Number of snowballing papers grouped by quartile and related to methodologies

Finally, Figure 29 also reports nothing new and doesn't show any particular correlations, other than the prevalence of the most relevant methodologies at the Q1 and Q2 quartiles. Therefore, it is assumed that the authors, who will delve into the influence of Big Data on the supply chain in the future, will continue to predominantly use these methodologies. At quartile Q3, the same prevalences occur, with the exception of the *Literature review* methodology. Reiterating what was said with the previous categories, again quartile Q4 is not relevant because it contains only one snowballing paper. *System Dynamics/ Agent Based modeling* and *Development of Tool/Platform/Computer system* have no columns in the graph given the absence of snowballing papers in those categories.

Thanks to the considerations made in the previous lines, it can be seen that papers with higher quartiles argue and present the categories that are already currently most popular in the literature. This aspect, not only outlines what might be future research trends, but at the same time affirms the relevance of most of the snowballing papers subjected to critical analysis.

In conclusion, the results that emerged in this section can be considered remarkable given even the small percentage of snowballing papers for which there is no quartile information.

3.4 Final Considerations

The objective of this research work is, as reported several times on the various pages, to study the influence of Big Data on the supply chain. Not only that, it also seeks to motivate and demonstrate how the use of Big Data can effectively support supply chain actors in various areas or activities, thereby incentivizing future researchers to overcome the limitations of the literature and obstacles that still characterize the transition of the traditional supply chain to digitization.

Before reviewing what are the actual results that have emerged from this thesis, therefore, it is to show how Big Data represents a trending technology. Not surprisingly, through an analysis of google trends, which allows to examine how frequently the term Big Data has been searched around the world, it emerges that there is a real upward trend from 2010 to 2017. In more recent years, on the other hand, there is no real trend. It is possible to observe this in Figure 30.

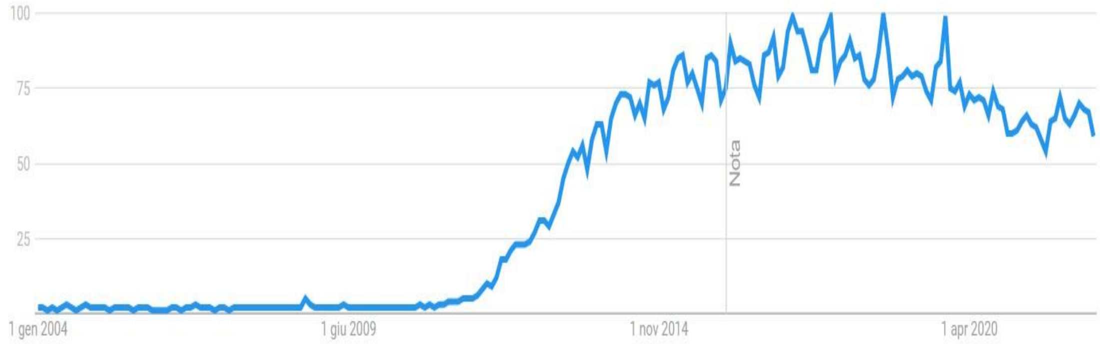

Figure 30 - Google trends of term "Big Data" [42]

In particular, it is interesting to see which countries have contributed most to the rise in the number of searches for the term Big Data. More details can be found in Figure 31. The map includes values calculated on a scale from zero to one hundred, where one hundred indicates the location with the highest frequency of searches in proportion to the total number of searches for that location, and the value fifty indicates a location with half as many searches. The value zero, in contrast, indicates a location for which insufficient data were collected. By reporting only the top five rankings, it is therefore inferred that the largest contribution is associated predominantly with the Asian territory.

Figure 31 - Map of Google trend for term "Big Data" [42]

In reality, it is also necessary to emphasize the need to attach importance to another country as well, namely *India*. It, in fact, isn't only the country of origin for most of the authors of the snowballing papers analyzed, but it is also the country that the literature has often focused on, highlighting the applications of Big Data in Indian manufacturing chains.

So, it is a matter of a developing country, *India* is the fastest growing economy in the world but at the same time it has a large number of inhabitants with a rather high poverty level [43].

