Complexity

Complexity in Industry 4.0 Systems and Networks

Lead Guest Editor: Dimitris Mourtzis
Guest Editors: Nikos Papakostas and Sotiris Makris

Complexity in Industry 4.0 Systems and Networks

Complexity in Industry 4.0 Systems and Networks

Lead Guest Editor: Dimitris Mourtzis
Guest Editors: Nikos Papakostas and Sotiris Makris

Editorial Board

Contents

Hindawi
Complexity
Volume 2019, Article ID 7817046, 2 pages
https://doi.org/10.1155/2019/7817046

Editorial

Complexity in Industry 4.0 Systems and Networks

Dimitris Mourtzis,[1] **Nikolaos Papakostas**,[2] **and Sotiris Makris**[1]

[1] *University of Patras, Patras, Greece*
[2] *University College Dublin, Dublin, Ireland*

Correspondence should be addressed to Dimitris Mourtzis; mourtzis@lms.mech.upatras.gr

Received 29 November 2018; Accepted 2 December 2018; Published 13 January 2019

Complexity has been studied under the perspective of both physical and functional domains. In the former one, it takes into account aspects of the system structure and configuration, the variety of products it can process, and the system's resources, such as humans, machines, buffers, and their interdependencies as well as the behavior of the system. In the latter case, complexity is considered as a measure of uncertainty in achieving functional requirements, taking into account aspects of manufacturing system design, its states, and the degree to which it can handle the variety of demand.

Today, manufacturing enterprises are often facing the challenge of having to manufacture highly customized products in small lot sizes. One solution to react to the rapidly varying demands and make the use of resources more flexible lies in the digitalization of the manufacturing systems; this still remains a major challenge in industry. Information and material flow over the production and distribution are becoming increasingly complex. The advent of Industry 4.0 initiative and the modern technologies, such as Cyber-Physical Systems (CPS) and Internet of Things (IoT), open new horizons towards the industrial digitalization by enabling automated procedures and communication by means that were not available in the past. The increased connectivity and interaction among systems, humans, and machines enable the integration of various automated or semiautomated systems, increasing flexibility and productivity. This will also lead to interconnected Industry 4.0-enabled manufacturing systems and networks that will constitute an integrated whole that delivers the needed products to the right place, at the right time, taking into consideration a number of newly created parameters and factors.

Although the adoption of Industry 4.0 and IoT paradigms in manufacturing reveals great potential, the increased complexity that may occur in different areas, including product, information, machine, shop-floor, and enterprise, as well as at a network level, is a main challenge. The technology foreseen in the context of Industry 4.0 paradigm aims to reduce the complexity of the systems; nevertheless managing the amount of generated data, dealing with the increased number of variables, and integrating different tools are all fields of further investigation. In particular, the way the generated data and information from different sources and levels can be integrated into methods and systems for adaptive and effective decision-making needs to be further analyzed, aiming to manage the different and conflicting decision variables. In Industry 4.0 systems and networks, advanced monitoring techniques and novel automation systems are integrated and the overall degree of complexity is highly affected. Existing approaches and methods, including chaos, nonlinear dynamics, and information theories as well as hybrid approaches (Heuristics-Indexes), need to be further exploited, taking into consideration the increased number of parameters and variables of Industry 4.0 systems and networks.

Therefore, the main objective of this special issue is to collect and consolidate innovative and high-quality research contributions, mainly focusing on methods and tools to model, quantify, and control Industry 4.0 complexity. This special issue provides insights on how Industry 4.0 technologies may support effective decision-making, while reducing systems' and networks' complexity through the submitted scientific contributions in the form of both research and review papers.

This special issue includes 5 original papers selected by the editors, featuring significant research contributions in the above-mentioned topics. The list of papers is as follows.

(1) "Multilayer Network-Based Production Flow Analysis" by T. Ruppert et al.

(2) "On the Design Complexity of Cyberphysical Production Systems" by L. Ribeiro and M. Hochwallner.

(3) "Smart Scheduling: An Integrated First Mile and Last Mile Supply Approach" by T. Bányai et al.

(4) "Topological Structure of Manufacturing Industry Supply Chain Networks" by S. S. Perera et al.

(5) "Green Supplier Selection for Process Industries Using Weighted Grey Incidence Decision Model" by J. Quan et al.

Conflicts of Interest

I hereby declare that neither I nor the other guest editors have any possible conflicts of interest or private agreements with companies which are involved in the submissions in this special issue.

Acknowledgments

We would like to thank all authors who submitted their works for this special issue.

Dimitris Mourtzis
Nikolaos Papakostas
Sotiris Makris

Hindawi
Complexity
Volume 2018, Article ID 4631670, 12 pages
https://doi.org/10.1155/2018/4631670

Research Article

Green Supplier Selection for Process Industries Using Weighted Grey Incidence Decision Model

Jing Quan[1], **Bo Zeng**[2], and **Dai Liu**[2]

[1] *School of Science, Chongqing University of Technology, Chongqing, China*
[2] *Rongzhi College of Chongqing Technology and Business University, Chongqing, China*

Correspondence should be addressed to Bo Zeng; bozeng@ctbu.edu.cn

Received 16 February 2018; Revised 7 May 2018; Accepted 14 September 2018; Published 8 October 2018

Guest Editor: Sotiris Makris

Proper supplier selection to meet production demand is a major aspect of all manufacturing and process industries. Green supplier selection has been one of the most critical factors for environmental protection on account of increasing consumption levels and for sustainable development as well. This paper aims at developing an applicable methodology for green supplier selection for the process industry. In this study, both economic and environmental criteria are considered and a comprehensive weighted grey incidence decision approach for green supplier evaluation and selection in a process industry is proposed. First, an overall green supplier selection index system for process industries is considered; then a weighted grey incidence decision-making model with improved grey incidence coefficients and weighted degree of grey incidence is provided. Improved grey incidence coefficients are defined using transformation sequences of the initial data. To eliminate the ill effects from the use of equal weights, the maximum entropy method is used to determine the weights of the improved grey incidence coefficients. An application example is proposed with the data collected for the chemical processing industry, which provides acceptable results in determining the better supplier. In the end appendix, some theory regarding the weights for grey incidence coefficients is proposed. The empirical results indicate that the model is of great practical value for green supplier selection in the process industry.

1. Introduction

The supply chain is a series of activities to procure materials, process them into intermediate goods and final products, and deliver them to customers, which encompasses all links from suppliers to customers. Supply chain management quests to conduct the physical and information flow exchanged amongst all actors in a supply chain [1] best. Effective supply chains can establish a long-term effective collaboration amongst different businesses. Supplier selection is the process in which companies identify, evaluate, and contract with suppliers [2]. In today's Internet world, competitive business circumstances have caused many companies to elevate the importance of supply chain management and supplier selection, and businesses pay particular attention to identification and selection amongst alternative supply sources. Once a supplier becomes a partner, the relationship between the buyer and supplier will have a critical effect on the competitiveness of the entire supply chain. As most companies expend a considerable amount of their revenues on purchasing, the supplier selection process has become one of the most important issues for establishing an effective supply chain management system [3]. As organizations become more and more dependent on suppliers, the direct and indirect consequences of poor decision-making in selecting suppliers will become more critical [4].

Supplier selection is a complex decision-making process. Before the companies made decisions based almost exclusively on price and quantity, most of the modern researchers consider that the combination of factors should fit not only the economic and technical requirements but also a company's strategy [5]. Ho et al. suggested that global criteria should be applied to effectively evaluate supplier selection [6]. Traditional supplier selection models focus on suppliers' economic and technical efficiency but ignore the ecological efficiency of the supplier. Today, organizations should consider the environmental awareness of suppliers and demand that their suppliers reduce their environmental impacts [7].

Apart from the common criteria such as cost and quality, the research [8] discussed the green issues which can play an important role in sourcing and proposed critical environmental variables which can be used in supplier selection.

With the advent of industrialization, green supply chain management can be regarded as a strategy in which all supply chain members attain more value. Today's competitive markets have forced companies to focus on environmental issues aligned with other critical factors (cost, quality, service level, etc.) to increase the supply chain value. Hence, green supplier selection to reduce the purchase risk is one of the most important decision-making issues. There are many studies dedicated to supplier selection based on conventional criteria [4], while more and more literature [9–14] regarded green supplier evaluation or works which considered environmental criteria.

There are many different ways based on data envelopment analysis, cluster analysis, categorical methods, case-based-reasoning systems, decision models for the final choice phase, linear weighting models, statistical models, mathematical programming models, and artificial intelligence models [15, 16] for supplier selection decision.

Grey theory was proposed by Deng [17–19], and it is a mathematical theory that was derived from the concept of the grey set theory. It is one of the effective methods used to solve uncertainty problems under sparse data and incomplete information. The major advantage of the grey theory is that it is suitable to handle both incomplete information and unclear problems. It is used as an analysis tool when there is not enough data. Grey theory has been widely applied to prediction analysis [20–22] and decision-making [23]. As a significant portion of grey theory, grey incidence analysis has been successfully applied to social, economic, industrial, and agricultural fields and has solved a large number of scientific and production problems [24].

Grey incidence analysis is largely applied in the areas of project selection, prediction analysis, and performance evaluation. Tsai et al. [25] have developed a supplier selection model for the garment industry using grey incidence analysis. Yang and Chen [26] proposed an integrated model by combining the analytical hierarchy process and grey incidence analysis to evaluate and select the supplier for a notebook computer manufacturer. Li et al. [27] used the rough set theory for supplier selection. They converted the linguistic variable assigned to the alternatives into a grey number; based on grey theory, the suppliers were ranked. Wu et al. [28] discussed the use of a grey multiobjective decision-making approach to select a better site for a hypermarket in a Taiwanese city. They compared their findings with the multiplicative competitive interaction model and found that the grey approach yielded the better solution. Cao et al. [29] developed an integrated grey decision model for the selection of the material of a product.

Supplier selection decisions have become one of the most significant responsibilities of managers as well as the most critical and complex issues they must deal with. Furthermore, green supply chain performance measures must be considered in the supplier selection process, and green supplier selection decision-making for the chemical process industry has seldom been studied. Most methods for supplier selection are analytic hierarchy processes [1, 3, 4], which depend on subjective data provided by experts. In [25–29], grey incidence analysis with equal weighting has been used for supplier selection, as it does not require large or regular data samples and hence has an important practical significance. However, their methods require subjective data from experts and equally weighted factors and initial data to form behavioural sequences are needed to calculate the degree of grey incidence in Deng's grey incidence analysis. However, certain grey information cannot be directly obtained, and the correlation coefficients of each sequence at different times are of different importance to the system. Therefore, in this context, a suitable green supplier selection decision-making method for the chemical process industry based on an improved grey incidence model is proposed, in which both economic and environmental criteria for chemical processing industry supplier selection are considered, and improved grey incidence coefficients are defined using transformation sequences of the initial data. The updated weighted grey coefficient optimization model is then established to determine the weights to obtain improved correlation coefficients.

The remainder of this paper is organized as follows: in Section 2, we construct a green supplier selection index system for process industries. The improved weighted grey incidence decision model is presented in Section 3. In Section 4, a practical application of the green supplier evaluation and selection decision method for process industries is demonstrated. The discussion of the advantage of the weighted grey incidence decision model compared to other related methodologies is given in Section 5. In Section 6, some relevant conclusions on the green supplier evaluation and selection decision model are discussed. Finally, the appendix contains the proofs of the mathematical statements in the paper.

2. Green Supplier Selection Index System for Process Industries

In this section, we discuss what criteria should be taken into account in green supplier selection processes. In reviewing the literature, several conventional criteria have been identified by scholars. Ho et al. [6] concluded in a literature review of supplier selection models that the most popular economic criteria amongst researchers are cost, quality, delivery, management, technology, and flexibility. With a focus on environmental awareness, more and more authors are addressing supplier selection in the light of environmental aspects, such as Handfield et al. [10], Lee et al. [13], and Sarkis [30]. Yang et al. [31] first used the Delphi method to determine the critical activities and developed the system of environmental management and green product. Environmental criteria may include pollution production, pollution control, resource consumption, green product, environmental management, ecodesign, green image, green competencies, staff environmental training, and management commitment.

However, a firm should consider both environmental and conventional factors in order to select the most appropriate

Table 1: Green supplier selection criteria.

Number	Criteria	Related attributes
1	Cost	Product price, logistics cost, manufacturing cost, waste disposal cost, cost reduction performance
2	Quality	Product stability, product qualification ratio, low toxicity, process improvement, quality assurance, quality management, quality-related certificates, rejection by customers
3	Delivery	Order processing speed, supplier lead time, delivery time, delivery reliability, delivery delays, waiting time, credible delivery
4	Service	Warranty, responsiveness, stock management, capability of design, capability of technology support, flexibility
5	Technique capability	Manufacturing capacity, technological development, technological compatibility, technology level supplier's speed in development, capability of preventing pollution
6	Green product	Green packaging, recycling, remanufacturing, reuse
7	Pollution control	Solid wastes, waste water, energy consumption, average volume of air pollutants, harmful materials released, use of harmful materials
8	Environmental management	Environment-related certificates such as ISO 14000, green process planning, internal control process, and low carbon measures

supplier for partnership. Existing works have generally considered environmental aspects only, for example, Lee et al. [13] and Kuo et al. [14]. This study intends to present a comprehensive green supplier selection framework by considering both economic and environmental criteria and is conducted to extract both classic and green performance measures. We concentrated on business issues such as cost, quality, delivery, service, and technology to create opportunities for long-term collaboration with suppliers. Green criteria such as pollution control, green product, and environmental management systems are also integrated with the selection criteria for better consideration of the supplier selection process [8]. Having compared different studies in the literature, in this study, we select eight criteria: cost, quality, delivery, service, technique capability, pollution control, green product, and environmental management, as the main factors of the green supplier selection process according to previous research. Table 1 presents the summary of the most important criteria in the research literature.

(1) Cost: cost can be viewed as one of the most important factors for supplier selection decision-making. Proper suppliers can reduce the cost and provide buyers with better competencies in the market. Cost consists of product price, logistics cost, manufacturing cost, inventory charge, energy charge, maintenance expense, inspection expense, and security expense, as well as waste disposal costs as an environmental factor and cost reduction performance

(2) Quality: to improve production quality, the management must consider quality control and process improvement. Quality assurance can meet customer demands for optimal utilization of resources and coincides with a firm's objectives. Total quality management and quality-related certificates such as ISO 9000, EN 29000, and BS 5750 are taken into account. Low toxicity and rejection by customers can also reflect the quality

(3) Delivery: to satisfy the customer market, firms must cooperate with the right supplier, who provides their requirements at the proper time, at the right place, under the right conditions, and with the right service. Order fulfilment capability consists of order fulfilment planning, product execution, distribution management, and cross application integration. Order processing speed, supplier lead time, delivery time, delivery reliability, delivery delays, waiting time, and credible delivery are the main components of lead time

(4) Service: in the 21st century's competitive business environment, companies not only have to try to satisfy the demand of customers for high-quality products at competitive prices but also must operate at a high service level with good stock management and design capability to achieve customer satisfaction. Firms can achieve this goal by fast deliveries, low costs, minimum wastage, quick response, high productivity, low stocks, no damage, few mistakes, high staff morale, and so on

(5) Technique capability: technique capability is the life of a company. Advanced manufacturing technology will assist the enterprise to become a leader in its field. Therefore, manufacturing capabilities and new technology development capabilities of the supplier are necessary to meet the current and future demands of the firm. Additionally, a new product design by the supplier, technological compatibility, capacity, and speed in development should be considered

(6) Green product: there has been a greater focus on a green competency amongst vendors and suppliers, which has strategic value and favours company image in recent years. Green packaging is a type of packaging which aims at protecting the environment by using environment-friendly materials which can be recycled or reused

(7) Pollution control: pollution control is an important reference indicator which influences the choice of a supplier. Suppliers' attitude toward pollution is becoming a determining factor for partnerships. Pollution not only destroys the environment with solid wastes and wastewater but also leads to superfluous energy consumption. Therefore, contamination control can lead to proper usage of energy for suppliers. To reduce pollution in general, the use of harmful materials must be confined

(8) Environmental management: the goals of environmental management systems are to compel firms to reduce the negative impacts of production to the environment and force buyers to be more aware of the environment, influencing firms in their decision-making. Environment-related certificates such as ISO 14000, green process planning, internal control process, and low carbon measures are the main considered indicators of environmental management

In summary, green supplier selection has become an interesting topic to researchers owing to increasing awareness of environmental protection and its long-term effect on business and marketing performance.

3. Proposed Weighted Grey Incidence Decision Model

Grey incidence analysis uses information from the grey system to dynamically compare each factor quantitatively. The relational analysis suggests how to make predictions and decisions and generates reports that offer suggestions for supplier selection. Grey incidence analysis is a method to analyse the relational grades for discrete sequences. This is unlike the traditional statistics analysis handling the relationships between variables. Statistical analysis works with sufficient data, and the data distribution must be typical. However, grey incidence analysis requires fewer data and can analyse many factors that can overcome the disadvantages of the statistics method.

The following preliminaries can be found in [23, 24]. In system analysis, one first chooses the quantity which reflects the characteristics of a system best. Afterwards, the factors that influence the behaviour of the system are determined. Then a quantitative analysis is considered; subsequently, one needs to process the chosen characteristic quantities, and the available data are converted to relevant nondimensional values of roughly equal magnitude by using sequence operators. The main procedure of grey incidence analysis first translates the performance of all alternatives into a comparability sequence. This step is called *grey relational generating*.

Definition 1 [24]. Assume that X_i is a factor of a system, and its observation value at the ordinal position k is $x_i(k)$, where $k = 1, 2, \cdots, n$. Then, $X_i = (x_i(1), x_i(2), \cdots, x_i(n))$ is called a behavioural sequence of factor X_i. If k stands for the time order, then $x_i(k)$ represents the observational value of factor X_i at the time moment k, and X_i is called a behavioural time sequence. If k is an ordinate of some criteria and $x_i(k)$ is the observation of the factor x_i at the criteria k, then X_i is called a behavioural criterion sequence. If k is the ordinal number of the object observed and $x_i(k)$ stands for the observation of the factor X_i of the kth objective, then X_i is called a behavioural horizontal sequence.

Regardless of the sequence (time, criterion, or horizontal), the required incidence analysis can always be conducted.

Definition 2 [23]. Let $X_i = (x_i(1), x_i(2), \cdots, x_i(n))$ be a behavioural sequence of factor X_i, where $i = 1, 2, \cdots, m$. D is a sequence operator such that $X_iD = (x_i(1)D, x_i(2)D, \cdots, x_i(n)D)$, where $x_i(k)D = (x_i(k)/x_i(1))$, $x_i(1) \neq 0$, and $k = 1, 2, \cdots, n$. Then, D is referred to as an initializing operator, and X_iD is the initial image of X_i..

According to these sequences, a reference sequence (ideal target sequence) can be defined.

Definition 3 [24]. Assume that $X_0 = (x_0(1), x_0(2), \cdots, x_0(n))$ is a sequence of data representing a system characteristic, and for all i, X_i is a sequence of relevant factors. For a given real number $\xi \in (0, 1)$, let

$$\begin{aligned} \gamma_{0i}(k) &= \gamma(x_0(k), x_i(k)) \\ &= \frac{\min\limits_j \min\limits_r |x_0(r) - x_j(r)| + \xi \max\limits_j \max\limits_r |x_0(r) - x_j(r)|}{|x_0(k) - x_i(k)| + \xi \max\limits_j \max\limits_r |x_0(r) - x_j(r)|}, \\ \gamma(X_0, X_i) &= \frac{1}{n}\sum_{k=1}^{n} \gamma(x_0(k), x_i(k)). \end{aligned} \tag{1}$$

Then, $\gamma(X_0, X_i)$ is called the degree of grey incidence of X_i with respect to X_0, $\gamma(x_0(k), x_i(k))$ is the incidence coefficient of X_i with respect to X_0 at point k, and ξ is the distinguishing coefficient.

The grey incidence coefficients between all comparability sequences and the reference sequence are then calculated. Based on these grey incidence coefficients, the grey incidence grade between the reference sequence and every comparability sequence is calculated. If a comparability sequence translated from an alternative has the highest grey incidence grade between the reference sequence and itself, that alternative will be the best choice.

In this study, the procedure used in the improved weighted grey incidence model for green supplier selection consists of eight steps: (I) determination of selection index data, (II) generation of referential series, (III) normalization of the data set, (IV) construction of the differential information data, (V) set up of the grey incidence coefficient, (VI) determination of the weight of grey incidence coefficient, (VII) calculation of the degree of grey incidence, and (VIII) determination of the grades of suppliers. The supplier with the highest degree of grey incidence is considered the optimal supplier.