For this reason, this result shouldn't suggest that *India* is the first country to apply the use of Big Data to the supply chain world, but it simply indicates a strong interest in the topic. There are many barriers that limit the implementation of Big Data Analytics in Indian manufacturing supply chains.

From the literature, it is learned that the most critical obstacles are (Raut, R.D. *et al.,* 2021):

- Lack of top management support;
- Lack of financial support;
- Lack of expertise;
- Lack of techniques or procedures.

In light of this situation, final considerations are offered below that allow the reader to be able to draw conclusions about what are the real benefits of Big Data on the supply chain.

As shown by the articles with higher relevance scores, Data-Driven Supply Chain (DDSC) allows for improvement in complex supply chain management, and most importantly, data-driven supply chains are positively correlated with multiple dimensions of production that consequently lead to improved customer satisfaction (Chavez *et al*., 2017). Production dimensions include the following aspects: quality, delivery, flexibility and cost. It is easy to imagine that a quality supply chain guarantees excellent results and incentivizes customer loyalty, which usually requires very short lead times. The supply chain must be able to demonstrate flexibility in case of changes in production volume or special requirements. Finally, organizations must be able to set the right cost level that allows products to be priced appropriately and that ensures a profit.

Big Data supports each of the aspects just described; in fact, data sharing enables product defects to be minimized through early product simulation. In addition, data sharing incentivizes collaboration among the various supply chain actors, which reduces design time and improves manufacturability and on-time delivery. Knowledge of the data enables demand forecasting or knowing customer needs in time, and through this, supply chains can be flexible. Improved manufacturability and demand forecasting results in reduced inventory and consequently reduced cost.

Overall, Big Data makes it possible to improve the relationship organizations have with customers and suppliers, thereby reducing inventory. Suppliers therefore can be more easily evaluated and selected (Zekhnini *et al.,* 2021). It is inferred that Big Data is also useful in inventory and logistics management.

In conclusion, because Big Data affects many aspects of the supply chain, it impacts the performance of organizations. In fact, Big Data and Big Data Analytics influence Supply Chain Performance (SCP) based on three roles: supportive facilitator, source of empowerment and game-changer (Xu *et al.,* 2021). Indeed, as a supportive facilitator, Big Data aims to facilitate process improvement, which consequently improves SCP performance. As a source of empowerment, Big Data enables new processes, while as a game-changer it aims for radical changes by integrating new goals.

The aspects discussed so far, on which Big Data impacts, thus without a doubt recall all the supply chain areas and activities discussed extensively in Chapters 3.2, 3.3. and 3.4. However, this is not sufficient to conclude the discussion of the results. In fact, one very important and widely influenced by Big Data hasn't yet been mentioned: supply chain *agility. Agility*, introduced in Section 1.4, can be considered as the key factor for a chain to survive in uncertain and highly fluctuating markets. Therefore, it is necessary to possess it in order to achieve competitive advantage. In this regard, Big Data supports and facilitates real-time supply chain monitoring and this consequently improves the speed and agility of organizations' decisions (Jayender and Kundu, 2021).

A final aspect anticipated in Section 1.5, but not yet discussed, is *Bullwhip Effect* (BWE). It refers to the phenomenon of demand amplification affecting the entire supply chain. The use of Big Data technology has a significant impact on Supply Chain Management (SCM) (Raman *et al*., 2018) and consequently on operational aspects such as demand signal processing, order batching, and price changes. All these aspects are in fact sources of the *Bullwhip Effect*.

4.Conclusions

The fourth and final chapter of this research thesis presents the conclusions of the work, highlighting its benefits and limitations. In addition, the chapter is intended to serve as a starting point for future researchers through exploring aspects that could be further discussed in the literature.

4.1 Benefits of the present study to the literature state of the art

Today, companies are immersed in the information age, which in turn is accompanied by a major digital revolution. The increasing evolution of this phenomenon has caused an increase in the amount of data that supply chains must be able to manage. So, in detail, it is possible to speak of an *era of Big Data*. For this reason, the goal of this research paper, as anticipated several times, is to study the applications of Industry 4.0 digital technologies to the supply chain. The various technologies were assigned to different members of the research team, so this thesis was devoted exclusively to Big Data. It is, as demonstrated extensively in Chapter 3, a very recent technology. Although it has in fact attracted the attention of literary scholars and supply chain managers only in recent years, Big Data can be considered one of the most important technologies. Big Data represents a technology that companies can't do without. Not surprisingly, the sheer volume of big data is leading supply chain digitization to no longer be considered a desire but a necessity. Thus, it is easy to deduce that the first benefit of this thesis is precisely its focus on the most important technology for companies. In fact, Big Data is able to improve the daily life of the company not only because of the large volume of information it provides, but also because of the possible decisions that can be made after the data analysis.

For example, data can be evaluated from different aspects:

- Cost reduction;
- Time reduction;
- Product development;
- Consumer preferences;
- Strategic decisions.

Big Data can revolutionize a company's methods of organization. In this regard, it is important to specify the fact that the application and use of Big Data actually involves two business revolutions and not just one. In fact, before Big Data can be used, it is necessary to subject the supply chain to a digital transformation. Such change requires a large investment in time, money and structure (Agrawal and Narain, 2021). Therefore, it is crucial for each organization before embarking on such a big change to analyze what its critical aspects are and what enabling factors the digitization process requires more attention. Additionally, after preparing the suppy chain for the application of this new technology, it revolutionizes strategic decisions and the entire performance due to its benefits. Making known the importance and magnitude of the changes that this technology requires of companies is an additional benefit brought by this thesis work through the content analysis of snowballing papers.

In no article analyzed was an overview presented that delved into the current state of Big Data application in all areas and activities of the supply chain. In this regard, this research paper fills the literature gap just described, providing the reader with the opportunity to explore topics and findings that have not been discussed until now. In fact, all seventy-six snowballing papers were categorized, as described in Section 2.4, according to the type of supply chain activities argued, according to the supply chain areas described, according to the research methodology followed in each paper, and according to the type of technologies discussed simultaneously in the same paper or theories used by the authors. This work therefore facilitates and directs future analysis of specific aspects of the supply chain.

The study turned its attention only to articles focused on the manufacturing sector. This further supports the reader and researchers in understanding the influence of Big Data closely related to the supply chain, unlike the previous literature containing only very general articles.

4.2 Limitations

Despite the benefits brought by this research work, discussed at length in Section 4.1, it is good to expose the limitations of the paper too.

As explained in Section 2.4, all literary articles resulting from the application of *Snowballing* passed a selection process in order to then be definitively considered *snowballing papers* and subjected to critical analysis. Exclusion criteria include inaccessibility and the presence of any conferences associated to the article. Further details in this regard can be observed in Figure 8. It can be inferred that the presence of articles excluded for the above-mentioned reasons may be a possible limitation towards the results obtained. In this regard, this limitation could have been eliminated or at least reduced with the use of more databases. Not surprisingly, the databases used as a source for searching articles were two: *Scopus* and *Google Scholar*. This aspect, therefore, is also a limitation for research.

As anticipated in Section 4.1, the study focused exclusively on the manufacturing sector and this limits the results. In fact, they could have included other sectors as well and thus would have allowed the reader more generalized conclusions that are valid for any type of supply chain.

Moving on to content analysis, the following are the main gaps that emerged from the seventy-six snowballing papers. From the most recurring gaps, it emerges that the literature lacks an explanation that can support and explain to companies the transition from the traditional supply chain to a digitized supply chain. No paper in fact exists on the way enabling factors are applied. In addition, few studies have investigated the impact of Big Data and Predictive Analytics (BDPA) on social and environmental sustainability in supply chains. The research findings discussed in Chapter 3 do not fill these gaps.

In order that Big Data can be implemented and analyzed in a supply chain, extensive knowledge is required, but as noted in the article entitled *Data science, predictive* *analytics, and big data: A revolution that will transform supply chain design and management*:

Typically, there is no single individual that can possibly have all of what is needed by a data scientist (Waller *et al*., 2013).