3.1. Step I: Determination of Selection Index Data. Green supplier selection is based on the critical factors comprising the selection index: cost, supply quality, lead time to delivery, warranty and capacity of the supplier, green product, pollution control, and environmental management. The decision criteria may be represented by $C_1, C_2, \cdots, C_n$. Experts were also asked to consider industry specifications within the decision-making process. The structure of the decision problem consists of m alternatives: supplier 1, supplier 2, ..., supplier m. Each alternative can be evaluated in terms of the decision criteria in both qualitative and quantitative terms.

$$\begin{cases} X_1 = (x_1(1), x_1(2), \cdots, x_1(n)), \\ X_2 = (x_2(1), x_2(2), \cdots, x_2(n)), \\ \cdots, \\ X_m = (x_m(1), x_m(2), \cdots, x_m(n)). \end{cases} \tag{2}$$

3.2. Step II: Generation of Referential Series X_0. The referential series X_0 is the optimal value for each criterion, where $x_0(k)$ is some observation value, and it may be the largest or the smallest of $\{x_1(k), x_2(k), \cdots, x_m(k)\}$. When the larger value of some criteria is better, we let $x_0(k)$ be the largest of $\{x_1(k), x_2(k), \cdots, x_m(k)\}$. When the smaller data of some criteria is better, we let $x_0(k)$ be the smallest of $\{x_1(k), x_2(k), \cdots, x_m(k)\}$. X_0 is used to compare with the value of all alternative suppliers.

3.3. Step III: Normalization of Data Set. The main idea of grey incidence is to determine whether the relationship between sequence curves is close according to the similarity of their shape. The aim is to display the correlation between factors quantitatively. Incidence analysis is the basis for grey system analysis and forecasting. To obtain grey information of the factors quantitatively, we perform the following grey transformation. The series data in this case can be normalized using one of two approaches; for a criteria whose larger value is better (quality, warranty and capacity), it can be normalized using formula (3a), while for a criteria whose smaller value is better (cost and delivery), it can be normalized using formula (3b).

$$y_i(k) = \frac{x_0(k)}{x_i(k)}, \quad \text{if } x_0(k) \text{ is the smallest one of } x_i(k), i = 1, 2, \cdots, m, k = 1, 2, \cdots, n, \tag{3a}$$

$$y_i(k) = \frac{x_i(k)}{x_0(k)}, \quad \text{if } x_0(k) \text{ is the biggest one of } x_i(k), i = 1, 2, \cdots, m, k = 1, 2, \cdots, n. \tag{3b}$$

Then, we can obtain the normalization sequences:

$$\begin{cases} Y_1 = (y_1(1), y_1(2), \cdots, y_1(n)), \\ Y_2 = (y_2(1), y_2(2), \cdots, y_2(n)), \\ \cdots, \\ Y_m = (y_m(1), y_m(2), \cdots, y_m(n)). \end{cases} \tag{4}$$

3.4. Step IV: Construction of the Differential Information Data. Now, we construct the differential information of grey incidence as follows:

$$\Delta_{0i}(k) = \frac{1}{1 + \lambda |1 - y_i(k)|}, \quad i = 1, 2, \cdots, m, k = 1, 2, \cdots, n, \tag{5}$$

where $\lambda > 0$ is some fixed number and $y_i(k)$ is defined by (3a) and (3b). Thus, we obtain the differential information sequences:

$$\begin{cases} \Delta_{01} = \left(\frac{1}{1 + \lambda |1 - y_1(1)|}, \frac{1}{1 + \lambda |1 - y_1(2)|}, \cdots, \frac{1}{1 + \lambda |1 - y_1(n)|} \right), \\ \Delta_{02} = \left(\frac{1}{1 + \lambda |1 - y_2(1)|}, \frac{1}{1 + \lambda |1 + y_2(2)|}, \cdots, \frac{1}{1 + \lambda |1 - y_2(n)|} \right), \\ \cdots, \\ \Delta_{0m} = \left(\frac{1}{1 + \lambda |1 - y_m(1)|}, \frac{1}{1 + \lambda |1 - y_m(2)|}, \cdots, \frac{1}{1 + \lambda |1 - y_m(n)|} \right). \end{cases} \tag{6}$$

3.5. Step V: Set-Up of Grey Incidence Coefficients. Assume that $X_0 = (x_0(1), x_0(2), \cdots, x_0(n))$ is a sequence of data representing a system's characteristics, and for all i, X_i is a sequence of relevant factors. For a given real number $\xi \in (0, 1)$, let

$$\gamma_{0i}(k) = \frac{\min_j \min_r \Delta_{0j}(r) + \xi \Delta_{0i}(k)}{\min_j \min_r \Delta_{0j}(r) + \xi \max_j \max_r \Delta_{0j}(r)}, \tag{7}$$

where $\Delta_{0i}(k)$ is defined by (5); then $\gamma_{0i}(k)$ is called the incidence coefficient of X_i with respect to X_0 at point k, where ξ is called the distinguishing coefficient. Therefore, we obtain the grey incidence coefficient sequences:

$$\begin{cases} \gamma_{01} = (\gamma_{01}(1), \gamma_{01}(2), \cdots, \gamma_{01}(n)), \\ \gamma_{02} = (\gamma_{02}(1), \gamma_{02}(2), \cdots, \gamma_{02}(n)), \\ \cdots, \\ \gamma_{0m} = (\gamma_{0m}(1), \gamma_{0m}(2), \cdots, \gamma_{0m}(n)). \end{cases} \tag{8}$$

3.6. Step VI: Determination of the Weight of Grey Incidence Coefficients. The basis of the grey incidence model is proximity and similar factors to determine the degree of incidence coefficient according to the curve. Past researches take equal weighted incidence coefficients for the degree of grey incidence at each time point. However, a serial incidence coefficient at different time points does not have the same influence on system behaviour. Therefore, the similarity of systematic development characterized by the degree of grey incidence is related to the weight of each point. Each factor has its own contribution in the evaluation. The weight value of a time point is the influence coefficient of the system. The larger the weight value is, the more important the coefficient is. Usually, the weight value depends on the decision maker's subjectivity, which may result in some errors. How to arrange the decision weights should be studied in the decision-making process. Determining the weight of incidence coefficient should cause the distribution density value of weighted grey incidence coefficients to balance; therefore, according to Proposition 1 and Theorem 1 in the appendix, we define the maximum constraint of weighted grey incidence entropy, and the weight vector is

$$\varpi_{0i} = \Gamma_{0i}{}^{-1} b, \tag{9}$$

where

$$\Gamma_{0i} = \begin{pmatrix} \Delta_{0i}(1) & -\Delta_{0i}(2) & & & & \\ & \Delta_{0i}(2) & -\Delta_{0i}(3) & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & \Delta_{0i}(n-1) & -\Delta_{0i}(n) \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix},$$

$$\varpi_{0i} = \begin{pmatrix} \varpi_{0i}(1) \\ \varpi_{0i}(2) \\ \cdots \\ \varpi_{0i}(n) \end{pmatrix},$$

$$b = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix}. \tag{10}$$

3.7. Step VII: Calculation of the Degree of Grey Incidence. Assume that $X_0 = (x_0(1), x_0(2), \cdots, x_0(n))$ is a sequence of data representing a system's characteristics and X_i are sequences of relevant factors, and let

$$\gamma(X_0, X_i) = \sum_{k=1}^{n} \varpi_{oi}(k)\gamma_{0i}(k), \tag{11}$$

where $\sum_{k=1}^{n} \varpi_{0i}(k) = 1$ and $\Delta_{0i}(k)$ are defined by (5). Then $\gamma(X_0, X_i)$ is called the weighted degree of grey incidence of X_i with respect to X_0.

3.8. Step VIII: Determination of the Grades of Suppliers. Based on the degree of grey incidence, the priority of the m suppliers is determinate, and the optimal supplier who has the maximal degree is selected.

4. Practical Application of Green Supplier Selection Decision Maker for Process Industries

In this section, we consider a practical application of green supplier selection decision maker for some chemical processing industry. The business of the chemical processing industry is to change the chemical structure of natural materials in order to derive products of value to other industries or in daily life. Chemicals are produced from these raw materials, principally minerals, metals, and hydrocarbons in a series of processing steps. Further treatment, such as mixing and blending, is often required to convert them into end-products (e.g., paints, adhesives, medicines, and cosmetics). It is known that the supplier selection of raw materials or natural materials is very important to the chemical processing industry. Surely, the cost is one of the most important factors for supplier selection. For quality, delivery, service, and technique capability, many suppliers can meet the requirements of the chemical processing industry. However, green production, pollution control, and environmental management may influence product environmental quality safety, even the development prospect of the enterprise. The data for this case study have been collected from the procurement department, which is accountable for the whole purchasing processes at a given chemical processing industry. An interview has been conducted for the purpose of data collection. Seven suppliers have responded with a quotation representing the prescribed product with their attributes. The proposed weighted grey incidence decision model is used to deal with small samples, and the data of supplier selection for the chemical processing industry is just a sample. Therefore, the methodology weighted grey incidence decision model can be used to select the green suppliers for process industries. Applying the proposed model to supplier selection decision implies that all the steps above have to be followed.

Table 2: Alternatives versus criteria.

Supplier	C1	C2	C3	C4	C5	C6	C7	C8
Supplier 1	2439	86%	12	4	170	0.10	10	6.9
Supplier 2	2567	90%	12	3	260	0.08	13	7.0
Supplier 3	2711	93%	14	3	280	0.11	12	7.2
Supplier 4	3099	94%	10	2	260	0.10	11	6.7
Supplier 5	2302	89%	13	2.5	290	0.09	12	7.0
Supplier 6	3115	92%	12	3	250	0.09	11	6.9
Supplier 7	2800	90%	14	2	250	0.08	10	7.0

(1) Determination of selection index data: the structure of the decision problem considered in this study consists of seven alternatives, e.g., suppliers 1–7. For brevity, we only select the following indexes: product price (yuan), product qualification ratio (%), delivery time (days), warranty (months), manufacturing capacity (kg per day), remanufacturing (%), harmful materials released (kg), and low carbon measures (g) as the decision criteria, and they are compiled as C_1, C_2, C_3, C_4, C_5, C_6, C_7, and C_8, respectively. Table 2 shows the alternatives versus the criteria matrix for all seven suppliers

(2) Generation of referential series X_0: the referential series X_0 is the optimal value for each criterion, and $x_0(k)$ is some observation value; it may be the largest

Table 3: Normalized data.

Supplier	$y_i(k)$	C1	C2	C3	C4	C5	C6	C7	C8
Supplier 1	$y_1(k)$	0.9438	0.9149	0.8333	1.0000	0.5862	0.9091	1.0000	0.9710
Supplier 2	$y_2(k)$	0.8968	0.9574	0.8333	0.7500	0.8966	0.7273	0.7692	0.9571
Supplier 3	$y_3(k)$	0.8491	0.9894	0.7143	0.7500	0.9655	1.0000	0.8333	0.9306
Supplier 4	$y_4(k)$	0.7428	1.0000	1.0000	0.5000	0.8966	0.9091	0.9091	1.0000
Supplier 5	$y_5(k)$	1.0000	0.9468	0.7692	0.6250	1.0000	0.8182	0.8333	0.9571
Supplier 6	$y_6(k)$	0.7390	0.9787	0.8333	0.7500	0.8621	0.8182	0.9091	0.9710
Supplier 7	$y_7(k)$	0.8221	0.9574	0.7143	0.5000	0.8621	0.7273	1.0000	0.9571

or the smallest of $\{x_1(k), x_2(k), \cdots, x_7(k)\}$. For indexes C_1, C_3, C_7, and C_8, the smallest is the best, and for indexes C_2, C_4, C_5, and C_6, the largest is the best. Therefore, $X_0 = (2302, 0.94, 10, 4, 290, 0.11, 10, 6.7)$ are used to compare the values of each corresponding supplier

(3) Normalization of data set: according to the above referential series X_0 and formulas (3a) and (3b)

$$\begin{aligned} y_i(k) &= \frac{x_0(k)}{x_i(k)}, \quad \text{if } x_0(k) \text{ is the smallest one of} \\ & \qquad x_i(k), i = 1, 2, \cdots, 7, k = 1, 2, \cdots, 8, \\ y_i(k) &= \frac{x_i(k)}{x_0(k)}, \quad \text{if } x_0(k) \text{ is the biggest one of} \\ & \qquad x_i(k), i = 1, 2, \cdots, 7, k = 1, 2, \cdots, 8, \end{aligned} \tag{12}$$

the series data in this case can be normalized. The normalized data are tabulated in Table 3

(4) Construction of the differential information data: from the normalized data and formula (5)

$$\Delta_{0i}(k) = \frac{1}{1 + \lambda |1 - y_i(k)|}, \quad i = 1, 2, \cdots, 7, k = 1, 2, \cdots, 8, \tag{13}$$

we take that $\lambda = 1.5$ is some fixed number, we calculate the differential information data, and the differential information data are tabulated in Table 4

(5) Set up of grey incidence coefficient: according to the differential information data and formula (7)

$$\gamma_{0i}(k) = \frac{\min\limits_j \min\limits_r \Delta_{0j}(r) + \xi \Delta_{0i}(k)}{\min\limits_j \min\limits_r \Delta_{0j}(r) + \xi \max\limits_j \max\limits_r \Delta_{0i}(r)}, \tag{14}$$

the grey incidence coefficient is calculated and tabulated in Table 5

(6) Determination of the weight of grey incidence coefficient: according to the differential information data and formula (9) $\varpi_{0i} = \Gamma_{0i}{}^{-1} b$, where

$$\Gamma_{0i} = \begin{pmatrix} \Delta_{0i}(1) & -\Delta_{0i}(2) & & & & \\ & \Delta_{0i}(0) & -\Delta_{0i}(3) & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & \Delta_{0i}(6) & -\Delta_{0i}(7) \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix},$$
$$\varpi_{0i} = \begin{pmatrix} \varpi_{0i}(1) \\ \varpi_{0i}(2) \\ \cdots \\ \varpi_{0i}(7) \end{pmatrix},$$
$$b = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix}, \tag{15}$$

we calculate the weights of grey incidence coefficients and tabulated in Table 6

(7) Calculation of the degree of grey incidence: according to the weights of grey incidence coefficients based on the grey relational coefficients and formula (11)

$$\gamma(X_0, X_i) = \sum_{k=1}^{8} \varpi_{0i}(k) \gamma_{0i}(k), \tag{16}$$

the degree of the grey incidence is calculated and the values are tabulated in Table 7

(8) Grade of the suppliers: based on the degree of the grey incidence, the optimal supplier was selected. The priorities of the seven suppliers were supplier 1 > supplier 3 > supplier 4 > supplier 5 > supplier

Table 4: Differential information data.

$\Delta_{0i}(k)$	C1	C2	C3	C4	C5	C6	C7	C8
$\Delta_{01}(k)$	0.9223	0.8868	0.8000	1.0000	0.6170	0.8800	1.0000	0.9583
$\Delta_{02}(k)$	0.8659	0.9400	0.8000	0. 7273	0.8657	0.7097	0.7428	0.9396
$\Delta_{03}(k)$	0.8155	0.9843	0.7000	0.7273	0.9508	1.000	0.8000	0.9057
$\Delta_{04}(k)$	0.7216	1.0000	1.0000	0.5714	0.8657	0.8800	0.8800	1.0000
$\Delta_{05}(k)$	1.0000	0.9261	0.7428	0.6400	1.0000	0.7857	0.8000	0.9396
$\Delta_{06}(k)$	0.7186	0.9691	0.8000	0.7273	0.8286	0.7857	0.8800	0.9583
$\Delta_{07}(k)$	0.7894	0.9400	0.7000	0.5714	0.8286	0.7097	1.0000	0.9396

Table 5: Grey incidence coefficient.

$\gamma_{0i}(k)$	C1	C2	C3	C4	C5	C6	C7	C8
$\gamma_{01}(k)$	0.9505	0.9279	0.8727	1.0000	0.7562	0.9236	1.0000	0.9735
$\gamma_{02}(k)$	0.9147	0.9618	0.8727	0.8264	0.9145	0.8152	0.8364	0.9616
$\gamma_{03}(k)$	0.8826	0.9900	0.8091	0.8264	0.9687	1.0000	0.8727	0.9400
$\gamma_{04}(k)$	0.8228	1.0000	1.0000	0.7273	0.9145	0.9236	0.9236	1.0000
$\gamma_{05}(k)$	1.0000	0.9530	0.8364	0.7709	1.0000	0.8636	0.8727	0.9616
$\gamma_{06}(k)$	0.8210	0.9803	0.8727	0.8264	0.8909	0.8636	0.9236	0.9735
$\gamma_{07}(k)$	0.8660	0.9618	0.8091	0.7273	0.8909	0.8152	1.0000	0.9616

Table 6: The weights of grey incidence coefficients.

$\varpi_{0i}(k)$	C1	C2	C3	C4	C5	C6	C7
$\varpi_{01}(k)$	0.1171	0.1217	0.1350	0.1080	0.1750	0.1227	0.1080
$\varpi_{02}(k)$	0.1176	0.1084	0.1273	0.1400	0.1177	0.1435	0.1371
$\varpi_{03}(k)$	0.1297	0.1075	0.1511	0.1455	0.1113	0.1058	0.1323
$\varpi_{04}(k)$	0.1449	0.1046	0.1046	0.1830	0.1208	0.1188	0.1188
$\varpi_{05}(k)$	0.1040	0.1123	0.1400	0.1625	0.1040	0.1323	0.1323
$\varpi_{06}(k)$	0.1433	0.1063	0.1288	0.1416	0.1243	0.1311	0.1171
$\varpi_{07}(k)$	0.1243	0.1044	0.1402	0.1717	0.1184	0.1383	0.0981

Table 7: Degree of grey incidence.

Supplier	$\gamma(X_0, X_i)$	Weighted degree of grey incidence
Supplier 1	$\gamma(X_0, X_1)$	0.9133
Supplier 2	$\gamma(X_0, X_2)$	0.8821
Supplier 3	$\gamma(X_0, X_3)$	0.9023
Supplier 4	$\gamma(X_0, X_4)$	0.8960
Supplier 5	$\gamma(X_0, X_5)$	0.8954
Supplier 6	$\gamma(X_0, X_6)$	0.8880
Supplier 7	$\gamma(X_0, X_7)$	0.8633

6 > supplier 2 > supplier 7. Therefore, supplier 1 is the optimal supplier because it has the maximal degree of 0.9133 in Table 7

Through the actual calculation results of the proposed model for green supplier selection decision, we see that the results are most up to the mustard for green consideration. The optimal values of the green factors green production, pollution control, and environmental management are 0.11, 10, and 6.7, respectively, and the corresponding factors of an optimal supplier, “supplier 1,” are 0.10, 10, and 6.9, respectively. Therefore, this proposed model can be applied to make the optimal green supplier select decision.

5. Discussion

Most of the methods for supplier selection processes rely on subjective index data of experts, such as TOPSIS technique, analytic hierarchy process, and analytic network process. However, the weighted grey incidence decision model proposed in this manuscript depends on the objective data collected from the procurement department. These data used for supplier selection are more authentic and effective than scores by experts. Equally weighted factors and initial data sequences are used for calculating the degree of grey incidence in classical Deng’s grey incidence analysis model. However, certain grey information cannot be directly obtained, and the correlation coefficients of each sequence at different times are of different importance to the system. Therefore, in this study, improved grey incidence coefficients using transformation sequences of the initial data and the maximum entropy method for determining the weights of improved grey incidence coefficients are proposed. Some grey information can be obtained more easily by using the grey transformation sequences, so the proposed model distinguished the new degree of grey incidence more effectively. The model was successfully used in the analysis of green supplier selection decision in the chemical processing industry. The examples demonstrate that the method is practical and reliable; it is also effective for decision-making in some other applied fields.

6. Conclusion

This paper presents an improved model for grey incidence to fix the weighted grey incidence coefficients for green supplier selection in process industries. New grey incidence coefficients are created by transformation sequences of the initial data. The weights are determinate in the objective entropy information method, which reflect the actual data. The method is based on the difference of index data to produce determinate weights. The proposed approach is suitable for decision-making under more uncertain environments. This paper proposes an evaluation method to determine the overall performance for each candidate supplier. The optimum decision can then be made based on the supplier's overall rating. The use of the proposed model can facilitate company-level decision-making by using a systematic approach to select the appropriate supplier.

Appendix

In this section, some mathematical statements and their proofs are listed for convenience and content integrity. The following can be found in [32–35].