Indeed, the articles reveal an almost complete absence of individuals capable of filling the role of Data Scientist or Chief Data Officer, and, in this regard, the study doesn't fill that gap by providing good tools for training staff who support the supply chain.
In conclusion, one last limitation that can be highlighted concerns a particular supply chain activity, namely *Warehousing*. This is an activity that is highly influenced by the use of Big Data, and despite this, only one article related to this category was found. Therefore, it can be inferred that the research work didn't report great results in this regard.

4.3 Future Research

Sections 4.1 and 4.2 provide a cue for future researchers to explore or undertake new research. It is therefore recommended that future researchers, in the event of a new systematic literature review, expand the number of databases used, extend the critical analysis to the non-manufacturing sector, develop a framework that can support supply chain actors in digitization, delve deeper into the influence of Big Data in warehousing, and above all, provide useful tools for training corporate figures similar to the Data Scientist or Chief Data Officer. In addition, as the critical analysis revealed, the citations associated with an article may increase with the passage of time, and this may represent the possibility of finding new snowballing papers. It is recommended, therefore, to resume systematic literature review in the future.

Future researchers, in order to further investigate aspects of the supply chain influenced by Big Data, could better study all the activities that have recorded a limited number of snowballing papers in current research. These are *Transportation*, *Inventory Management*, *Manufacturing* and *Warehousing*. The same applies to areas and methodologies little mentioned in the analyzed papers, namely *Revenue/cost sharing, Financial*, *Business model* and *Scheduling*. The methodologies with little in-depth study, on the other hand, are: *Simulation*, *Dynamic system/agent based modeling*, *Model/architecture development*, *Tool/platform/computer system development*.

In conclusion, speaking instead of technological combinations, in the future scholars may further explore the connections between Big Data and technologies not mentioned so far. In fact, Big Data to date has no connections with the following technologies: *Augmented Reality* (AR), *Sensor Technology* (ST), *Self Driving Vehicles* (SDV), and *Unmanned Aerial Vehicles* (UAV). Therefore, it would be useful to understand whether the total absence of such technological combinations is due to the lack of compatibility of Big Data with the aforementioned technologies, or simply the lack of relevant articles with respect to this topic.

References

Agrawal, P., Narain, R. (2021). "Analysis of enablers for the digitalization of supply chain using an interpretive structural modelling approach", *International Journal of Productivity and Performance Management*.

Bag, S., Dhamija, P., Luthra, S., Huisingh, D. (2021). "How big data analytics can help manufacturing companies strengthen supply chain resilience in the context of the COVID-19 pandemic", *International Journal of Logistics Management*.

Bamel, N., Bamel, U. (2020). "Big data analytics based enablers of supply chain capabilities and firm competitiveness: a fuzzy-TISM approach", *Journal of Enterprise Information Management*.

Cetindamar, D., Shdifat, B., Erfani, E. (2021). "Understanding Big Data Analytics Capability and Sustainable Supply Chains", *Information Systems Management.*

Chalmeta, R., Barqueros-muñoz, J.-E. (2021). "Using big data for sustainability in supply chain management", *Sustainability*.

Chavez, R., Yu, W., Jacobs, M.A., Feng, M. (2017). "Data-driven supply chains, manufacturing capability and customer satisfaction", *Production Planning and Control*.

Frehe, V., Kleinschmidt, T., Teuteberg, F. (2014). "Big data in logistics-identifying potentials through literature, case study and expert interview analyses", *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*.

Hallikas, J., Immonen, M., Brax, S. (2021). "Digitalizing procurement: the impact of data analytics on supply chain performance", *Supply Chain Management*.

Hazen, B.T., Skipper, J.B., Ezell, J.D., Boone, C.A. (2016). “Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda”, *Computers and Industrial Engineering*.

Jalali S., Wohlin C. (2012). “Systematic Literature Studies: Database Searches vs. Backward Snowballing”, *Proceedings of the 2012 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement*.

Jayender, P., Kundu, G.K. (2021). “Intelligent ERP for SCM agility and graph theory technique for adaptation in automotive industry in India”, *International Journal of Systems Assurance Engineering and Management*.

Liu, P. (2021). “Pricing rules of Green Supply Chain considering Big Data information inputs and cost-sharing model”, *Soft Computing*.

Meriton, R., Bhandal, R., Graham, G., Brown, A. (2021). “An examination of the generative mechanisms of value in big data-enabled supply chain management research”, *International Journal of Production Research*.

Raman, S., Patwa, N., Niranjan, I., (...), Moorthy, K., Mehta, A. (2018). “Impact of big data on supply chain management”, *International Journal of Logistics Research and Applications*.

Ran, W., Wang, Y., Yang, L., Liu, S. (2020). “Coordination Mechanism of Supply Chain considering the Bullwhip Effect under Digital Technologies”, *Mathematical Problems in Engineering*.

Raut, R.D., Yadav, V.S., Cheikhrouhou, N., Narwane, V.S., Narkhede, B.E. (2021). “Big data analytics: Implementation challenges in Indian manufacturing supply chains”, *Computers in Industry*.

Thekkoote, R. (2021). “Understanding big data-driven supply chain and performance measures for customer satisfaction”, *Benchmarking*.

Waller, M. A., & Fawcett, S. E (2013). “Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management”, *Journal of Business Logistics*.

Xu, J., Pero, M.E.P., Ciccullo, F., Sianesi, A. (2021). “On relating big data analytics to supply chain planning: towards a research agenda*”, International Journal of Physical Distribution and Logistics Management*.

Yang, M., Fu, M., Zhang, Z. (2021). “The adoption of digital technologies in supply chains: Drivers, process and impact”, *Technological Forecasting and Social Change*.

Zekhnini, K., Cherrafi, A., Bouhaddou, I., Benghabrit, Y., Garza-Reyes, J.A. (2021). “Supply chain management 4.0: a literature review and research framework”, *Benchmarking*.

List of visited websites

[1] https://www.bucap.it/news/approfondimenti-tematici/gestione-del-magazzino/supply-chain-cos-e.htm#:~:text=Una%20definizione.%20Per%20supply%20chain%20s%E2%80%99intende%3A%20Un%20sistema,prodotto%20o%20un%20servizio%20dal%20fornitore%20al%20cliente.?msclkid=885839dea8d811ec9631cde73dee2725

[2] https://www.entersoftware.it/supply-chain-management/?msclkid=11b61dbaa8d811ecbfc77389e9fd73da

[3] https://paolotofoni.wordpress.com/2008/08/23/storia-della-supply-chain/?msclkid=11e93485a8dd11ecac71a87b41bccec8

[4] https://www.oracle.com/it/scm/what-is-supply-chain-management/?msclkid=11b5a467a8d811eca777dbdd424569cb#link3

[5] https://www.digital4.biz/executive/industria-40-storia-significato-ed-evoluzioni-tecnologiche-a-vantaggio-del-business/

[6] https://www.mise.gov.it/images/stories/documenti/Piano_Industria_40.pdf

[7] https://www.mecalux.it/blog/tecnologie-abilitanti#:~:text=Definizione%20Le%20tecnologie%20abilitanti%20o%20KETs%20%28Key%20Enable,il%20valore%20aggiunto%20necessario%20per%20generare%20vantaggio%20competitivo.?msclkid=c8986d18a9a711eca075451f862779a7

[8] https://www.make-consulting.it/industria-4-0-tecnologie-abilitanti/?msclkid=09e77b07a9a511ec8ced6407d6e2c223

[9] https://www.kilometrorosso.com/news/ladditive-manufacturing-che-da-forma-al-futuro-la-risposta-di-kilometro-rosso-con-lisa-tech/#:~:text=L%E2%80%99Additive%20Manufacturing%2C%20in%20realt%C3%A0%2

C%20a%20differenza%20di%20quanto,risale%20al%201984%20per%20mano%20di%20Chuck%20Hull.?msclkid=9bc12de1b3f211ec91ce8ade4a2c4bc2

[10] https://ahead-research.com/simulazione-digital-twin/?msclkid=acf88cacb3f811ec971ac5052477515e

[11] https://www.focusindustria40.com/integrazione-verticale-e-orizzontale/#:~:text=L%E2%80%99integrazione%20verticale%20invece%2C%20riguarda%20l%E2%80%99internalizzazione%20di%20tutte%20le,rispondere%20pi%C3%B9%20rapidamente%20alle%20nuove%20opportunit%C3%A0%20di%20mercato.?msclkid=a2bc5dc4b40111ec8e6d3b4061e93508