The distribution of the weighted grey incidence coefficients is defined below.

Definition 4. The distribution of the weighted grey incidence coefficients is defined so that

$$p_k = \frac{\varpi_{0i}(k)\Delta_{0i}(k)}{\sum_{k=1}^{n}\varpi_{0i}(k)\Delta_{0i}(k)} \tag{17}$$

is its density value.

According to [32, 33], $\{p_k\}$ is a grey sequence.

Definition 5. The function

$$H_{\otimes}(R_i) = -\sum_{k=1}^{n} p_k \ln p_k \tag{18}$$

is called the weighted grey incidence entropy of X_i, where $\forall k, p_k \geq 0, \sum_{k=1}^{n} p_k = 1$.

The following proposition holds.

Proposition 1. *[34, 35] If the density sequence of the distribution of the weighted grey incidence coefficient is $p = \{p_k \mid \forall k, p_k \geq 0, \sum_{k=1}^{n} p_k = 1\}$ and $H \otimes (R_i)$ is the incidence entropy, then $H \otimes (R_i)$ is maximised when p_k are equal.*

Corresponding to a uniform distribution, the optimal weights of the incidence coefficients yield the maximum weighted grey incidence entropy by Proposition 1.

Theorem 1. *[34, 35] The solution to the problem*

$$\begin{aligned} \max \quad & H_{\otimes}(R_i) = -\sum_{k=1}^{n} p_k \ln p_k \\ \text{s.t.} \quad & \sum_{k=1}^{n} \varpi_{0i}(k) = 1, \quad \varpi_{0i}(k) > 0, \\ & p_k = \frac{\varpi_{0i}(k)\Delta_{0i}(k)}{\sum_{k=1}^{n}\varpi_{0i}(k)\Delta_{0i}(k)} \end{aligned} \tag{19}$$

is

$$\begin{aligned} p_k &= \frac{1}{n}, \\ \varpi_{0i} &= \Gamma_{0i}^{-1} b, \end{aligned} \tag{20}$$

where

$$\begin{aligned} \Gamma_{0i} &= \begin{pmatrix} \Delta_{0i}(1) & -\Delta_{0i}(2) & & & & \\ & \Delta_{0i}(2) & -\Delta_{0i}(3) & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & \Delta_{0i}(n-1) & -\Delta_{0i}(n) \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix}, \\ \varpi_{0i} &= \begin{pmatrix} \varpi_{0i}(1) \\ \varpi_{0i}(2) \\ \cdots \\ \varpi_{0i}(n) \end{pmatrix}, \\ b &= \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix}. \end{aligned} \tag{21}$$

Proof. The Lagrange function of problem (19) is

$$L(p_k, \lambda) = -\sum_{k=1}^{n} p_k \ln p_k + \lambda\left(\sum_{k=1}^{n} p_k - 1\right). \tag{22}$$

The necessary condition for the existence of extreme values implies that

$$\begin{cases} \dfrac{\partial L}{\partial p_k} = -\ln p_k + \lambda - 1 = 0, & k = 1, 2, \cdots, n, \\ \dfrac{\partial L}{\partial \lambda} = \sum_{k=1}^{n} p_k - 1 = 0. \end{cases} \tag{23}$$

Hence, $p_k = e^{\lambda-1} (k = 1, 2, \cdots, n)$. This equal and $\sum_{k=1}^{n} p_k = 1$ lead to $p_k = 1/n$. From (17) and (19), we obtain

$$\begin{cases} \dfrac{\varpi_{0i}(k)\Delta_{0i}(k)}{\sum_{k=1}^{n}\varpi_{0i}(k)\Delta_{0i}(k)} = \dfrac{1}{n}, \\ \sum\limits_{k=1}^{n} \varpi_{0i}(k) = 1. \end{cases} \tag{24}$$

The above can be expressed as

$$\begin{cases} \varpi_{0i}(1)\Delta_{0i}(1) - \varpi_{0i}(2)\Delta_{0i}(2) = 0, \\ \varpi_{0i}(2)\Delta_{0i}(2) - \varpi_{0i}(3)\Delta_{0i}(3) = 0, \\ \cdots, \\ \varpi_{0i}(n-1)\Delta_{0i}(n-1) - \varpi_{0i}(n)\Delta_{0i}(n) = 0, \\ \sum\limits_{k=1}^{n} \varpi_{0i}(k) = 1. \end{cases} \tag{25}$$

And the matrix form is $\Gamma_{0i}\varpi_{0i} = b$, where

$$\Gamma_{0i} = \begin{pmatrix} \Delta_{0i}(1) & -\Delta_{0i}(2) & & & & \\ & \Delta_{0i}(2) & -\Delta_{0i}(3) & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & \Delta_{0i}(n-1) & -\Delta_{0i}(n) \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{pmatrix},$$
$$\varpi_{0i} = \begin{pmatrix} \varpi_{0i}(1) \\ \varpi_{0i}(2) \\ \cdots \\ \varpi_{0i}(n) \end{pmatrix},$$
$$b = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ 1 \end{pmatrix}. \tag{26}$$

The determinant of Γ_{0i} is

$$|\Gamma_{0i}| = \begin{vmatrix} \Delta_{0i}(1) & -\Delta_{0i}(2) & & & & \\ & \Delta_{0i}(2) & -\Delta_{0i}(3) & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & & \Delta_{0i}(n-1) & -\Delta_{0i}(n) \\ 1 & 1 & 1 & 1 & 1 & 1 \end{vmatrix}. \tag{27}$$

We add the $(n-1)$th row to the $(n-2)$th row, add the $(n-2)$th row to the $(n-3)$th row, until we add the third row to the second row, and add the second row to the first row; then we have

$$|\Gamma_{0i}| = \begin{vmatrix} \Delta_{0i}(1) & & & & & -\Delta_{0i}(n) \\ & \Delta_{0i}(2) & & & & -\Delta_{0i}(n) \\ \cdots & \cdots & \cdots & \cdots & \cdots & -\Delta_{0i}(n) \\ & & & & -\Delta_{0i}(n-1) & -\Delta_{0i}(n) \\ 1 & 1 & 1 & 1 & 1 & 1 \end{vmatrix}. \tag{28}$$

Now, for $k = 1, 2, \ldots, n-1$, we add the kth column multiplied by $(\Delta_{0i}(k))/(\Delta_{0i}(k))$ to the nth column, and we obtain

$$|\Gamma_{0i}| = \begin{vmatrix} \Delta_{0i}(1) & & & & & \\ & \Delta_{0i}(2) & & & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & & \Delta_{0i}(n-1) & \\ 1 & 1 & 1 & 1 & 1 & 1 + \sum\limits_{k=1}^{n-1} \dfrac{\Delta_{0i}(n)}{\Delta_{0i}(k)} \end{vmatrix}, \tag{29}$$

which is

$$\|\Gamma_{0i}\| = \prod_{k=1}^{n-1} \Delta_{0i}(k) \left(1 + \prod_{k=1}^{n-1} \frac{\Delta_{0i}(n)}{\Delta_{0i}(k)}\right). \tag{30}$$

From (3a) and (3b), we know that $\|\Gamma_{0i}\| \neq 0$. Therefore, the system of the linear equations has a unique solution. Hence, the weight vector is $\varpi_{0i} = \Gamma_{0i}{}^{-1} b$.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors' work was supported by the Chongqing Municipal Social Science Planning Commission Project of China (2016WT37), Science and Technology Research Project of Chongqing Education Commission (KJ1601902), and the open research fund of Chongqing Key Laboratory of Electronic Commerce Supply Chain System (1456026).

References

[1] A. Aguezzoul, "Overview on supplier selection of goods versus 3PL selection," *Journal of Logistics Management*, vol. 1, no. 3, pp. 18–23, 2012.

[2] D. Beil, "Supplier selection," in *Encyclopedia of Operations Research and Management Science*, Wiley, New York, NY, USA, 2010.

[3] R. Singh, H. Rajput, V. Chaturvedi, and J. Vimal, "Supplier selection by technique of order preference by similarity to ideal solution (TOPSIS) method for automotive industry," *International Journal of Advanced Technology & Engineering Research*, vol. 2, no. 2, pp. 157–160, 2012.

[4] F. T. S. Chan, N. Kumar, M. K. Tiwari, H. C. W. Lau, and K. L. Choy, "Global supplier selection: a fuzzy-AHP approach," *International Journal of Production Research*, vol. 46, no. 14, pp. 3825–3857, 2008.

[5] S. Hossein Cheraghi, M. Dadashzadeh, and M. Subramanian, "Critical success factors for supplier selection: an update," *Journal of Applied Business Research*, vol. 20, no. 2, pp. 91–108, 2011.

[6] W. Ho, X. Xu, and P. K. Dey, "Multi-criteria decision making approaches for supplier evaluation and selection: a literature review," *European Journal of Operational Research*, vol. 202, no. 1, pp. 16–24, 2010.

[7] A. Kumar, V. Jain, and S. Kumar, "A comprehensive environment friendly approach for supplier selection," *Omega*, vol. 42, no. 1, pp. 109–123, 2014.

[8] O. Gurel, A. Z. Acar, I. Onden, and I. Gumus, "Determinants of the green supplier selection," *Procedia-Social and Behavioral Sciences*, vol. 181, pp. 131–139, 2015.

[9] G. Noci, "Designing green vendor rating systems for the assessment of a suppliers environmental performance," *European Journal of Purchasing & Supply Management*, vol. 3, no. 2, pp. 103–114, 1997.

[10] R. Handfield, S. V. Walton, R. Sroufe, and S. A. Melnyk, "Applying environmental criteria to supplier assessment: a study in the application of the analytical hierarchy process," *European Journal of Operational Research*, vol. 141, no. 1, pp. 70–87, 2002.

[11] P. Humphreys, R. McIvor, and F. Chan, "Using case-based reasoning to evaluate supplier environmental management performance," *Expert Systems with Applications*, vol. 25, no. 2, pp. 141–153, 2003.

[12] P. K. Humphreys, Y. K. Wong, and F. T. S. Chan, "Integrating environmental criteria into the supplier selection process," *Journal of Materials Processing Technology*, vol. 138, no. 1-3, pp. 349–356, 2003.

[13] A. H. I. Lee, H. Y. Kang, C. F. Hsu, and H. C. Hung, "A green supplier selection model for high-tech industry," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7917–7927, 2009.

[14] R. J. Kuo, Y. C. Wang, and F. C. Tien, "Integration of artificial neural network and MADA methods for green supplier selection," *Journal of Cleaner Production*, vol. 18, no. 12, pp. 1161–1170, 2010.

[15] L. de Boer, E. Labro, and P. Morlacchi, "A review of methods supporting supplier selection," *European Journal of Purchasing & Supply Management*, vol. 7, no. 2, pp. 75–89, 2001.

[16] K. Govindan, S. Rajendran, J. Sarkis, and P. Murugesan, "Multi criteria decision making approaches for green supplier evaluation and selection: a literature review," *Journal of Cleaner Production*, vol. 98, pp. 66–83, 2015.

[17] D. Ju-Long, "Control problems of grey systems," *Systems & Control Letters*, vol. 1, no. 5, pp. 288–294, 1982.

[18] J. L. Deng, "Grey information space," *Journal of Grey System*, vol. 1, no. 2, pp. 103–117, 1989.

[19] J. L. Deng, "Introduction to grey system theory," *The Journal of Grey System*, vol. 1, pp. 1–24, 1989.

[20] B. Zeng, Y. Tan, H. Xu, J. Quan, L. Wang, and X. Zhou, "Forecasting the electricity consumption of commercial sector in Hong Kong using a novel grey dynamic prediction model," *Journal of Grey System*, vol. 30, no. 1, pp. 157–172, 2018.

[21] B. Zeng, H. Duan, Y. Bai, and W. Meng, "Forecasting the output of shale gas in China using an unbiased grey model and weakening buffer operator," *Energy*, vol. 151, pp. 238–249, 2018.

[22] B. Zeng and S. F. Liu, "A self-adaptive intelligence gray prediction model with the optimal fractional order accumulating operator and its application," *Mathematical Methods in the Applied Sciences*, vol. 23, no. 1, pp. 1–15, 2017.

[23] P. Golinska, M. Kosacka, R. Mierzwiak, and K. Werner-Lewandowska, "Grey decision making as a tool for the classification of the sustainability level of remanufacturing companies," *Journal of Cleaner Production*, vol. 105, no. 15, pp. 28–40, 2015.

[24] S. F. Liu, Y. J. Yang, and L. F. Wu, *Grey System Theory and Its Application*, Beijing: Science Press, 2014.

[25] C. H. Tsai, C. L. Chang, and L. Chen, "Applying grey relational analysis to the vendor evaluation model," *International Journal of The Computer, The Internet and Management*, vol. 11, no. 3, pp. 45–53, 2003.

[26] C. C. Yang and B. S. Chen, "Supplier selection using combined analytical hierarchy process and Grey relational analysis," *Journal of Manufacturing Technology Management*, vol. 17, no. 7, pp. 926–941, 2006.

[27] G.-D. Li, D. Yamaguchi, and M. Nagai, "A grey-based rough decision-making approach to supplier selection," *International Journal of Advanced Manufacturing Technology*, vol. 36, no. 9-10, pp. 1032–1040, 2008.

[28] W.-Y. Wu, C.-J. Bai, and O. K. Gupta, "A hypermarket site selection model using the grey multi-objective decision method," *International Journal of Logistics Systems and Management*, vol. 2, no. 1, pp. 68–77, 2006.

[29] H. J. Cao, F. Liu, C. B. Li, and C. Liu, "An integrated method for product material selection considering environmental factors and a case study," *Materials Science Forum*, vol. 532-533, pp. 1032–1035, 2006.

[30] J. Sarkis, *Greening the Supply Chain*, Springer, London, 2006.

[31] C. C. Yang, K. J. Yang, and S. Y. Peng, "Exploration strategies and key activities for the system of environmental management," *Total Quality Management & Business Excellence*, vol. 22, no. 11, pp. 1179–1194, 2011.

[32] Q. S. Zhang, X. J. Guo, and J. L. Deng, "Grey relation entropy method of grey relation analysis," *Systems Engineering-Theory & Practice*, vol. 8, no. 8, pp. 7–11, 1996.

[33] Q. S. Zhang, *Grey Difference Theory of Hazy Set*, The Press of Petroleum Industry, Beijing, 2002.

[34] L. L. Pei, W. M. Chen, J. F. Cui, and G. B. Zhuang, "Weighted degree of grey incidence based on optimization entropy," in *2009 IEEE International Conference on Grey Systems and Intelligent Services (GSIS 2009)*, pp. 10–12, Nanjing, China, 2009.

[35] J. Quan, B. Zeng, and L. Y. Wang, "Maximum entropy methods for weighted grey incidence analysis and applications," *Grey Systems: Theory and Application*, vol. 8, no. 2, pp. 144–155, 2018.

Hindawi
Complexity
Volume 2018, Article ID 3924361, 23 pages
https://doi.org/10.1155/2018/3924361

Research Article

Topological Structure of Manufacturing Industry Supply Chain Networks

Supun S. Perera [1], **Michael G. H. Bell** [1], **Mahendrarajah Piraveenan** [2], **Dharshana Kasthurirathna** [3], **and Mamata Parhi** [4]

[1] *Business School Institute of Transport and Logistics Studies (ITLS), University of Sydney, Australia*
[2] *School of Civil Engineering Complex Systems Research Group (CSRG), University of Sydney, Australia*
[3] *Faculty of Computing, Sri Lanka Institute of Information Technology (SLIIT), Sri Lanka*
[4] *Roehampton Business School, University of Roehampton, UK*

Correspondence should be addressed to Supun S. Perera; supun.perera@sydney.edu.au

Received 19 February 2018; Revised 29 April 2018; Accepted 8 August 2018; Published 3 October 2018

Academic Editor: Nikos Papakostas

Empirical analyses of supply chain networks (SCNs) in extant literature have been rare due to scarcity of data. As a result, theoretical research have relied on arbitrary growth models to generate network topologies supposedly representative of real-world SCNs. Our study is aimed at filling the above gap by systematically analysing a set of manufacturing sector SCNs to establish their topological characteristics. In particular, we compare the differences in topologies of undirected contractual relationships (UCR) and directed material flow (DMF) SCNs. The DMF SCNs are different from the typical UCR SCNs since they are characterised by a strictly tiered and an acyclic structure which does not permit clustering. Additionally, we investigate the SCNs for any self-organized topological features. We find that most SCNs indicate disassortative mixing and power law distribution in terms of interfirm connections. Furthermore, compared to randomised ensembles, self-organized topological features were evident in some SCNs in the form of either overrepresented regimes of moderate betweenness firms or underrepresented regimes of low betweenness firms. Finally, we introduce a simple and intuitive method for estimating the robustness of DMF SCNs, considering the loss of demand due to firm disruptions. Our work could be used as a benchmark for any future analyses of SCNs.

1. Introduction

Due to the increasingly complex and interconnected nature of the global supply chain networks (SCNs), recent research has focussed on modelling supply chains as complex adaptive systems using network science concepts [1, 2]. Following on from the work published by [3], who used network science techniques to generate a network topology and investigate its topological robustness, a large number of theoretical research papers have appeared in this area [4–12]. These studies have theoretically formulated generalizable growth mechanisms underlying the firm-partnering process in SCN formation. Subsequently, the network topologies generated based on various growth models have been investigated for their topological characteristics, such as robustness and efficiency.

Despite the large number of theoretical papers published within the past few years on network modelling of SCNs, the effort on empirical validation of the theoretical findings has been limited. This is mainly due to difficulty in obtaining large-scale datasets on supplier-customer relationships, which are often proprietary and confidential. Papers which systematically analyse the topologies of real-world SCNs, the conclusions of which can then be used to inform modelling efforts, have been relatively scarce.

In light of the above, this study presents a comprehensive analysis of two sets of SCN datasets, namely,

(1) The dataset of Indian automobile manufacturers [13]. This SCN includes contractual relationships between various firms (therefore it is modelled as an undirected network).

(2) Twenty-six SCNs across various manufacturing industry sectors, based on the dataset presented in Willems [14]. These SCNs include material flows between firms (as such, it is modelled as a directed network). All the SCNs in this dataset include the full depth in terms of tiers (from suppliers to retailers).

In particular, this study is aimed at addressing the following key research questions pertaining to SCN topology and robustness:

(1) What common topological characteristics (if any) can be expected from the directed material flow SCNs in the manufacturing sector?

(2) What are the key differences (if any) in the topology of directed material flow and undirected contractual relationship SCNs?

(3) Are there any self-organized features present in any of the SCNs?

(4) Are there any correlations between node attributes and the topological features of nodes in the directed material flow SCNs?

(5) How can we determine the robustness of SCNs, considering the inherent differences between the contractual relationship and material flow networks?

The remainder of this manuscript is structured as follows. Section 2 provides the background to this study and introduces key theoretical concepts in terms of network topological analysis. Section 3 describes the structure and limitations of each dataset considered, while Section 4 presents the data analysis methodology and results. Section 5 provides a discussion of the results obtained, and Section 6 concludes the paper.

2. Background

2.1. Topology of Undirected and Directed SCNs. The interfirm relationships in SCNs are commonly modelled using undirected links. However, the links between nodes in a SCN can include a direction, depending on the specific type of relationship being modelled. The interfirm relationships in a SCN can be broadly categorised into three classes, namely, (1) material flows, (2) financial flows, and (3) information exchanges. Material flows are usually unidirectional from suppliers to retailers, while financial flows are unidirectional in the opposite direction. Both material and financial flows mostly occur vertically, across the functional tiers of a SCN (however, in some cases, two firms within the same tier, such as two suppliers, could also exchange material and finances) [15]. In contrast, information exchanges are bidirectional (i.e., undirected) and include both vertical and horizontal connections (i.e., between firms across tiers and between firms within the same tier). Therefore, the same SCN can include different topologies based on the specific type of relationship denoted by the links in the model. For instance, unlike material and financial flows, SCN topology for information exchanges can exhibit shorter path lengths and high clustering due to a relatively larger number of horizontal connections [16].

Compared to undirected network representation, in directed networks, the adjacency matrix is no longer symmetric. As a result, the degree of a node in a directed network is characterised by both in-degree and out-degree. On this basis, the degree distribution of directed networks is analysed separately for in- and out-degrees. Also, unlike undirected networks, in directed networks the distance between node i and node j is not necessarily the same as the distance between node j and node i. In fact, in directed networks, the presence of a path from node i to node j does not necessarily imply the presence of a path from node j to node i [17]. This has implications on node centrality metrics, such as closeness and betweenness. In addition, many dynamics such as synchronizability and percolation are different in directed networks compared to undirected networks [18, 19]. Therefore, when modelling SCNs, it is important to first identify the specific type of relationship denoted by the links, so that the network can be correctly represented as undirected or directed.