[12] https://www.internet4things.it/industry-4-0/cyber-physical-systems-cps-cosa-sono-come-stanno-rivoluzionando-il-mondo-industriale/

[13] https://blog.osservatori.net/it_it/industrial-iot-definizione-applicazioni?hsLang=it-it

[14] https://www.articolando.net/generici/linternet-of-things-entra-nellarredamento-per-ufficio/#:~:text=L%27Internet%20of%20Things%20che%20entra%20in%20ufficio%20significa,con%20una%20conseguente%20riduzione%20di%20consumi%20e%20costi.?msclkid=bd2b3c91b41311ecae241b24ae51d8c9

[15] https://www.eurosw.it/il-cloud-e-la-logistica-supply-chain/?msclkid=3368945eb41711ec90fa6093d76d03fc

[16] https://www.blumeglobal.com/learning/artificial-intelligence/?msclkid=b830497aaac111ec93c77003b88e6aa6

[17] https://www.ibm.com/it-it/cloud/learn/machine-learning?msclkid=9c5eeb28aac611ecaf8b631df05451b2#toc-casi-di-ut-Cv579DJu

[18] https://www.bucap.it/news/approfondimenti-tematici/gestione-del-magazzino/blockchain-supply-chain.htm?msclkid=4d32cb9daabe11ec9d42b178fe8da821

[19] https://focus.namirial.it/cyber-security-supply-chain/?msclkid=bff34acab42411ecaac3b8bbbe7a93ec

[20] https://enterprise.teamsystem.com/blog/dematerializzazione-e-cloud/digital-supply-chain?msclkid=d6d240d0aaae11ec810114e1425576aa

[21] https://www.finoit.com/blog/digital-supply-chain-in-industry4-0/?msclkid=a5ce251eaaae11ec842dc706d556a0a5

[22] https://www.zerounoweb.it/software/erp-crm-scm/digital-supply-chain-management-da-dove-partire-per-essere-smart/

[23] https://www.ibm.com/it-it/analytics/hadoop/big-data-analytics?msclkid=6c92f6c2b49e11ec933221e1254ca2e0

[24] https://blog.osservatori.net/it_it/le-5v-dei-big-data?msclkid=376f91a4ab4111eca609021ad5762ec7

[25] https://it.wikipedia.org/wiki/Dato#:~:text=Un%20dato%20%28dal%20latino%20datum%20che%20significa%20dono%2C,una%20transazione%2C%20di%20un%20avvenimento%20o%20di%20altro.?msclkid=7d99160db4a311ec94c38bc428c243f9

[26] https://www.purestorage.com/it/knowledge/big-data/structured-vs-unstructured-data.html?msclkid=daf2c1aeab5411eca043cbc59f123a30

[27] https://megamarketing.it/che-cosa-sono-i-dati-semi-strutturati/?msclkid=740b112bb4a811ec9c45b93892661e59

[28] https://www.nextre.it/tipi-di-analisi-dei-big-data/?msclkid=e092ad9bab5811ec8477a6df350fc13d#I_benefici_dellanalisi_dei_big_data

[29] https://ichi.pro/it/spiegazione-dell-analisi-descrittiva-predittiva-e-prescrittiva-273367038492876?msclkid=ca65dfe1ab7911ec938441c9a4d040a2

[30] https://www.talend.com/resources/big-data-supply-chain/?msclkid=0c0f00dbaa9d11ec964604d6255fc0a3

[31] https://www.pragmaetimos.com/big-data-e-green-data-le-basi-digitali-per-la-sostenibilita/?msclkid=a44ad16dab9711ec8198db41b6777921

[32] https://www.greelane.com/it/scienza-tecnologia-matematica/scienze-sociali/snowball-sampling-3026730/?msclkid=a026416baf1d11ecaafd48296d28812d

[33] https://www.mathsly.it/wordpress/systematic-review-10-passi-per-orientarsi-ed-eseguirla-al-meglio/

[34] https://www.treccani.it/vocabolario/bias/?msclkid=7a5fbaf9b57411ec88403498674ce3fc