In this paper, we consider two types of SCNs, namely, (1) the SCN of Indian automobile manufacturers, which is modelled as an undirected network (referred to as the undirected contractual relationship SCN or UCR-SCN throughout this paper) and (2) twenty-six material flow SCNs across various manufacturing industry sectors, based on the dataset presented in Willems [14], which are modelled as directed networks (referred to as the directed material flow SCN or DMF-SCN throughout this paper).

2.2. Characterising SCNs Using Network Science Metrics. Mathematical analysis of a network requires it to be represented through an adjacency matrix (A). An element A_{ij} of the adjacency matrix A, for an undirected network, is given as

$$A_{ij} = \begin{cases} 1 & \text{if there is a link between nodes } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

In contrast, an element of the adjacency matrix, for a directed network, is given as

$$A_{ij} = \begin{cases} 1 & \text{if there is a link } from \text{ node } j \text{ to node } i, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The key difference between the adjacency matrix of an undirected and a directed network is that the adjacency matrix of an undirected network has two entries for each link (i.e., the adjacency matrix of undirected networks is symmetric). Therefore, the total number of links (L) for an undirected network is calculated as $L = 1/2\sum_{ij} A_{ij}$, while for a directed network it is calculated as $L = \sum_{ij} A_{ij}$.

In this paper, we have modelled all SCNs as unweighted networks where nodes represent individual firms. For the UCR-SCN, the links represent undirected contractual relationships between firms, while in the DMF-SCNs, the links represent the directed material flows between firms.

The metrics used to characterise the topology of complex networks can be classified into node- and network-level metrics (see Costa et al. [20] and Rubinov and Sporns [21] for a comprehensive range of measurements used for characterization of complex networks). Table 5 in Appendix A presents the list of network-level metrics used for analysis in this study, and their implications within a SCN context.

Node-level metrics characterise, in various ways, the importance of a particular node for the functionality of the overall network, based on its embedded position in the broader relationship network (known as the centrality of the node). Depending on the context, various centrality measures can be adopted to identify the key players of a given network. Table 6 in Appendix A presents the list of node-level metrics used in this study, and their implications within a SCN context.

2.3. Data-Driven Studies. Even though most theoretical modelling efforts of SCNs have focussed on variants of preferential attachment to generate network topologies for analysis [22], there have been some studies which have adopted a data-driven approach. For instance, Kim et al. [23] have undertaken a node- and network-level topological analysis, using three case studies of automotive supply networks (namely, Honda Accord, Acura CL/TL, and Daimler Chrysler Grand Cherokee) presented by Choi and Hong [24]. Although the SCNs used in this study are complete, the SCNs are small in size (largest network includes only 34 firms), which limits the observations of emergent network topological properties. Kito et al. [25] have constructed a SCN for Toyota using the data available within an online database operated by MarkLines Automotive Information Platform. By analysing the SCN topology, the authors have identified the tier structure of Toyota to be barrel-shaped, in contrast to the previously hypothesized pyramidal structure. Another fundamental observation reported in this study is that Toyota SCN topology was found to be not scale-free (even with finite-size effects taken into account). Although the dataset used in this study is sufficiently large (with 3109 firms), it is limited to only the top three tiers of the overall SCN.

More recently, using Bloomberg data, Brintrup et al. [26] and Orenstein [27] have undertaken topological analysis of various SCNs. Brintrup et al. [26] have studied the SCN of Airbus and have reported that this SCN displays assortative mixing and communities based on geographic locations of the firms. Orenstein [27] has undertaken topological analysis of retail and food industry SCNs by considering the suppliers within the top three tiers. The SCNs considered in this study were found to have scale-free topologies with degree exponents below 2. Although the dataset used in this study is sufficiently large and allows observation of temporal variations to the SCN topology, consideration of only a part of the SCN depth in terms of tiers has limited the generalizability of the results. It is noted that the key limitation in using the Bloomberg database, for constructing SCNs, is that the data are not exhaustive since the database only includes publicly listed firms. Therefore, the SCNs constructed using Bloomberg data may only provide a part of the full picture.

Although the above studies have provided a number of key insights about the topological structure of various SCNs, no study to date has systematically investigated a large collection of directed material flow SCNs in various industry sectors and compared the results against an undirected SCN of contractual (interfirm) relationships.

By considering a collection of twenty-six SCNs from the manufacturing industry, our study is able to investigate and establish the general topological properties of these SCNs. This effort will complement the large body of theoretical literature on modelling SCN topologies through various growth models, by revealing what specific topological characteristics are needed to be captured in an appropriate growth model. In addition, the correlation analysis presented in this study, between various node-level centrality measures and two exogenous factors (stage cost and stage time), can be powerful in demonstrating how the position of firms can influence the overall functionality of the SCN (and vice versa).

Finally, this study has used the reliable dataset provided in published work by Willems [14] in relation to the DMF-SCNs. The data collection procedure for the UCR-SCNs is detailed in Section 3.1.1. Our study offers distinct insights from previous studies because (i) it is based on a large collection of real-world SCNs belonging to the manufacturing sector, (ii) most of the SCNs are large (have a relatively high number of nodes) so that various emergent properties can be sufficiently demonstrated, (iii) we extract and compare the topological properties of SCNs where links represent either the directed material flows or undirected contractual relationships, (iv) our correlation analysis of node centrality measures with exogenous factors in DMF-SCNs provides insights into the impact of the position of firms in the dynamics of SCNs, and (v) we investigate the self-organization in all SCNs, which helps us identify nonrandom features of these systems. Thus, this study is unique on several counts from previous studies described above.

2.4. Generation of Null Models for Hypothesis Testing. An important question when testing hypothesis related to network topologies is whether the degree distribution on its own is sufficient to describe the structure of a network, i.e., whether the topological features observed in the network are explained by the ensembles of networks generated by its degree distribution while preserving the degree vector. In this regard, degree-preserving randomisation (DPR) plays an important role in generation of null models.

DPR involves rewiring the original network, to generate an ensemble of null models, while preserving the degree vector [28, 29]. At each time step, the DRP process randomly picks two connected node pairs and switches their link targets. This switching is repeatedly applied to the entire network until each link is rewired at least once. The resulting network represents a null model where each node still has the same degree, yet the paths through the network have been randomised. Comparison of properties of a given network, with the properties of an ensemble of networks generated by DPR, allows one to identify if the properties observed in the real network are unique and meaningful or whether they are common to all networks with that degree sequence [30].

In this study, we use DPR to investigate the presence of self-organized topological features in the undirected contractual relationships SCN. For the DMF-SCNs, we introduce a novel DPR procedure to generate null models by preserving both the degree of nodes and the numbers of links present between adjacent tiers.

2.5. Robustness of SCNs. Past studies have gained insights into the robustness of SCNs by using (1) analytical measurements available in network science, such as network centralisation, percolation threshold, and assortativity, or (2) simulations to investigate how various network-level metrics are affected when nodes are sequentially removed, either randomly (known as random failures) or based on degree (known as targeted attacks) [3, 8, 11, 12].

While the above methods may be suitable for obtaining high-level insights into the robustness structure of UCR-SCNs, a more specific method needs to be formulated for assessing the robustness of tiered DMF-SCNs. In particular, this method should consider both the topological structure of the network and its ability to meet the consumer demands. In this regard, we introduce a simple and intuitive method to investigate the robustness of DMF-SCNs, considering the demand at the retailer nodes.

3. Data Sources, Structure, and Limitations

3.1. Data Sources and Structure

3.1.1. Undirected Contractual Relationship SCN. The undirected contractual relationship SCN consists of customer-supplier contractual linkages in the Indian automotive industry. The network is constructed from the citations of the autocomponent firms about the firms they offer their products and services to (in this regard, a firm could be a supplier to one firm and a customer of another firm). The list of firms is taken from the Automotive Component Manufacturers Association (ACMA) of India annual publication “Buyers Guide.” The data corresponds to the year 2001–2002.

The basic topological analysis of this data (excluding the robustness analysis and self-organization analysis presented in this paper) was presented in Parhi [13]. Although this data was originally created from a 2002 directory, it was compared with recent records as a part of a study titled “Dynamics of Distribution and Diffusion of New Technology” [31] and the authors demonstrated that there is no qualitative difference in the industry structure compared to the 2002 dataset. In fact, the actors and their topography in the supply chain in the Indian Automotive Industry has remained more or less the same over the years (note that this dataset only includes the organized sector of the Indian automotive industry), where the only change has occurred with regard to the dynamics of the interactions, thanks to the fast integration of the information system in a highly integrated world.

3.1.2. Directed Material Flow SCNs. Willems [14] provides a dataset of real-world multiechelon supply chains, used for inventory optimization purposes. The overall dataset includes a total of 38 multiechelon supply chains, from various industries. The chains described in this paper comprise actual supply chain maps created by either company analysts or consultants. Since these maps have been implemented in practice, they demonstrate how users have modelled actual supply chains.

The above-mentioned dataset includes the following key information:

(1) The industry sector of each supply chain network

(2) For each supply chain network:

  (i) The stages (nodes) representing each firm involved

  (ii) The arcs (links) representing precedence relationship between stages

(3) For each stage:

  (i) Its classification and tier based on its function within the overall supply chain

  (ii) The direct cost added at the stage (stage cost)

  (iii) The average processing time at the stage (stage time)

  (iv) The average daily demand at each retailer stage

Note that in the network models developed in this paper, we denote stages (i.e., firms) as nodes and the arcs between stages (which represent the precedence relationships between firms) as directed links.

From the original dataset, networks with more than 100 firms (i.e., nodes) were selected for our analysis, and there were twenty-eight such large networks. Smaller networks were omitted in this analysis since they do not offer any interesting insights into emergence of various complex topological features. Then, using the industry sector information, these SCNs were categorised into six main groups as illustrated in Table 1. As can be seen, the set of SCNs considered by us vary in size (with a minimum of 108 to a maximum of 2025 nodes).

3.2. Limitations of the Datasets

3.2.1. Undirected Contractual Relationship SCN. The SCN data for the Indian automotive industry covers the organized sector of the industry, as there is no comprehensive database on the unorganized segment. The list of customers includes

Table 1: Classification of directed material flow SCNs considered in the study.

Reference	Industry group	Subindustry classification	Number of tiers	Total nodes	Total links
1	Aircraft engines and engine parts	N/A	6	468	605
2		N/A	4	2025	16,225
3	Arrangement of transportation of freight and cargo	N/A	4	116	119
4		N/A	5	626	632
5	Chemical	Soap and other detergents, except specialty cleaners	4	133	164
6		Perfumes, cosmetics, and other toilet preparations	7	186	359
7		Pharmaceutical preparations	9	253	253
8		Paints, varnishes, lacquers, enamels, and allied products	3	271	524
9		Primary batteries, dry and wet	5	617	753
10		Perfumes, cosmetics, and other toilet preparations	5	844	1685
11		Perfumes, cosmetics, and other toilet preparations	8	976	1009
12		Industrial organic chemicals	4	1479	2069
13	Computer-related	Semiconductors and related devices	2	108	452
14		Computer peripheral equipment	5	152	211
15		Computer peripheral equipment	8	154	224
16		Computer peripheral equipment	9	156	263
17		Computer peripheral equipment	10	156	169
18		Computer storage devices	8	577	2262
19	Electrical	Electromedical and electrotherapeutic apparatus	7	145	224
20		Power-driven handtools	6	334	1245
21		Electromedical and electrotherapeutic apparatus	5	482	941
22		Telephone and telegraph apparatus	3	1206	4063
23		Electromedical and electrotherapeutic apparatus	6	1386	1857
24	Farm machinery and equipment	N/A	3	409	853
25		N/A	4	706	908
26		N/A	4	1451	4812

only the firms that have been self-reported by the autocomponent firms as their principal customers.

3.2.2. Directed Material Flow SCNs. A key limitation of the available dataset is the lack of information in relation to the geographical locations of individual firms. This information was not provided in the original dataset in Willems [14] due to confidentiality reasons. Unlike the virtual networks (such as WWW or social networks), the SCN structure is largely influenced by geographical aspects (since the congregation or dispersion of the suppliers depends on the raw material distribution over various geographic regions). Therefore, if geographic location information was available, in-depth conclusions could have been made about various observed structural features of the SCNs.

In addition, this study is unable to investigate the dynamic nature of the SCNs since the dataset does not provide any information pertaining to temporal changes in the SCN topology. Lastly, the relationship strength between firms is not captured in the dataset in terms of the amount material flow. Although specific production capabilities of firms within each tier are known, no information is available in relation to how much each upstream firm supplies to the downstream firms.

Nevertheless, the size of the dataset, both in terms of the number of networks available and the size of each network, as well as the cost and time data associated with nodes make this a very attractive dataset to study.

4. Data Analysis Methodology and Results

4.1. Topological Results. Using each dataset, we constructed SCNs, where the nodes represent the individual firms and the links represent either the undirected contractual

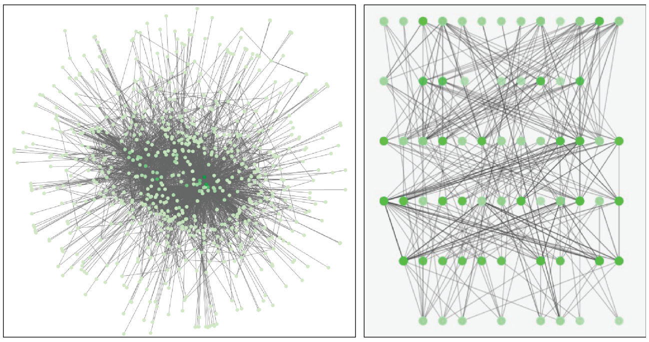


Figure 1: Visualisation of (1) the UCR-SCN (left), and (2) the tiered DMF-SCN in reference 6 (right). The shade of green is set proportional to the node degree.

Table 2: Basic topological features of the undirected contractual relationship SCN.

Topological feature	Value
No. of nodes	687
No. of links	3283
Average degree	9.557
Network centralisation	0.296
Network density	0.014
Network heterogeneity	1.815
Network diameter	7
Characteristic path length	3.07
Clustering coefficient	0.065
Assortativity (degree)	−0.292
γ (R^2 for power law)	2.31 (0.95)

relationships or the directed material flows between firms (see Figure 1). Cytoscape software and JAVA programming language were used to visualise and analyse the SCNs. The results are presented in the following subsections.

4.1.1. Undirected Contractual Relationship SCN. Table 2 presents the network-level topological features of the UCR-SCN of Indian automobile manufacturers.

4.1.2. Directed Material Flow SCNs. Table 3 presents the topological features of the 26 directed material flow SCNs. The three key observations that can be made from the table are that (i) the degree distributions of the majority of SCNs indicate good R^2 correlations against the power law fit, and most of them have degree exponents in the range of 1 to 3, (ii) most SCNs are disassortative, in terms of degree as well as stage cost and stage time, and (iii) none of the SCNs indicated clustering (not presented in the table), due to the links only being present between functional tiers. These observations have important implications, which are discussed in more detail in Section 5.

In addition, for this dataset, we have investigated the node-level centrality metrics. Since node-level metrics themselves provide information about individual nodes rather than networks as a whole, here we choose to primarily study correlations between them and other node-level attributes available in the dataset. In particular, we studied the correlation coefficients between the centrality metrics for each node (namely, in- and out-degree centrality, betweenness centrality, and closeness centrality) and its corresponding (1) stage cost and (2) stage time. The correlation plots are presented in Appendix B. The results of this assessment are discussed in detail in Section 5.

4.2. Identifying Self-Organized Topological Features Using Degree-Preserving Randomisation. DPR can help establish whether or not the observed topological property in a network is simply an artefact of the network's inherent structural properties or a property unique to the nodes. Comparison of the original betweenness and closeness centrality distributions with the average distributions of the same metrics, for an ensemble of randomised (through DPR) networks, can reveal whether the centrality distributions observed in the original network are structural or not. In particular, if the original (i.e., observed) and randomised trends are identical, then the centrality distributions observed in the original network are purely structural, i.e., they can entirely be explained by the degree distribution without attribution to any other external mechanism. However, if the original centrality distribution diverges from the average trend obtained for the randomised ensembles, there is an underlying mechanism which induces this deviation in the original network [17].

Although the divergence between the observed centrality distribution and the randomised ensemble average is generally found through visual inspection of the plots [17, 32], in this study we have used the Kolmogorov-Smirnov test (KS test) to establish any statistically significant deviations. The KS test is a powerful statistical test that allows one to compare two distributions (the null hypothesis of the test is that there exists no difference between the two distributions). Importantly, the KS test does not make any assumption about the underlying distribution of the data (i.e., it is a nonparametric and distribution-free test statistic), thus allowing comparisons between arbitrary distributions.

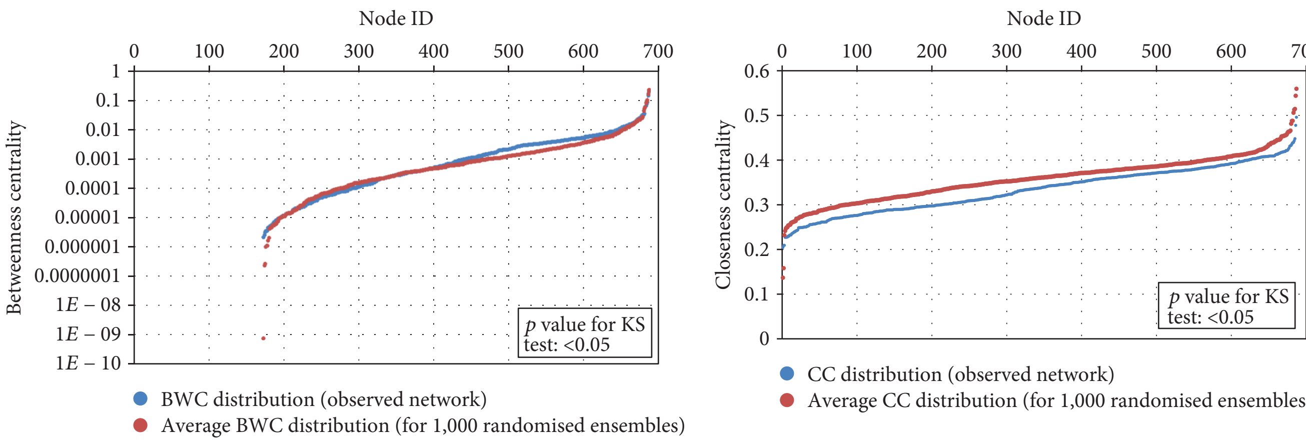
4.2.1. Undirected Contractual Relationship SCN. For the UCR-SCN, we applied DPR where, at each time step, two connected node pairs are picked and their links targets switched. This switching is repeatedly applied to the entire network until every link is rewired at least once (without allowing the creation of self-loops or multilinks between node pairs). The resulting network represents a null model where each node still has the same degree, yet the paths through the network have been randomised. Using this process, we generated 1000 randomised networks. We have then compared the betweenness and closeness centrality distributions of the original network with the average betweenness and closeness values obtained for the 1000 randomised networks (see Figure 2).

Table 3: Summary of results obtained for each DMF-SCN analysed.

Reference	Industry group	Subindustry classification	Average degree	Network centralisation	Network density	Network heterogeneity	Network diameter	Characteristic path length	Assortativity (degree)	Assortativity (cost)	Assortativity (time)	γ for in-degree (R^2 for power law)	γ for out-degree (R^2 for power law)
1	Aircraft Engines and Engine Parts	N/A	2.585	0.201	0.0006	3.032	5	1.812	−0.464	0.217	−0.336	1.75 (0.94)	5.27 (0.89)
2		N/A	16.025	0.111	0.008	2.453	3	2.383	−0.722	−0.553	−0.98	1.86 (0.44)	1.76 (0.98)
3	Arrangement of Transportation of Freight and Cargo	N/A	2.052	0.07	0.018	0.991	3	1.834	−0.055	−0.109	−0.169	3.50 (1.00)	2.33 (0.88)
4		N/A	2.019	0.074	0.003	1.167	4	2.043	0.063	−0.05	0.025	8.48 (1.00)	2.47 (0.86)
5	Chemical	Soap and Other Detergents, Except Specialty Cleaners	2.466	0.196	0.019	1.089	3	1.806	0.031	−0.4	0.203	2.47 (0.92)	2.36 (0.96)
6		Perfumes, Cosmetics, and Other Toilet Preparations	3.86	0.083	0.021	1.061	6	2.825	0.251	0.13	0.055	2.15 (0.82)	2.16 (0.79)
7		Pharmaceutical Preparations	2	0.128	0.0008	1.68	8	3.464	−0.249	0.474	0.053	9.98 (1.00)	2.14 (0.91)
8		Paints, Varnishes, Lacquers, Enamels, and Allied Products	3.867	0.079	0.014	1.205	2	1.434	−0.313	−0.161	−0.071	2.20 (0.70)	2.63 (0.99)
9		Primary Batteries, Dry and Wet	2.441	0.037	0.004	1.506	4	2.867	−0.644	−0.622	−0.122	2.95 (0.92)	2.56 (0.82)
10		Perfumes, Cosmetics, and Other Toilet Preparations	3.993	0.074	0.005	1.56	4	1.964	−0.137	−0.029	−0.03	2.40 (0.79)	2.17 (0.89)
11		Perfumes, Cosmetics, and Other Toilet Preparations	2.068	0.019	0.0002	0.973	7	3.328	0.002	−0.019	0.024	3.01 (0.85)	3.40 (0.89)
12		Industrial Organic Chemicals	2.798	0.031	0.0002	1.573	3	2.271	−0.059	−0.186	−0.458	2.89 (0.72)	2.80 (0.93)
13	Computer Related	Semiconductors and Related Devices	8.37	0.739	0.078	1.828	1	1	−0.821	−0.756	−0.816	1.43 (0.61)	2.62 (0.57)
14		Computer Peripheral Equipment	2.776	0.203	0.018	1.717	4	3.084	−0.181	0.194	−0.224	2.96 (0.98)	1.95 (0.97)
15		Computer Peripheral Equipment	2.909	0.067	0.019	1.021	7	2.621	0.155	−0.036	0.057	2.43 (0.84)	2.28 (0.92)
16			3.372	0.161	0.022	1.279	8	2.754	−0.254	−0.215	−0.022	2.10 (0.90)	2.08 (0.88)

Table 3: Continued.

Reference	Industry group	Subindustry classification	Average degree	Network centralisation	Network density	Network heterogeneity	Network diameter	Characteristic path length	Assortativity (degree)	Assortativity (cost)	Assortativity (time)	γ for in-degree (R^2 for power law)	γ for out-degree (R^2 for power law)
		Computer Peripheral Equipment											
17		Computer Peripheral Equipment	2.167	0.195	0.014	1.647	9	3.096	−0.283	−0.168	−0.109	2.04 (0.85)	4.91 (1.00)
18		Computer Storage Devices	7.841	0.342	0.014	2.47	7	2.561	−0.47	−0.031	−0.752	1.81 (0.89)	2.98 (0.93)
19	Electrical	Electromedical and Electrotherapeutic Apparatus	3.09	0.084	0.021	0.922	6	3.493	−0.08	−0.091	−0.293	2.49 (0.99)	2.36 (0.83)
20		Power-Driven Handtools	7.455	0.183	0.022	1.828	5	2.697	−0.604	−0.031	−0.241	1.64 (0.79)	2.30 (0.62)
21		Electromedical and Electrotherapeutic Apparatus	3.905	0.18	0.008	2.675	4	1.963	−0.614	−0.62	−0.584	1.72 (0.82)	3.17 (0.86)
22		Telephone and Telegraph Apparatus	6.738	0.124	0.006	2.75	2	1.075	−0.726	0.115	−0.553	1.71 (0.58)	2.65 (0.90)
23		Electromedical and Electrotherapeutic Apparatus	2.68	0.04	0.002	2.057	5	2.793	−0.293	−0.037	−0.195	1.88 (0.74)	3.54 (0.94)
24	Farm Machinery and Equipment	N/A	4.171	0.339	0.01	1.911	2	1.538	−0.122	0.023	−0.122	2.46 (0.67)	2.24 (0.95)
25		N/A	2.572	0.039	0.004	1.782	3	2.647	−0.805	−0.256	−0.707	2.64 (0.94)	1.96 (0.58)
26		N/A	6.633	0.401	0.005	2.683	3	1.66	−0.106	−0.047	−0.008	1.91 (0.53)	2.34 (0.84)



(a) (b)

FIGURE 2: (a) Betweenness centrality distribution of the observed SCN against the average betweenness centrality distribution obtained from 1000 randomised ensembles. (b) Closeness centrality distribution of the observed SCN against the average closeness centrality distribution obtained from 1000 randomised ensembles.

for ***each*** *tier pair* T_i, T_{i+1} *in the network, from* $i = 0$ to $i = n - 1$ **do.**
  $//n$ = number of tiers
  for ***each*** *link pair between* T_i, T_{i+1}, *that is not rewired* **do.**
    Ramdomly pick two links (a, b) and (c, d) between T_i, T_{i+1};
    /* The following condition ensures the uniqueness of. the two links selected, in terms of source and. destination nodes */.
    if $a = c$ or $c = d$ **then.**
      | Continue to pick another link pair;
    /* The following condition prevents creation of. multi-links between node pairs */.
    if *Link* (a, d) *or link* (c, b) *already exists* **then.**
      | Continue to pick another link pair;
    Remove links (a, b) and (c, d) and create links (a, d) and (c, b);

ALGORITHM 1: Modified degree-preserving randomizing algorithm for tiered networks.

4.2.2. Directed Material Flow SCN. For the DMF-SCNs, we have developed a modified DPR procedure, which we refer to as the tier constrained DPR (TC-DPR). At each time step, this process picks a pair of links which lie across the same two tiers and swaps their target nodes. In particular, the following algorithm has been used.

The above process is repeatedly applied to the original network until every link is rewired at least once. By choosing link pairs between the same two tiers and swapping their targets, in addition to preserving the degree of nodes, we also preserve the number of intertier links. Therefore, the resulting network is a null model whose degree distribution and tier structure are identical to the original SCN.

Using the TC-DPR process, we generated 1000 randomised networks. We have then compared the betweenness and closeness centrality distributions of the original network with the average betweenness and closeness values obtained for the 1000 randomised network ensembles. The p values obtained for the KS test for each SCN is presented in Table 4. Figures 3–6 illustrate some scenarios where self-organization in terms of betweenness and closeness was identified.

4.3. Robustness Analysis. In this section, we investigate the robustness of the UCR-SCN. Subsequently, a simple and intuitive methodology for assessing the robustness of tiered DMF-SCNs with demand considerations is introduced.

4.3.1. Undirected SCN. Since the UCR-SCN dataset does not include any node attributes, we resort to the generic topological robustness analysis technique commonly employed in network science literature. In particular, we remove nodes, either randomly (to simulate random failures) or sequentially based on their degree (to simulate targeted attacks). In each iteration, we measure the size of the largest connected component (LCC) of the network. As nodes are successively removed, the overall network disintegrates into numerous subnetworks. The number of nodes in the LCC of

Table 4: p values obtained for the KS test for observed centrality distributions and the average centrality distributions obtained for 1000 randomised ensembles, for various SCNs.

Reference	Industry group	Subindustry classification	Observed BWC against randomised ensemble averages (p value for the KS test)	Observed CC against randomised ensemble averages (p value for the KS test)
1	Aircraft engines and engine parts	N/A	>0.1	≤0.05
2		N/A	≤0.05	>0.1
3	Arrangement of transportation of freight and cargo	N/A	≤0.05	>0.1
4		N/A	≤0.05	>0.1
5	Chemical	Soap and other detergents, except specialty cleaners	≤0.05	>0.1
6		Perfumes, cosmetics, and other toilet preparations	>0.1	>0.1
7		Pharmaceutical preparations	≤0.05	≤0.05
8		Paints, varnishes, lacquers, enamels, and allied products	>0.1	≤0.05
9		Primary batteries, dry and wet	>0.1	>0.1
10		Perfumes, cosmetics, and other toilet preparations	>0.1	>0.1
11		Perfumes, cosmetics, and other toilet preparations	≤0.05	≤0.05
12		Industrial organic chemicals	≤0.05	>0.1
13	Computer-related	Semiconductors and related devices	>0.1	>0.1
14		Computer peripheral equipment	≤0.05	>0.1
15		Computer peripheral equipment	>0.1	≤0.05
16		Computer peripheral equipment	>0.1	≤0.05
17		Computer peripheral equipment	>0.1	>0.1
18		Computer storage devices	>0.1	≤0.05
19	Electrical	Electromedical and electrotherapeutic apparatus	0.1–0.05	>0.1
20		Power-driven handtools	>0.1	≤0.05
21		Electromedical and electrotherapeutic apparatus	>0.1	>0.1
22		Telephone and telegraph apparatus	>0.1	>0.1
23		Electromedical and electrotherapeutic apparatus	≤0.05	≤0.05
24	Farm machinery and equipment	N/A	≤0.05	>0.1
25		N/A	>0.1	>0.1
26		N/A	≤0.05	≤0.05

the fragmented network therefore provides insights into its structural integrity in terms of overall connectivity.

Figure 7 illustrates the topological robustness assessment results. The size of the LCC has been plotted against the percentage of removed nodes, under random and targeted node removal scenarios. Note that for the random node removal scenario, the result has been obtained by averaging 100 runs of the simulation.

4.3.2. Directed SCNs. Unlike the UCR-SCN, the Willems [14] dataset for DMF-SCNs includes average daily demand levels at retailer nodes. Therefore, in order to more accurately assess the robustness of the DMF-SCNs, we develop a simple and intuitive methodology. In particular, we establish the robustness of each network as a function of lost average daily demand, as firms are sequentially removed. In developing the robustness assessment method, we assume that the firms within the same tier have substitutable capabilities with no capacity restrictions, i.e., the supplies lost due to failure of one firm can be fully replenished by another firm in the same tier.

For instance, consider the hypothetical DMF-SCN scenario presented in the Figure 8. In order for retailer B to satisfy its average daily demand, it relies on three supply chains, namely, (1) B-F-J-N, (2) B-F-J-M, and (3) B-F-I-M. Now consider a scenario where the manufacturing firm J

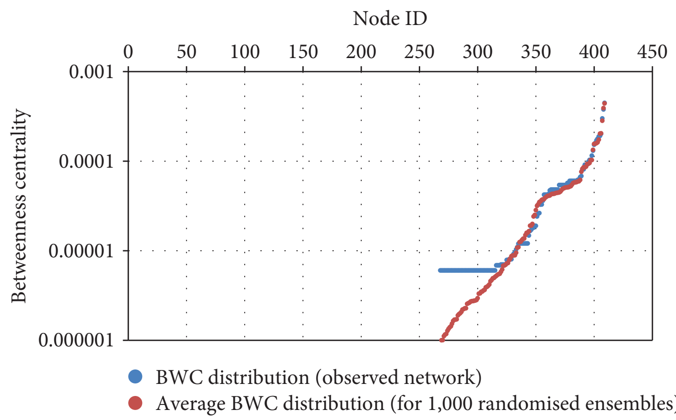


FIGURE 3: Betweenness centrality distribution of the observed SCN (ref #: 24—Farm Machinery and Equipment) against the average betweenness centrality distribution obtained from 1000 randomised ensembles.

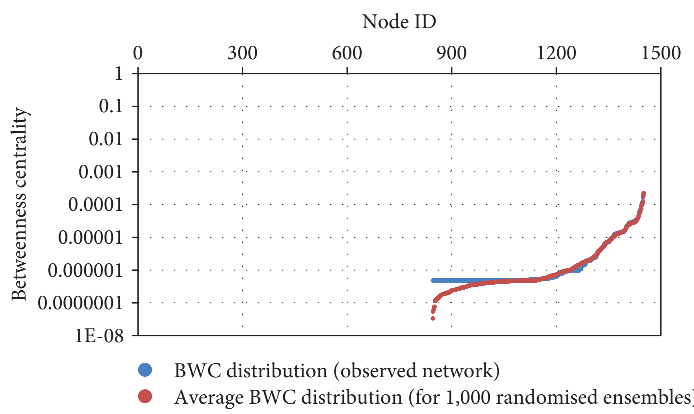


FIGURE 4: Betweenness centrality distribution of the observed SCN (ref #: 26—Farm Machinery and Equipment) against the average betweenness centrality distribution obtained from 1000 randomised ensembles.

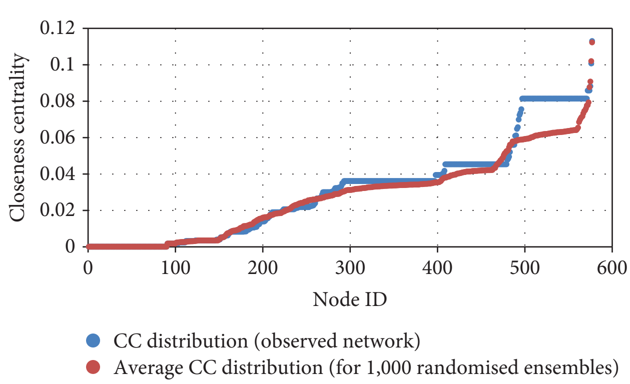


FIGURE 5: Closeness centrality distribution of the observed SCN (ref #: 18—Computer Storage Devices) against the average closeness centrality distribution obtained from 1000 randomised ensembles.

is removed from the network—in this case, retailer B will lose two out of the three supply chains it relies on to satisfy its average daily demand. However, retailer B will still be able to satisfy its average demand through the supply chain B-F-I-M, which is not affected by removal of firm J. Therefore, as long as the retailer node has access to an upstream most supplier through a complete supply chain, it will be able to satisfy its average daily demand. However, when firm J is removed, retailer D will not be able to satisfy its average daily demand—since both supply chains for D (D-G-J-M and D-G-J-N) are reliant on firm J. In this regard, it is important to note that removal of any retailer node implies that it will not be able to satisfy its respective average daily demand.

Based on the above idea, we develop a robustness metric termed "Robustness Score" which is defined as follows:

$$\text{Robustness Score} = \sum_{r=1}^{R} \delta_r D_r, \tag{3}$$

where R is the set of retailers in the SCN and D_r is the average daily demand at retailer r. δ_r captures the availability of paths to the upstream most suppliers at each retailer and is 1 if there exists at least one path connecting the retailer to an upstream most supplier and 0 otherwise.

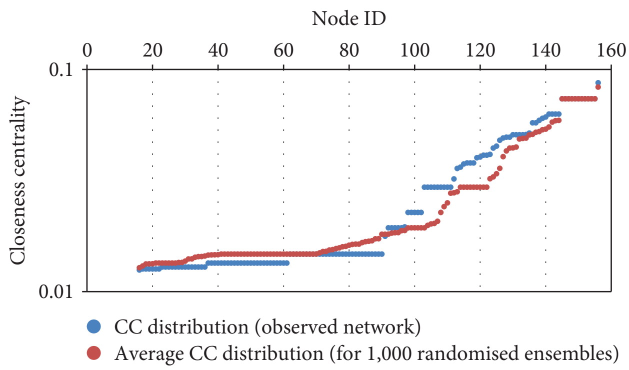


Figure 6: Closeness centrality distribution of the observed SCN (ref #: 16—Electromedical and Electrotherapeutic Apparatus) against the average closeness centrality distribution obtained from 1000 randomised ensembles.

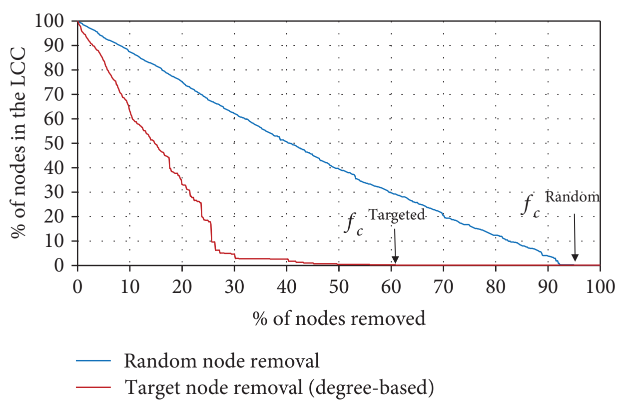


Figure 7: Percentage of nodes in the largest connected component under random and targeted (degree-based) node removal.

We applied the above idea of Robustness Score to the SCN ref 20 (power-driven handtools), where nodes were removed randomly and sequentially based on their degree (i.e., targeted attacks). For each scenario, after the removal of each node, the drop in Robustness Score was recorded to generate a profile. This result is presented in Figure 9.

5. Discussion

5.1. Topological Structure of the SCNs. The degree distributions for the majority of the DMF-SCNs indicate a good fit with power law. In particular, 14 out of the 26 networks analysed display 80% or higher R^2 correlation with a power-law fit for both their in- and out-degree distributions The degree exponents of all DMF-SCNs were generally found to lie in the range of 1 to 3: note that $\gamma = 2$ is the boundary between hub and spoke ($\gamma < 2$) and scale-free ($2 < \gamma < 3$) network topologies. Similar findings were observed for the UCR-SCN, which displayed a 95% R^2 correlation with a power-law fit for its degree distribution with a degree exponent of 2.31.

These results are in good agreement with the empirical findings on SCNs reported in recent data-driven studies [26, 27, 33, 34], which indicate the topologies of SCNs tend to have degree distributions which can be satisfactorily modelled by power law, with degree exponent ≈ 2 [22].

Although a number of past theoretical studies have relied on the Barabási-Albert (BA) model to generate topologies representative of SCNs [3, 10, 12], based on our results, it is evident that the BA model cannot sufficiently explain the intricacies of real-world SCN topologies. In particular, the BA model generates network topologies with $\gamma = 3$ [35], while the real-world SCNs indicate γ in the range of 1 to 3. Also, the assortative (or disassortative) mixing observed in real SCNs is not a feature of networks generated by the BA model, as shown analytically (in the limit of large network size) by Newman [36]. Finally, the BA model cannot generate networks with pronounced community structure which has been observed in real SCNs, since all nodes in the network belong to a single weakly connected component [37].

Indeed, a range of network growth models are available in extant literature, and broadly speaking, they can be categorised as either evolving models or generative models [17]. The evolving models are aimed at capturing the microscopic mechanisms underlying the temporal evolution of a network topology. In this regard, the BA model can be regarded as an evolving model. In contrast, generative models can be used to generate a snapshot of a network topology. Since the SCNs may have evolved based on various nongeneralizable principles, rather than attempting to understand and model the underlying growth mechanism through an evolving network growth model, it would be beneficial to simply mimic the observed topologies from data-driven studies using a generative network model.

In this regard, fitness-based generative models have recently gained prominence in theoretical research ([38–40]; Smolyarenko, 2014; [41]). In fitness-based models, the fitness distribution and the connection rules are given by a priori arbitrary functions, which enable a considerable amount of tuning (Smolyarenko, 2014). Indeed, this tunability makes such models a useful and practical modelling tool. For example, Ghadge et al. (2010) have proposed a purely statistical method for generating a range of network topologies by randomly allocating the nodes with fitness values sampled from a log-normal distribution. The propensity of each node to attract links is determined proportionally to its fitness. This method is referred to as the log-normal fitness attachment (LNFA), and it includes a tunable parameter σ (the shape parameter of the log-normal distribution), which can be manipulated to generate a large spectrum of networks. At one extreme, when σ is zero, all nodes have the same fitness and therefore at the time a new node joins the network, it chooses any existing node as a neighbour with equal probability, thus replicating the random graph model. On the other hand, when σ is increased beyond a certain threshold, very few nodes will have very large levels of fitness while the overwhelming majority of nodes have extremely low levels of fitness. As a result, the majority of new connections will be made to a few nodes which have high levels of fitness. The resulting network therefore resembles a monopolistic/"winner-take-all" scenario, which can sometimes be observed in the real world. Between the above two extremes (random

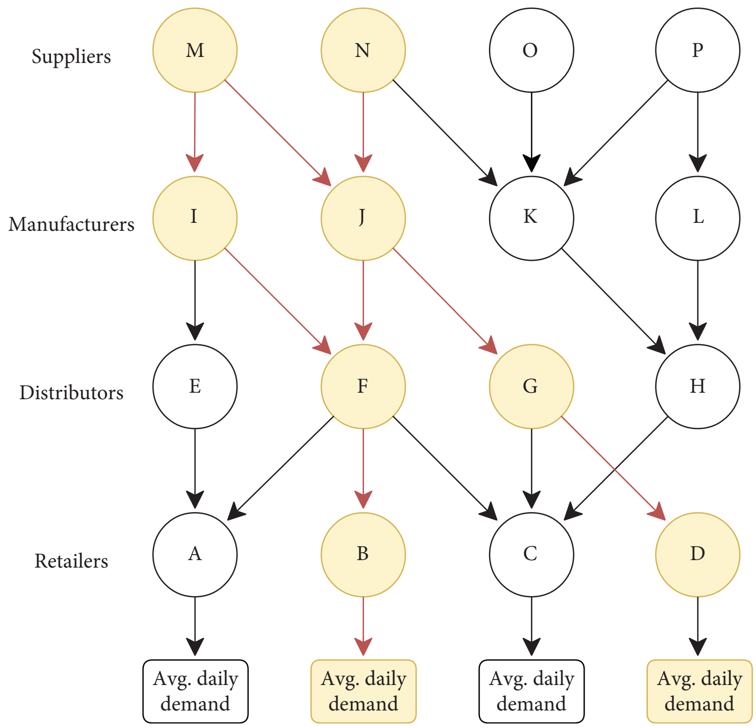


Figure 8: Path structure of a typical DMF-SCN.