[35] https://www.bing.com/images/search?view=detailV2&ccid=yIcMoaXq&id=82799D92FEF638DB892F86D046CDD04D77480DDF&thid=OIP.yIcMoaXqWFxPL-WJ_2oRlgHaEI&mediaurl=https%3a%2f%2fwww.unipa.it%2fbiblioteche%2f.content%2fimmagini%2ffare-ricerca%2fbibliometria%2fbibliometria_prospetto_indicatori.png&cdnurl=https%3a%2f%2fth.bing.com%2fth%2fid%2fR.c8870ca1a5ea585c4f2fe589ff6a1196%3frik%3d3w1Id03QzUbQhg%26pid%3dImgRaw%26r%3d0&exph=447&expw=800&q=+indicatori+bibliometrici&simid=607991619404328846&FORM=IRPRST&ck=8252405838CB601CEB98452B86BE6D56&selectedIndex=15&ajaxhist=0&ajaxserp=0

[36] http://biblioteche.unipv.it/home/risorse/indicatori-bibliometrici#:~:text=Gli%20indicatori%20bibliometrici%20vengono%20utilizzati%20nell%E2%80%99ambito%20dell%E2%80%99analisi%20scientometrica,messi%20a%20punto%20applicando%20tecniche%20matematiche%20e%20statistiche.

[37] https://it.wikipedia.org/wiki/Fattore_di_impatto#:~:text=Il%205-year%20Journal%20Impact%20Factor%20%C3%A8%20un%20indicatore,citazioni%20degli%20articoli%20pubblicati%20nei%20cinque%20anni%20precedenti.

[38] https://www.corriere.it/speciali/ospedali_eccellenza/pop_metodo.shtml

[39] https://www.roars.it/online/limpact-factor-nella-valutazione-della-ricerca-e-nello-sviluppo-delleditoria-scientifica/

[40] https://faseb.onlinelibrary.wiley.com/doi/abs/10.1096/fj.08-107938

[41] https://www.internauka.org/en/blog/scopus-journal-quartile-ranking

[42] https://trends.google.it/trends/?geo=IT

[43] https://www.ekuota.com/it/fare-business-in-india/#:~:text=%E2%80%9Cl%E2%80%99India%20ha%20un%20modello%20di%20sviluppo%20solido%20%28%E2%80%A6%29,lavoro%2C%20in%20una%20economia%20sulla%20strada%20dello%20sviluppo%E2%80%9D.

Thanks

I thank my parents for granting me this achievement, for supporting me on this path. A path not always formed by descents, on the contrary, it can be said that this master's degree was formed by many ascents. I thank my parents for accompanying me to the top.

I thank my brother for supporting and putting up with me especially in these last two years. Aware that I haven't always been a perfect roommate, I hope I can always be a perfect sister, life partner and reference point for you.

I thank my Angel. Of course, this is not just your first name. You have been a true angel to me, a point of reference, a support and a companion with whom to share moments of both joy and difficulty. Turin is just one of the many milestones added to our achievements. Yes, that's right, "ours" because this isn't just my accomplishment, it is only by thinking about our future that I have been able to get this far.

I thank Valeria, Anna and Giorgia. Companions of adventure, university colleagues and friends. I was really lucky to have met you, so much so that I think every time that meeting should have happened from day one. Unfortunately, these past two years have divided us a bit, but I hope we can participate in each other's joys in the future too.

I thank Federica, you have been a really wonderful discovery for me. You have often given voice to my thoughts without me expressing them to you, you have strengthened and supported me even at times when I thought I couldn't cope. I think I have never found such a strong understanding with a friend. An understanding that often turned out to be successful. I believe that fate couldn't have hindered our meeting, which occurred right near one of the most difficult obstacles to overcome during our bachelor's degree. This cannot be pure coincidence. I wish us not to lose each other during the life journey ahead.

I thank Rossella. Despite the miles that have divided us, you have always been able to show and make your closeness felt. I thank you for choosing to share this journey of yours with me and for being a friend in addition to a colleague.

Finally, I thank all my friends and relatives who have known me all my life and who, although they haven't experienced my journey closely, have always been a source of courage.