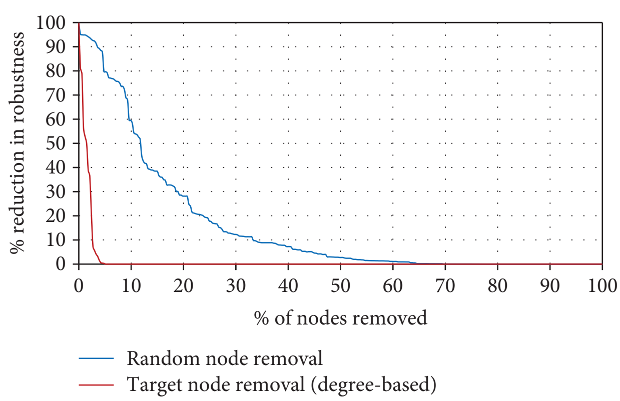


Figure 9: Effect of random and targeted (degree based) node removals on overall SCN robustness.

and monopolistic) lies a spectrum of power law networks which can closely represent many real SCNs (Ghadge et al., 2010). Nguyen and Tran [42] have illustrated that the LNFA model can indeed generate network topologies with γ in the range of 2–3 (and beyond), which can realistically represent the observed SCN topologies. Recent work by Bell et al. [43] has shown that a risk-averse behaviour by SCN firms can lead to a fitness-based network growth.

It is also interesting to note that all the DMF-SCNs indicate relatively lower network centralisation values llustrating the largely distributed and decentralised nature of modern SCNs. This lower centralisation could also be due to the recent supply chain practice known as modular assembly, where manufacturers obtain pre-assembled modules from a reduced base of suppliers (such as through intermediate subassemblers), in contrast to the traditional approach in which individual components are procured and assembled by the manufacturer [44]. In general, the SCNs which involve more complex and specialised manufacturing and assembly processes (such as aircraft parts, computer equipment and, electrical and farm machinery) were found to be more centralised than the other SCNs such as cargo and chemicals (which is also indicated by its low level of heterogeneity).

The network centralisation for the UCR-SCN was found to be higher than the average network centralisation for the DMF-SCN dataset. This could be due to the inherent differences between the two types of the SCNs, i.e., the contractual relationships are generally centralised through a leader firm whereas the flow networks generally link firms who wish to exchange material/goods.

In terms of degree assortativity, majority of the DMF-SCNs (21 out of 26) were found to be slightly or strongly disassortative, where the highly connected hubs tend to avoid each other, instead linking to lower-degree nodes. As a result, the network structure of these SCNs tends to display a hub-and-spoke character (as opposed to core periphery structure observed in assortative networks). Some SCNs in industries such as "Perfumes, Cosmetics, and Other Toilet Preparations" and "Semiconductors and Related Devices" were strongly disassortative. We found no SCN which was strongly assortative. Additionally, the UCR-SCN was also found to be disassortative, in terms of node degree. This type of disassortative mixing has been observed commonly in economic systems [17] where trade typically takes place between individuals or organizations of different skills and specialities.

An unfavourable implication of the disassortativity observed in the SCNs (in terms of degree) is that since high degree nodes are less connected to one another, many paths between nodes in the network are dependent on high degree nodes. Therefore, failure of a high degree node in a disassortative network would have a relatively large impact on the overall connectedness of the network [29]. On the other hand, disassortative networks are generally resilient against cascading impacts arising from targeted attacks—since hub nodes are not connected with each other, the likelihood of disruption impacts cascading from one hub node to another is minimised [45].

Furthermore, the DMF-SCNs mostly displayed slight or strong disassortative tendencies in terms of stage-cost and stage-time attributes (20 out of 26 SCNs were found to be disassortative in terms of stage cost and stage time). That is, firms which contribute high stage cost are on average more likely to be connected with firms that contribute low stage cost, and vice versa, and the same is true for stage time. No SCN that we studied displayed a strong assortative tendency in terms of these attributes. The strictly tiered structure of these SCNs could be responsible for this stage-cost and stage-time disassortativity as it separates the functional capabilities of firms which are linked with each other between tiers.

An important difference between the UCR-SCN and the DMF-SCNs was observed in their clustering properties. The clustering coefficient indicates the degree to which firms in a SCN tend to cluster together around a given firm. For example, it can indicate how various suppliers behave with respect to the final assembler at the global level [23]. Therefore, the higher the clustering coefficient, the more dependent suppliers are on each other for production [46]. The UCR-SCN indicated low levels of clustering while none of the DMF-SCNs indicated any clustering. It is noted that in the DMF-SCN dataset considered, no horizontal connections were observed (i.e., no connections between firms within the same functional tier). Due to this inherent structural limitation, which prohibits triadic closure, these SCNs do not indicate any clustering (therefore, clustering values are not reported in Table 2). This implies that the suppliers in the material flow networks are independent of each other (since they are likely to be competitors). While this structure may limit diffusion of knowledge through the SCN, it has favourable implications in terms of network robustness, since disruptions to one supplier is unlikely to impact another. Hearnshaw and Wilson (2010) note that SCNs with a low clustering coefficient are likely to experience a more opportunistic behaviour and less collaboration which may lead to system inefficiencies due to the difficulty in system-wide coordination. Therefore, SCN efficiency can be improved by creating new horizontal connections, such as through social relationships. Choi and Wu [47] report a real-world example of such a scenario where Honda has built clustering in their supply chain by directing and facilitating the relationships of its first-tier suppliers with some second-tier suppliers.

5.2. Correlation between Topological Centrality Metrics and Other Firm Attributes in Directed Material Flow SCNs. For the DMF-SCNs, node centrality correlation assessments were carried out against node attributes. In general, for the majority of SCNs (regardless of the industry), we observed the following:

(1) The degree of firms was found to correlate positively with the stage cost while no correlation was identified with the stage time. This implies that the firms with higher number of connections generally add higher levels of direct costs to the market price of the final product.

   (i) The in-degree of firms was found to correlate positively with the stage cost and negatively with stage time. This was particularly evident for the SCNs in aircraft engine and electrical industries. It implies that in these SCNs, the firms with higher number of upstream suppliers tend to add higher levels of direct costs to the market price of the final product and they tend to have relatively lower average processing times (these are firms which play the role of assemblers, which bring together many components from various suppliers for assembly purposes).

   (ii) With the exception of SCNs in aircraft engine and electrical industries, the out-degree of firms was found to correlate positively with the stage cost and the stage time, suggesting that the firms with high number of downstream customers (such as major distributors) tend to add higher levels of direct costs to the market price of the final product and they tend to have relatively higher average processing times. This was particularly evident for the two SCNs in the transportation of freight and cargo industry.

(2) The betweenness and closeness centralities of firms were found to correlate positively with their stage cost, which implies that those firms which are more involved in the relationships between other pairs of firms and those firms which are active information generators tend to add higher values to the final product.

(3) The SCNs in electrical, computer-related, farm machinery and equipment, and aircraft engine industries indicated negative correlations between the betweenness centrality and stage time. This implies that, in these SCNs, the firms which are involved in the relationships between two other firms tend to require lower average processing times. Such central firms generally play the role of assemblers or distributors, thus requiring less processing times.

The above insights demonstrate how, from a SCN point of view, the position of an individual firm with respect to the others can influence both strategy and behaviour [48].

5.3. Self-Organized Features

5.3.1. Undirected SCN. Comparison of the betweenness and closeness centrality distributions of the original network with the average betweenness and closeness distributions obtained for the 1000 randomised networks (shown in Figure 2) reveals that there exists a significant difference (KS test p value is <0.05) between the respective centrality distributions in the observed SCN and the average for the naturally occurring scheme.

In terms of betweenness centrality, we observe that moderate betweenness firms were overrepresented in the observed SCN compared to the average of the randomised ensembles. Therefore, we can conclude that the UCR-SCN topologies are nonrandom and favour moderate betweenness firms. A similar result was reported by Becker et al. [32] who constructed a manufacturing system network model from real-world data (where nodes represent separate work stations and links represent material flows between work stations). By applying the DPR process to generate an ensemble of networks with the same degree distribution, they observed that work stations with a particularly high betweenness centrality are overrepresented in the manufacturing system studied. They concluded that the manufacturing system topology is therefore nonrandom and favours the existence of a few highly connected work stations. Betweenness of a firm in the context of a UCR-SCN indicates the extent to which it can intervene over interactions among other firms in the SCN by being a gatekeeper for relationships [23]. Indeed, this has specific advantages in a networked economy, since it enables the firms to acquire more market intelligence and control by playing the role of an intermediator. Therefore, it is reasonable to see firms self-organizing themselves in order to increase their betweenness in the SCN.

In relation to closeness centrality, we observed that all firms in the UCR-SCN included lower closeness centralities when compared with the average of the randomised ensembles. Closeness indicates the proximity of a given firm with respect to others in the network. Complex manufacturing industries, such as the automobile manufacturing sector indicated in the UCR-SCN, include longer supply and service chains which can place the firms peripherally and far away from each other (as indicated by the network diameter of 7 for the UCR-SCN, which is higher than the majority of the DMF-SCNs).

5.3.2. Directed SCNs. Our results indicate that only some DMF-SCNs include self-organized features in terms of betweenness and closeness. Those SCNs with a p value ≤ 0.05 for the KS test, as outlined in Table 4, indicate that there exists the significant difference between the centrality distributions of the observed SCN and the average of the randomised network ensembles. This implies that in these SCNs, there exists an external mechanism beyond the degree distribution, which has driven the firms to adopt various centrality levels. In general, as illustrated in Figures 3 and 4, comparison of observed betweenness profiles against the average of randomised ensembles reveals that firms with low betweenness are underrepresented in the observed DMF-SCNs. This is an interesting observation which complements the findings obtained for the UCR-SCNs which indicated the moderate betweenness firms to be overrepresented. This implies that in material flow SCNs, the firms need to maintain a certain level of betweenness to function (as evident from the plateau of low betweenness nodes in the profiles of observed SCNs presented in Figures 3 and 4).

In contrast, comparison of observed closeness profiles against the average of randomised ensembles reveals that firms with moderate closeness are overrepresented in the observed DMF-SCNs (Figures 5 and 6). While it is closely related to betweenness centrality, closeness is more relevant in situations where a firm acts as a generator of information rather than a mere mediator/gatekeeper. For example, due to various hindrances, the market demand information can easily be distorted when it flows from the downstream firms towards upstream firms. Such distortions can lead to undue deviation between production plans of manufacturers and supply plans of suppliers, leading to a phenomenon known as the bullwhip effect. Firms with high closeness centrality levels therefore play a major role in sharing the actual market demand information with upstream firms in the SCN, thus diminishing the adverse impacts arising from the bullwhip effect [9]. This could be a reason why the firms in the DMF-SCNs have self-organized to have higher than random levels of closeness in the moderate closeness regime.

5.4. Robustness Character

5.4.1. Undirected SCN. The point identified as f_c in Figure 7 indicates the critical threshold at which the LCC disappears and the network is fully fragmented into individual nodes. As can be seen from this figure, the network breaks apart relatively rapidly when the nodes are removed sequentially by targeting higher degree ones first, compared with random node removals. It is evident that f_c^{Targeted} occurs at 60.93% while f_c^{Random} occurs at 95.04%.

It is interesting to investigate how different the observed robustness (against random node removals) of the network at hand is, when compared against a randomly wired network of the same size (in terms of the number of nodes and links). This question can be answered using the Molloy-Reed criterion which identifies the presence of a LCC within a network (Reed, 1995). This is on the basis that for a network to have a LCC, most nodes that belong to it must be connected to at least two other nodes [17]. In particular, the Molloy-Reed criterion states that a network has a LCC if

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} > 2, \tag{4}$$

where $\langle k \rangle$ and $\langle k^2 \rangle$ are the first (mean) and second moment of the network's degree distribution, respectively.

Based on the above, networks with $\kappa < 2$ lack a LCC and are composed of many disconnected clusters. Applying the insight provided by the Molloy-Reed criterion to a network with an arbitrary degree distribution, one can predict the fraction of nodes required to be randomly removed from the network in order to destroy its LCC (i.e., the critical threshold, f_c) as follows [49];

$$f_c = 1 - \frac{1}{\langle k^2 \rangle / \langle k \rangle - 1} \tag{5}$$

In contrast, the critical threshold of an Erdős–Rényi network (i.e., a randomly wired network) is given by [17]

$$f_c^{\mathrm{ER}} = 1 - \frac{1}{\langle k \rangle}. \tag{6}$$

A network is considered to display enhanced robustness if its critical threshold is higher than that of a randomly wired network of the same size (in terms of the number of nodes and links). The UCR-SCN is characterised by $\langle k \rangle = 9.56$ and $\langle k^2 \rangle = 392.16$. Using the above formulae, we can determine f_c to be 0.975 and f_c^{ER} to be 0.895. This implies that, in order to destroy the LCC of the UCR-SCN by fragmenting it into many disconnected components, one would need to remove 97.5% of the nodes (note that this theoretical prediction is generally in agreement with f_c of 95.04% established through the average of 100 simulations). Also, since $f_c > f_c^{\mathrm{ER}}$, we can conclude that the UCR-SCN displays enhanced robustness against random failure of firms. Indeed, the enhanced robustness of this SCN against random firm removals manifests owing to its hub structure. Random node removals, by definition, affect nodes irrespective of their degrees. Since scale-free networks, such as the UCR-SCN considered here, comprise predominantly of less connected nodes and a few hubs, the chance of randomly removing a hub is almost negligible. Therefore, random node removals are likely to affect mainly the less connected nodes, which, although numerous, play a limited role in maintaining a network's integrity [17].

5.4.2. Directed SCN. For the DMF-SCNs, we demonstrate a simple and intuitive method which considers the impact of firm failure on the output capability of the SCN. This method can be used by practitioners to establish the robustness of their SCNs and also to compare the robustness of various SCN systems against each other.

Since the above robustness concept for the DMF-SCNs depends on the demand profiles at the retailers, it may not provide generalizable insights. Therefore, we have not attempted to investigate the robustness character of the full dataset. However, it is evident from the robustness profile presented for the reference SCN 20 (power-driven hand tools) that the targeted removal of nodes based on their degree has a drastic impact on the overall network robustness—only 5% of the firms need to be removed for the entire SCN to be incapable to meet any demand at the retailers. In comparison, the SCN is generally much more robust against random removal of firms—about 70% of the firms need to be removed randomly, before the SCN is incapable of satisfying any demand at the retailers. The above result highlights the importance of hub nodes, through which the majority of the paths traverse. Therefore, the removal of these nodes will have significant impacts on the ability of SCN to meet the demands.

Sheffi and Rice [50] note the importance of building flexibility and redundancies into SCNs as a way of improving the robustness of these systems. In this regard, parallel supply paths with minimal dependencies could be incorporated into SCNs, so that a disruption in one firm does not impact the operations of the other.

6. Conclusions and Future Directions

In general, in both DMF and UCR SCNs, we observed degree distributions which conform to power law indicating the existence of hub/leader firms. In both types of SCNs, disassortative mixing was observed in terms of the degree of firms. Interestingly, majority of DMF SCNs also showed disassortative mixing in terms of firm cost and time attributes. Since DMF SCNs are characterised by various functional tiers, no clustering was evident in these systems, while some level of clustering was identified in the UCR SCN. The node centrality correlation assessment carried out against firm attributes (cost and time) for DMF-SCNs reveals a relationship between the position and the function of firms within the system.

Additionally, we identified that the UCR SCN included enhanced topological robustness against random firm failures and self-organized topological features in terms of both betweenness and closeness of firms. However, self-organized features were only present in some of the DMF SCNs.

Finally, we have developed a simple and intuitive method for assessing the robustness of DMF SCNs considering the demand loss at retailers as firms are removed. This analysis outlines the importance of hub firms, through which the majority of the supply paths traverse. Therefore, the removal of these nodes will have significant impacts on the ability of SCN to meet the demands. The robustness score concept can indeed be used to identify the most critical firms in a DMF SCN, which will provide insights beyond the purely topology-based centrality metrics. Future research could investigate the application of TC-DPR we introduced here to identify the configuration, of a particular tiered SCN, which will maximise the robustness score by allocating supply paths based on demand levels at retailers.

Our work for the first time attempted to generalize the topological features of a large number of SCNs from the manufacturing sector. It is notable that since we only considered relatively large networks, finite size effects are minimal. While some topological features were indeed network-specific, the topological similarities between the networks were striking. Therefore, this work could be used as a benchmark for developing generalized growth mechanisms for SCNs in future.

Appendix

A. Network and Node-Level Metrics

Table 5: Network-level metrics used and their SCN implications.

Mathematical representation	SCN implication
Average degree	
For an undirected network: $<k> = 1/N\sum_{i=1}^{N} k_i = 2L/N$ For a directed network: $<k^{\text{in}}> = 1/N\sum_{i=1}^{N} k_i^{\text{in}} = <k^{\text{out}}> = 1/N\sum_{i=1}^{N} k_i^{\text{out}} = L/N$, where L and N are the total number of links and nodes in the network.	Indicates, on average, how many connections a given firm has. A higher average degree implies good interconnectivity among the firms in the SCN.
Network diameter	
$\text{diameter} = \max_{i,j} l(i, j)$, where l is the number of hops traversed along the shortest path from node i to j.	The diameter of a SCN is the largest distance between any two firms in the network (i.e., the maximum shortest path length). More complex manufacturing processes can include large network diameters (i.e., many stages of production) indicating difficulty in governing the overall SCN under a centralised authority.
Network density (D)	
$D = <k>/N - 1$, where $<k>$ is the mean degree of all the nodes and N is the total number of nodes, in the network.	Density of a SCN indicated the level of interconnectivity between the firms involved. SCNs with high density indicate good levels of connectivity between firms which can be favourable in terms of efficient information exchange and improved robustness due to redundancy and flexibility [50].
Network centralisation (C)	
$C = (N/N - 2)(\max (k)/N - 1 - \text{density})$, where N is the total number of nodes in the network and $\max (k)$ is the maximum degree of a node within the network. Density is determined as per the equation below.	Network centralisation provides a value for a given SCN between 0 (if all firms in the SCN have the same connectivity) and 1 (if the SCN has a star topology). This indicates how the operational authority is concentrated in a few central firms within the SCN. Highly centralised SCNs can have convenience in terms of centralised decision implementation and high level of controllability in production planning. However, highly centralised SCNs lack local responsiveness since relationships between firms in various tiers are decoupled [23].
Network heterogeneity (H)	
$H = \sqrt{\text{variance}(k)}/<k>$, where $<k>$ is the mean degree and $\text{variance}(k)$ is the variance of the degree, of all the nodes in the network.	Heterogeneity is the coefficient of variation of the connectivity. Highly heterogeneous SCNs exhibit hub firms (i.e., firms with high number of contractual connections). In extreme cases, there may be many super large hubs (winner-take-all scenario, indicating centralised control of the overall SCN through a very few firms).
Average clustering coefficient ($<C>$)	
$<C> = \sum_i C_i/N$, where N is the total number of nodes in the network and C_i is the number of triangles connected to node i divided by the number of triples centered around node i.	The clustering coefficient indicates the degree to which firms in a SCN tend to cluster together around a given firm. For example, it can indicate how various suppliers behave with respect to the final assembler at the global level [23]. Therefore, the higher the clustering coefficient, the more dependent suppliers are on each other for production [46].
Characteristic (or average) path length ($<l>$)	
The characteristic path length $<l>$ is $<l> = 1/N(N-1)\sum_{i\neq j} l_{i,j}$, where N is the total number of nodes in the network and $l_{i,j}$ is the shortest topological distance between nodes i and j.	Characteristic (or average) path length is the average topological distance between all pairs of firms (along the shortest path) in a SCN. It measures how efficiently information can be transferred between pairs of firms within a SCN.
Degree exponent (γ) [35]	
The degree distribution P_k of an undirected scale-free network is approximated with power law as follows: $P_k \sim k^{-\gamma}$, where k is the degree of the node and γ is the degree	SCNs with $\gamma < 2$ include very large hubs which acquire control through contractual relationships with other firms at a rate faster than the growth of the SCN in terms of new firm additions. As γ continues to increase beyond 2, the SCNs include smaller and less numerous hubs, which

Table 5: Continued.

Mathematical representation	SCN implication
exponent (also known as the power law or scale-free exponent). Directed networks generally include two separate degree distributions, one for the in-degree and another for the out-degree. In such cases, there will be two separate degree exponents, i.e., γ_{in} and γ_{out}.	ultimately leads to a topology similar to that of a random network where all firms have almost the same number of connections. In particular, when γ is less than or equal to 2, the network topology is referred to as a "hub and spoke" topology; when γ is higher than 2 but less than 3, the network topology is referred to as scale-free; and when γ is higher than 3, the network topology is random.
Assortativity (ρ) [36]	
Assortativity is formally defined as a correlation function of excess degree distributions and link distribution of a network. For undirected networks, when degree distribution is denoted as p_k and excess degree (remaining degree) distribution is denoted as q_k, one can introduce the quantity $e_{j,k}$ as the joint probability distribution of the remaining degree distribution of the remaining degrees of the two nodes at either end of a randomly chosen link. Given these distributions, the assortativity of an undirected network is defined as $\rho = 1/\sigma_q^2[\sum_{jk} jk(e_{jk} - q_j q_k)]$, where σ_q is the standard deviation of q_k which is given as $\sigma_q^2 = \sum_k k^2 q_k - [\sum_k k q_k]^2$. Assortativity, ρ is a value between −1 and 1. For $\rho > 0$, the network is assortative; for $\rho = 0$, the network is neutral; and for $\rho < 0$, the network is disassortative.	Positive assortativity means that the firms with similar connectivity would have a higher tendency to connect with each other (for example, highly connected firms could be managing subcommunities in certain areas of production and then connect to other high-degree firms undertaking the same function). This structure can lead to cascading disruptions—where a disruption at one leaf node can spread quickly within the network through the connected hubs [46]. In contrast, a negative assortativity indicates that it is the firms with dissimilar connectivity that tend to pair up in the given network. Note that assortativity can also be defined in terms of node attributes other than the degree.

Table 6: Node-level metrics used and their SCN implications.

Mathematical representation	SCN implication
Degree (k)	
In undirected networks, the degree of node i is given as $k_i = \sum_{j=1}^{N} A_{ji} = \sum_{i=1}^{N} A_{ji}$. In directed networks, the degree of node i is separated into in- and out-degrees, as follows: $k_i^{\text{in}} = \sum_{j=1}^{N} A_{ij}, \quad k_i^{\text{out}} = \sum_{j=1}^{N} A_{ji}$, where A_{ij} is any element of the adjacency matrix A.	Represents the number of direct neighbours (connections) a given firm has. For instance, in a given SCN, the firm with the highest degree (such as the integrators that assemble components) is deemed to have the largest impact on operational decisions and strategic behaviours of other firms in that particular SCN. Such a firm has the power to reconcile the differences between various other firms in the SCN and align their efforts with greater SCN goals [23]. In directed networks, the firms which have high in-degree are considered to be "integrators" who collect information from various other firms to create high-value products. In contrast, the firms which have high out-degree are considered to be "allocators" who are generally responsible for distribution of high-demand resources to other firms and/or customers.
Betweenness centrality (normalised) [51]	
The betweenness centrality of a node n is defined as $C_b(n) = 2/(N-1)(N-2)\sum_{s\neq n\neq t}(\sigma_{s,t}(n)/\sigma_{s,t})$, where s and t are nodes in the network, which are different from n, $\sigma_{s,t}$ denotes the number of shortest paths from s to t, and $\sigma_{s,t}(n)$ is the number of shortest paths from s to t that n lies on.	Betweenness centrality of a firm is the number of shortest path relationships going through it, considering the shortest path relationships that connect any two given firms in the SCN. Therefore, it indicates the extent to which a firm can intervene over interactions among other firms in the SCN by being a gatekeeper for relationships [23]. Those firms with high levels of betweenness generally play a vital role in SCNs—mainly owing to their ability to increase the overall efficiency of the SCN by smoothing various exchange processes between firms.
Closeness centrality [52]	
The closeness centrality of a node n is defined as $C_c(n) = 1/<L(n,m)>$, where $<L(n,m)>$ is the length of the shortest path between two nodes n and m (note that for unweighted graphs with no geodesic distance information, each link is assumed to be one unit of distance). The closeness centrality of each node is a number between 0 and 1.	Closeness centrality is a measure of the time that it takes to spread the information from a particular firm to the other firms in the network. While it is closely related to betweenness centrality, closeness is more relevant in situations where a firm acts as a generator of information rather than a mere mediator/gatekeeper. Firms with high closeness centrality levels enable the overall SCN to be more market sensitive (i.e., responsive) by spreading the actual market demand information with the other upstream firms [9].

B. Node-Level Centrality Correlations with Stage Time and Stage Cost

B.1 Correlation Analysis for Total Degree.

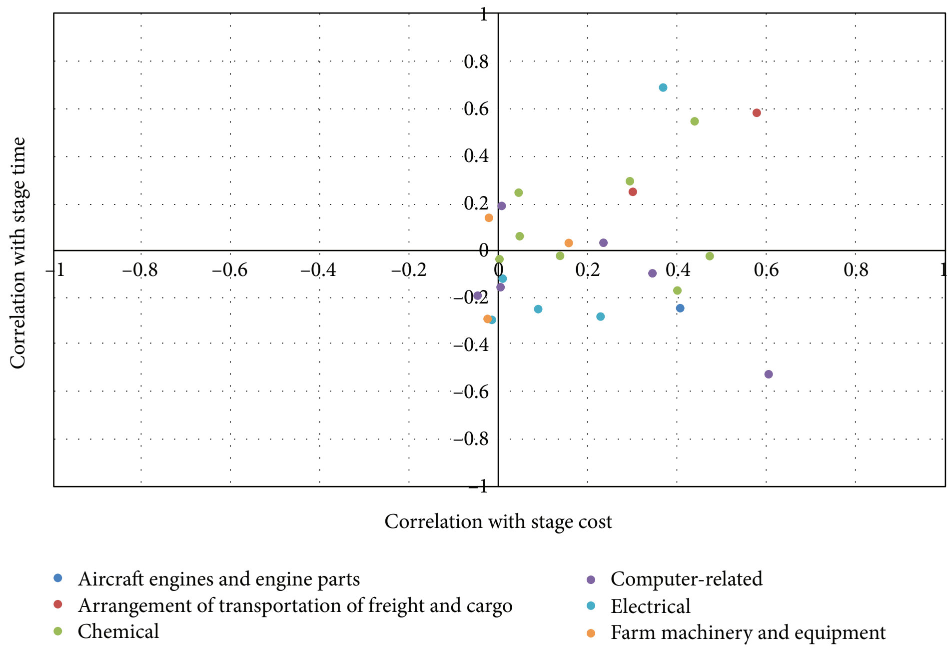


Figure 10

B.2 Correlation Analysis for In-Degree.

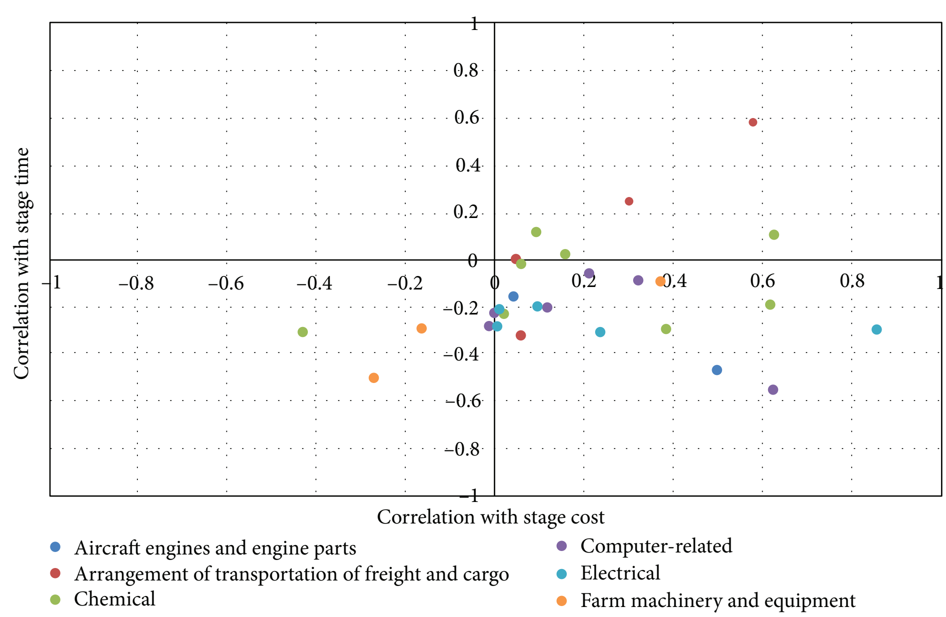


Figure 11

B.3 Correlation Analysis for Out-Degree.

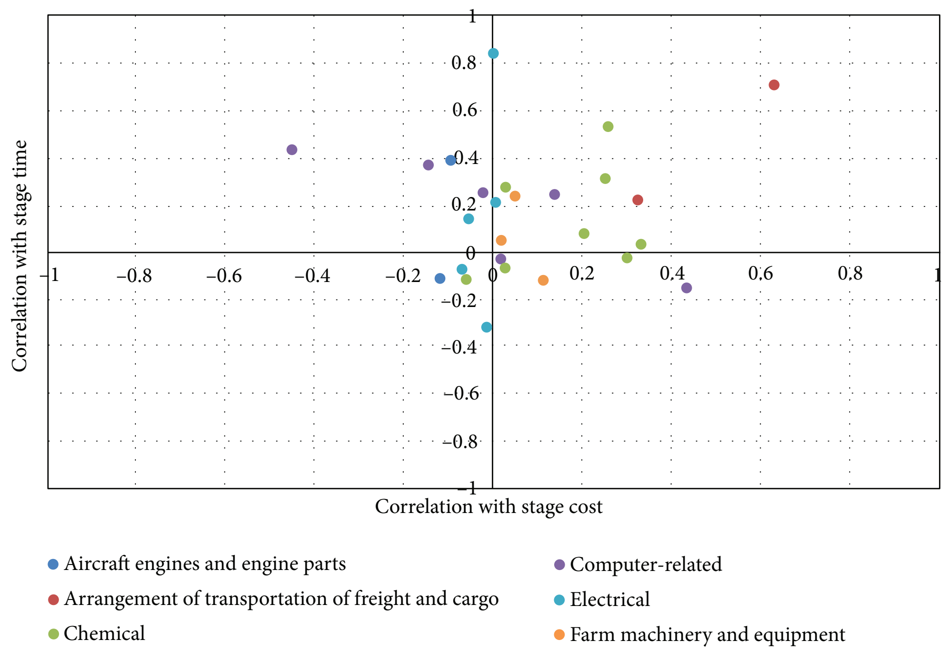


Figure 12

B.4 Correlation Analysis for Betweenness Centrality.

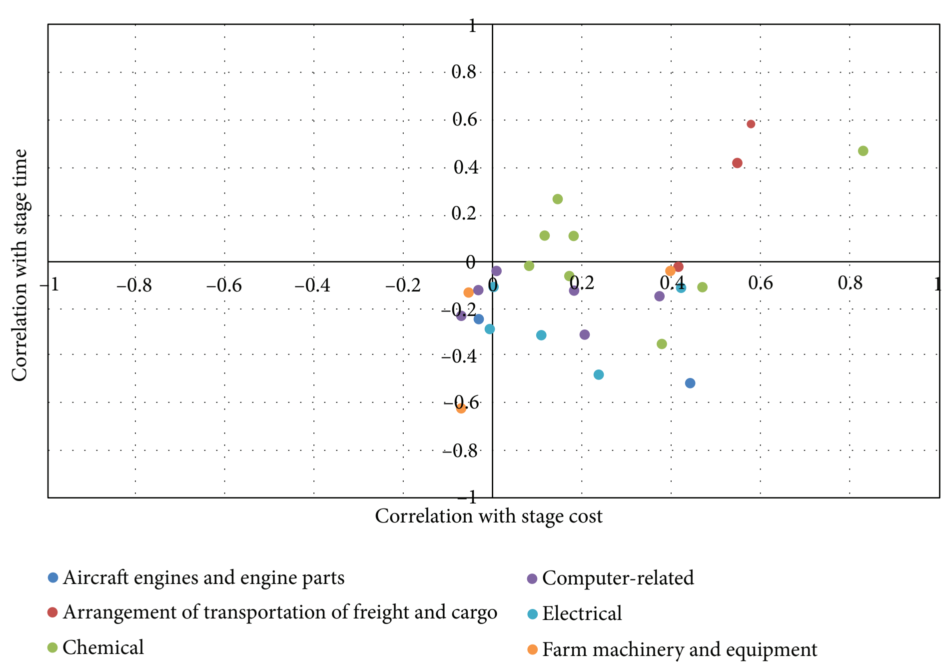


Figure 13

B.5 Correlation Analysis for Closeness Centrality.

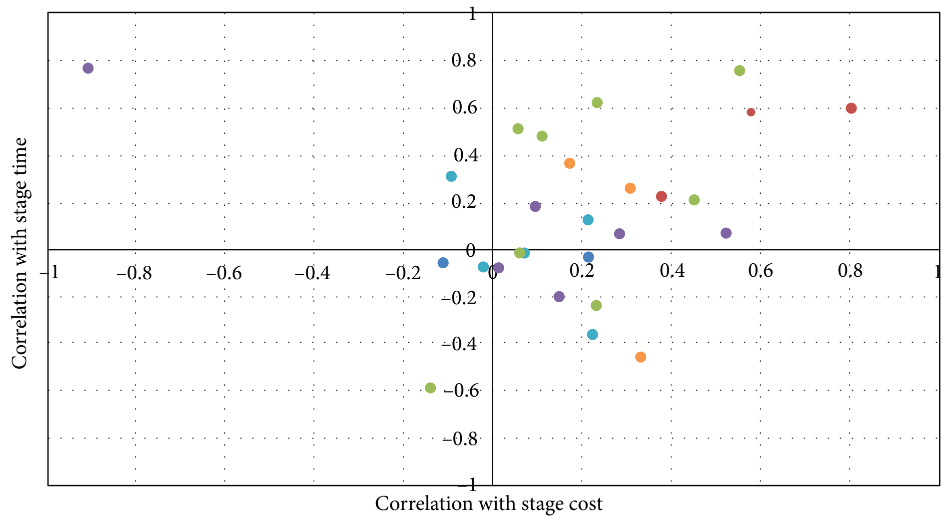


Figure 14

Data Availability

The DMF SCN dataset is publicly available under the published work of Willems [14]. The adjacency matrix for the UCR SCN can be provided upon request by the corresponding author.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

The authors acknowledge the Australian Research Council (ARC) for funding this work under grant DP140103643.

References

[1] T. Y. Choi, K. J. Dooley, and M. Rungtusanatham, "Supply networks and complex adaptive systems: control versus emergence," *Journal of Operations Management*, vol. 19, no. 3, pp. 351–366, 2001.

[2] S. D. Pathak, J. M. Day, A. Nair, W. J. Sawaya, and M. M. Kristal, "Complexity and adaptivity in supply networks: building supply network theory using a complex adaptive systems perspective*," *Decision Sciences*, vol. 38, no. 4, pp. 547–580, 2007.

[3] H. P. Thadakamalla, U. N. Raghavan, S. Kumara, and A. Albert, "Survivability of multiagent-based supply networks: a topological perspective," *IEEE Intelligent Systems*, vol. 19, no. 5, pp. 24–31, 2004.

[4] Y. Kim, Y. S. Chen, and K. Linderman, "Supply network disruption and resilience: a network structural perspective," *Journal of Operations Management*, vol. 33-34, pp. 43–59, 2015.

[5] Y. Li, "Networked analysis approach of supply chain network," *Journal of Networks*, vol. 9, no. 3, 2014.

[6] S. I. Mari, Y. H. Lee, M. S. Memon, Y. S. Park, and M. Kim, "Adaptivity of complex network topologies for designing resilient supply chain networks," *International Journal of Industrial Engineering*, vol. 22, no. 1, pp. 102–116, 2015.

[7] L. Wen, M. F. Guo, and L. J. wang, "Statistic characteristics analysis of directed supply chain complex network," *International Journal of Advancements in Computing Technology*, vol. 4, no. 21, pp. 84–91, 2012.

[8] M. Xu, X. Wang, and L. Zhao, "Predicted supply chain resilience based on structural evolution against random supply disruptions," *International Journal of Systems Science: Operations & Logistics*, vol. 1, no. 2, pp. 105–117, 2014.

[9] N. R. Xu, J. B. Liu, D. X. Li, and J. Wang, "Research on evolutionary mechanism of agile supply chain network via complex network theory," *Mathematical Problems in Engineering*, vol. 2016, Article ID 4346580, 9 pages, 2016.

[10] Q. Xuan, F. Du, Y. Li, and T. J. Wu, "A framework to model the topological structure of supply networks," *IEEE Transactions on Automation Science and Engineering*, vol. 8, no. 2, pp. 442–446, 2011.

[11] K. Zhao, A. Kumar, and J. Yen, "Achieving high robustness in supply distribution networks by rewiring," *IEEE Transactions on Engineering Management*, vol. 58, no. 2, pp. 347–362, 2011.

[12] K. Zhao, A. Kumar, T. P. Harrison, and J. Yen, "Analyzing the resilience of complex supply network topologies against random and targeted disruptions," *IEEE Systems Journal*, vol. 5, no. 1, pp. 28–39, 2011.

[13] M. Parhi, "Impact of the changing facets of inter-firm interactions on manufacturing excellence: a social network perspective of the Indian automotive industry," *Asian Journal of Technology Innovation*, vol. 16, no. 1, pp. 117–141, 2008.

[14] S. P. Willems, "Data set-real-world multiechelon supply chains used for inventory optimization," *Manufacturing & Service Operations Management*, vol. 10, no. 1, pp. 19–23, 2008.

[15] S. Lazzarini, F. Chaddad, and M. Cook, "Integrating supply chain and network analyses: the study of netchains," *Journal on Chain and Network Science*, vol. 1, no. 1, pp. 7–22, 2001.

[16] E. J. S. Hearnshaw and M. M. J. Wilson, "A complex network approach to supply chain network theory," *International Journal of Operations & Production Management*, vol. 33, no. 4, pp. 442–469, 2013.

[17] A. Barabási, *Network Science*, Cambridge University Press, 2016.

[18] S. M. Park and B. J. Kim, "Dynamic behaviors in directed networks," *Physical Review E*, vol. 74, no. 2, article 026114, 2006.

[19] N. Schwartz, R. Cohen, D. Ben-Avraham, A. L. Barabási, and S. Havlin, "Percolation in directed scale-free networks," *Physical Review E*, vol. 66, no. 1, article 015104, 2002.

[20] L. D. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas, "Characterization of complex networks: a survey of measurements," *Advances in Physics*, vol. 56, no. 1, pp. 167–242, 2007.

[21] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010.

[22] S. Perera, M. G. H. Bell, and M. C. J. Bliemer, "Network science approach to modelling the topology and robustness of supply chain networks: a review and perspective," *Applied Network Science*, vol. 2, no. 1, p. 33, 2017.

[23] Y. Kim, T. Y. Choi, T. Yan, and K. Dooley, "Structural investigation of supply networks: a social network analysis approach," *Journal of Operations Management*, vol. 29, no. 3, pp. 194–211, 2011.

[24] T. Y. Choi and Y. Hong, "Unveiling the structure of supply networks: case studies in Honda, Acura and DaimlerChrysler," *Journal of Operations Management*, vol. 20, no. 5, pp. 469–493, 2002.

[25] T. Kito, A. Brintrup, S. New, and F. Reed-Tsochas, *The Structure of the Toyota Supply Network: An Empirical Analysis*, SSRN Scholarly Paper ID 2412512, Social Science Research Network, Rochester, NY, 2014.

[26] A. Brintrup, Y. Wang, and A. Tiwari, "Supply networks as complex systems: a network-science-based characterization," *IEEE Systems Journal*, vol. 11, no. 4, pp. 2170–2181, 2017.

[27] P. Orenstein, "How does supply network evolution and its topological structure impact supply chain performance?," in *2016 Second International Symposium on Stochastic Models in Reliability Engineering, Life Science and Operations Management (SMRLO)*, pp. 562–569, Beersheba, Israel, February 2016.

[28] S. Maslov and K. Sneppen, "Specificity and stability in topology of protein networks," *Science*, vol. 296, no. 5569, pp. 910–913, 2002.

[29] R. Noldus and P. Van Mieghem, "Assortativity in complex networks," *Journal of Complex Networks*, vol. 3, no. 4, pp. 507–542, 2015.

[30] B. K. Fosdick, D. B. Larremore, J. Nishimura, and J. Ugander, "Configuring random graph models with fixed degree sequences," 2016, http://arxiv.org/abs/1608.00607.

[31] C. Diebolt, T. Mishra, and M. Parhi, "Dynamics of distribution and diffusion of new technology," in *India Studies in Business and Economics ReDIF-Book*, Springer, 2016.

[32] T. Becker, M. Meyer, and K. Windt, "A manufacturing systems network model for the evaluation of complex manufacturing systems," *International Journal of Productivity and Performance Management*, vol. 63, no. 3, pp. 324–340, 2014.

[33] Z. Gang, Y. Ying-Bao, B. Xu, and P. Qi-Yuan, "On the topological properties of urban complex supply chain network of agricultural products in mainland China," *Transportation Letters*, vol. 7, no. 4, pp. 188–195, 2015.

[34] W. Keqiang, Z. Zhaofeng, and S. Dongchuan, "Structure analysis of supply chain networks based on complex network theory," in *2008 Fourth International Conference on Semantics, Knowledge and Grid*, pp. 493-494, Beijing, China, December 2008.

[35] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[36] M. E. J. Newman, "Assortative mixing in networks," *Physical Review Letters*, vol. 89, no. 20, article 208701, 2002.

[37] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.

[38] C. Bedogne' and G. J. Rodgers, "Complex growing networks with intrinsic vertex fitness," *Physical Review E*, vol. 74, no. 4, article 046115, 2006.

[39] G. Bianconi and A. L. Barabási, "Competition and multiscaling in evolving networks," *Europhysics Letters (EPL)*, vol. 54, no. 4, pp. 436–442, 2001.

[40] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz, "Scale-free networks from varying vertex intrinsic fitness," *Physical Review Letters*, vol. 89, no. 25, article 258702, 2002.

[41] S. Perera, M. G. H. Bell, and M. C. J. Bliemer, "Resilience characteristics of supply network topologies generated by fitness based growth models," *Transportation Research Board 95th Annual Meeting*, 2016, Washington DC, United States, January 2016, 2016.

[42] K. Nguyen and D. A. Tran, "Fitness-based generative models for power-law networks," in *Handbook of Optimization in Complex Networks*, pp. 39–53, Springer, Boston, MA, 2012.

[43] M. Bell, S. Perera, M. Piraveenan, M. Bliemer, T. Latty, and C. Reid, "Network growth models: a behavioural basis for attachment proportional to fitness," *Scientific Reports*, vol. 7, no. 1, 2017.

[44] S. J. Hu, X. Zhu, H. Wang, and Y. Koren, "Product variety and manufacturing complexity in assembly systems and supply chains," *CIRP Annals*, vol. 57, no. 1, pp. 45–48, 2008.

[45] C. Song, S. Havlin, and H. A. Makse, "Origins of fractality in the growth of complex networks," *Nature Physics*, vol. 2, no. 4, pp. 275–281, 2006.

[46] A. Brintrup, A. Ledwoch, and J. Barros, "Topological robustness of the global automotive industry," *Logistics Research*, vol. 9, no. 1, 2016.

[47] T. Y. Choi and Z. Wu, "Taking the leap from dyads to triads: buyer–supplier relationships in supply networks," *Journal of Purchasing and Supply Management*, vol. 15, no. 4, pp. 263–266, 2009.

[48] S. P. Borgatti and X. Li, "On social network analysis in a supply chain context*," *Journal of Supply Chain Management*, vol. 45, no. 2, pp. 5–22, 2009.

[49] R. Cohen, K. Erez, D. Ben-Avraham, and S. Havlin, "Resilience of the Internet to random breakdowns," *Physical Review Letters*, vol. 85, no. 21, pp. 4626–4628, 2000.

[50] Y. Sheffi and J. B. Rice Jr, "A supply chain view of the resilient enterprise," *MIT Sloan Management Review*, vol. 47, no. 1, p. 41, 2005.

[51] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.

[52] G. Sabidussi, "The centrality index of a graph," *Psychometrika*, vol. 31, no. 4, pp. 581–603, 1966.

Hindawi
Complexity
Volume 2018, Article ID 5180156, 15 pages
https://doi.org/10.1155/2018/5180156

Research Article

Smart Scheduling: An Integrated First Mile and Last Mile Supply Approach

Tamás Bányai, Béla Illés, and Ágota Bányai

University of Miskolc, Miskolc, Hungary

Correspondence should be addressed to Tamás Bányai; alttamas@uni-miskolc.hu

Received 21 February 2018; Accepted 9 May 2018; Published 26 July 2018

Academic Editor: Dimitris Mourtzis

Supply chain management applies more and more Industry 4.0 innovations to increase their availability, elasticity, sustainability, and efficiency. In interconnected logistics networks, operations are integrated from suppliers through 3rd party logistics providers to customers. There are different delivery models depending on the time and cost. In the last few years, a wide range of customers is willing to pay an extra fee for the same delivery or instant delivery. This fact led to the increased importance of the optimized design and control of first mile/last mile (FMLM) delivery solutions. Cyberphysical system-based service innovations make it possible to enhance the productivity of FMLM delivery in the big data environment. The design and operation problems can be described as NP-hard optimization problems. These problems can be solved using sophisticated models and methods based on heuristic and metaheuristic algorithms. This research proposes an integrated supply model of FMLM delivery. After a careful literature review, this paper introduces a mathematical model to formulate the problem of real-time smart scheduling of FMLM delivery. The integrated model includes the assignment of first mile and last mile delivery tasks to the available resources and the optimization of operations costs, while constraints like capacity, time window, and availability are taken into consideration. Next, a black hole optimization- (BHO-) based algorithm dealing with a multiobjective supply chain model is presented. The sensitivity of the enhanced algorithm is tested with benchmark functions. Numerical results with different datasets demonstrate the efficiency of the proposed model and validate the usage of Industry 4.0 inventions in FMLM delivery.

1. Introduction

The technologies of Industry 4.0 affect the connection of products, customers, production, and service companies. Digitization makes the supply chain solutions more efficient, flexible, and customer-focused. Key technologies, like smart logistics and warehousing and advanced analysis of information, will lead to the digital supply chain. In complex supply chain solutions, like first mile or last mile logistics, the application of these key technologies transforms separated supply chain solutions into interconnected logistics systems.

The increased complexity of these interconnected logistics networks needs new algorithms to solve the NP-hard optimization problems of these large-scale networks.

The design and operation of first mile and last mile supply chains include a huge number of problems: facility location, routing, scheduling, design of loading unit building and packaging processes, budgeting, warehousing, and assignment or queuing.

The model presented in this work combines the first mile and last mile operations of different package delivery companies and shows an optimization method to solve the real-time smart scheduling problem of open tasks in the interconnected logistics network. To our best knowledge, the design of FMLM supply chain has not been considered in the current literature.

The main contributions of this work include (1) an integrated model of FMLM delivery, (2) an enhanced black hole optimization-based algorithm, (3) a test of the modified black hole algorithm (BHA) with different datasets and test functions, and (4) computational results of FMLM delivery problems with different datasets.

This paper is organized as follows: Section 2 presents a literature review, which systematically summarizes the research background of FMLM supply. Section 3 describes the model framework of the FMLM supply chains including the use of Industry 4.0 innovations. Section 4 presents an enhanced black hole optimization and supposes some modification to improve the convergence and enhance its efficiency. Section 5 demonstrates the sensitivity analysis of the algorithm based on CEC 2014 functions. For our study, in Section 6, we focus on the optimization results with numerical analysis. Conclusions and future research directions are discussed in Section 7.

2. Literature Review

There exist a huge number of articles related to the FMLM supply chain design. To build a link between literatures and this research work, we are focusing on the previous research works and results to find research gaps. Our methodology of a structured literature review includes four main steps [1]:

(i) Search for articles in databases and other sources, like Scopus, Science Direct, and Web of Science.

(ii) Reduce the number of articles by reading the abstract and identify the main topic.

(iii) Define a methodology to analyze the chosen articles.

(iv) Describe the main scientific results and identify the scientific gaps and bottlenecks.

Firstly, the relevant terms were defined. It is a crucial phase of the review, because there are excellent review articles in the field of supply chain design and we did not want to produce an almost similar review, but we applied the presented methodology. We used the following keywords to search in the Scopus database: "last mile" OR "first mile" AND "supply." Initially, 137 articles were identified. This list was reduced to 50 articles selecting journal articles only. Our search was conducted in January 2018; therefore, new articles may have been published since then.

In the following step, the 50 articles were reduced after reading the abstracts. We excluded articles, in which the topics did not catch our interest and the smart scheduling of FMLM delivery cannot be addressed. After this reduction, we got 34 articles.

The reduced articles can be classified depending on the subject area. Figure 1 demonstrates the classification of these 34 articles considering ten subject areas. This classification shows the importance of multidisciplinary approach and the majority of mathematics, optimization and decision making in the design, and operation of complex, interconnected systems.

As Figure 2 demonstrates, the FMLM delivery as a new trend in supply chain solutions has been researched in the past decade. The first article in this field was published in 2005 [2], and it was focusing on customer satisfaction with order fulfillment in retail supply chains. The number of published papers has been increased in the last years; it shows the importance of this research field.

The distribution of the most frequently used keywords is depicted in Figure 3. As the keywords show, FMLM delivery is especially important for e-commerce solutions and humanitarian supply chains, and the optimization is based on many cases in integer programming, but in the case of NP-hard problems heuristic and metaheuristic solutions have to be taken into consideration.

We analyzed the articles from the point of view of scientific impact. The most usual form to evaluate articles from the point of view of scientific impact is the citation. Figure 4 shows the 10 most cited articles with their number of citations.

It remains a key challenge for supply chain solutions to make the best decisions related to FMLM delivery. In recent years, there have been many studies solving optimization problems of FMLM supply related to design and operation using bilevel multiobjective optimization with time window [3], Benders decomposition-based branch-and-cut algorithm to solve two-level stochastic problems of the last mile relief network [4], or a visual interactive simulation application for minimizing risk and improving outbound logistical efficiency in time-sensitive attended home deliveries and services [5]. Empirical evidences are also used by researches [2, 6] and comparative studies also analyze e-commerce solutions and traditional trade channels from last mile solutions, costs, energy efficiency [7], and carbon emission point of view [8]. The comparative studies show that FMLM solutions have a high share in total logistics costs [9].

Last mile delivery and last mile supply are living optimization problems in a wide range of the economy [10, 11], but the literature includes a great number of articles in the field of humanitarian logistics. Sudden onset disaster such as hurricane or earthquake creates a stochastic, chaotic environment to distribute humanitarian relief to the victims [12]. The relief items are usually delivered from temporary warehouses to the points of distribution (POD). The most important objective function is the minimization of the response time to provide required items to the victims. Integrated models and algorithms make it possible to solve multiobjective optimization problems, like facility location [13], inventory assignment, resource optimization, and routing [14]. Disaster operation can be modelled with two-stage relief chain consisting of a single staging area where donations arrive over time in uncertain quantities, which are distributed to victims located at the POD [15]. The last mile humanitarian efforts can be based on microretailers as last mile nodes, especially in Asia, where it is possible to use social enterprises as coordinators of supply chains for the distribution of items for victims [16]. Technology-enabled cooperation based on Industry 4.0 innovations can help facilitate collaboration through the supply chain [17, 18]. The supply chain operation reference (SCOR) can be applied for the metrics of logistics processes in the disaster area, but the metrics does not cover last mile options [19].

Complexity is the most important characteristic of supply chain solutions. Complexity can be measured in different fields of supply chain and logistics. The design and operation of complex supply chain processes can be described as NP-hard optimization problems [20]. These problems can be

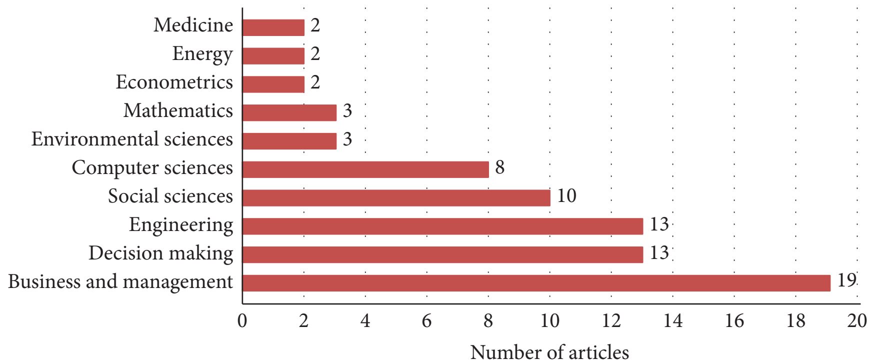


FIGURE 1: Classification of articles considering subject areas based on search in the Scopus database.

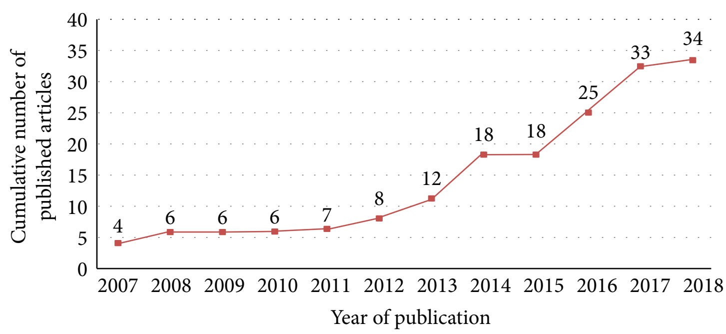


FIGURE 2: Classification of articles by year of publication.

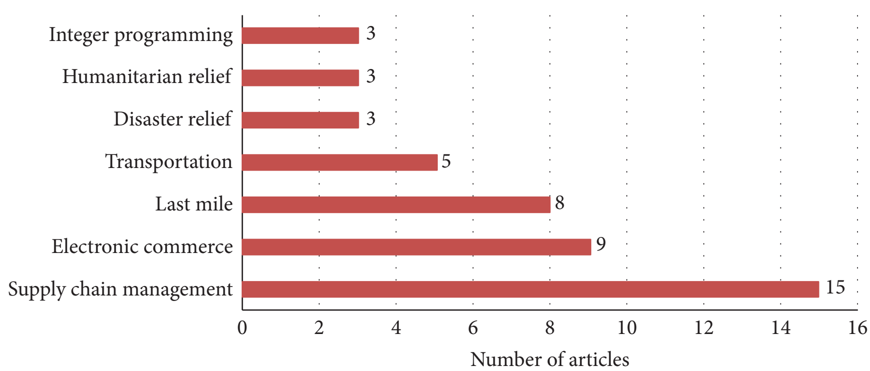


FIGURE 3: Classification of articles considering the used keywords.

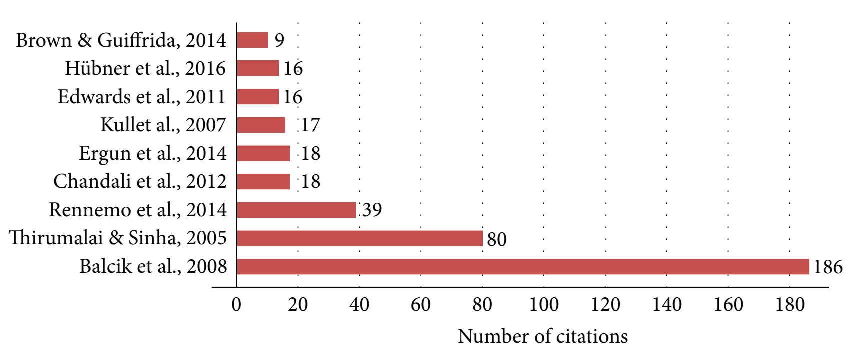


FIGURE 4: The most cited articles based on search in the Scopus database.

solved using sophisticated models and methods based on metaheuristic algorithms, like black hole optimization [21]. Inventory management systems represent another important research field, where stochastic reliability measurement and design optimization are the main research directions [22]. Different types of biomass, like energy grass or forest biomass, have gained increasing interest in the recent years as a renewable source of energy, where predictive control methodology approaches in multilayer models become more and more important [23]. Not only design but also control strategies can be represented as NP-hard optimization problems, where simulation methods can also be used to find optimal solutions [24]. Complex network topology metrics is popular to analyze social networks, like Facebook, Twitter, or ResearchGate, but there are new novel complexity indicators, like line balancing rate and number of intercell and intracell flows to describe manufacturing layouts [25]. The complexity can be influenced by many factors, but in today's economy, one of the most important influencing factors is the price competition, which influences the pricing strategy. The design and operation of blending technologies represent a special field of supply chain management because of the significant approach of technological and logistic aspects. Approaches take technological and logistic aspects into consideration, and research results confirm that outsourcing is a valuable cost-cutting tool for blending technologies [26].

Intelligent retail environment, product enhancement, interconnected supply chain, and data-driven business are the key factors of successful FMLM processes. The last mile logistics has been characterized as the most expensive part of the supply chain, featuring negative impacts on pollution and operation costs. However, in densely populated areas, the number of possible material flow paths is extremely increased, but in this case, there is a great possibility to optimize the last mile solutions. The integration of last mile and first mile solutions increased the complexity of supply chains. Improving the efficiency of FMLM delivery is a major driver for the success of e-commerce [27]. Multichannel approach, mobile commerce, and data-driven marketing have accelerated the improvement of e-commerce solutions; e-commerce intelligent systems have a great impact on the reengineering of the e-order fulfilment process [28].

The recent interest in the expansion of FMLM supply in emerging economies in Africa and Asia has been greatly debated in the literature; firm expansion [29], distribution of medicines [30], development of urban transportation and freight parking system in densely populated areas [31], and food supply chain improvement [32] are the main research topics.

The analysis, evaluation, and development of FMLM solutions are areas of research which have shown remarkable growth over the last few years. Researchers investigated new evaluation and measurement methods and tools to improve the design methods including integrated facility location and routing problems [33, 34], development of strategic planning framework [35, 36], or stochastic facility routing [37]. Third-party logistics (3PL) providers allow the management to outsource processes of supply chain functions. Their services can include a wide range of first mile or last mile operations, like warehousing, loading and unloading, packaging, shipment, and collection [38].

The optimal design and operation of FMLM processes and operations have a great environmental impact; therefore, it is important to integrate environmental aspect into the design aspects, objective functions, and constraints of the engineering and business problem. It is especially true in densely populated areas. Researches examine different solutions of online and traditional retail supply chains in order to assess their relative environmental impacts [39]. The most important field of environmental impact-related researches is focusing on services through road transport. Researches show that first mile and last mile trips with transit may increase multimodal trip emissions significantly, mitigating potential impact reductions from transit usage [40]. Environmental impact can be measured in city logistics solutions; pollution and urbanistic considerations led to a change in the use of private vehicles, and the application of new technologies in dense city centers is used more and more often, like accelerating moving walkways [41].

Heuristics and metaheuristics are used in all fields of engineering [42]. BHO is inspired by the behavior of exotic and powerful places in the space, where gravitation forces are so high that all particles were trapped reaching the event horizon. BHO is a simplified particle swarm optimization (PSO) method with inertia weight represented by gravitational forces of black hole and stars [43]. BHO is an efficient global search technique, which can be implemented in a binary, discrete, or continuous form. A binary black hole algorithm (BBHA) is used for feature selection and classification on biological data [44]. A continuous black hole algorithm is used to investigate the critical slip surface of soil slope [45] and to support the nondestructive diagnosis of wiring networks. In this research, the combination of BHA and time domain reflectometry shows that BHA can be implemented in real-world systems [46]. BHO can be combined with other heuristic algorithms to improve its efficiency and convergence. The integration of the core of swarming optimization methods can increase the efficiency of preprocessing, transfer functions, and the discretization [47].

More than 50% of the articles were published in the last 4 years. This result indicates the scientific potential of this research field. The articles that addressed the optimization of first mile and last mile deliveries are focusing on routing, facility location, strategy framework, environmental impact, and comparison, but none of the articles aimed to identify the real-time optimization aspects of FMLM supply. Therefore, smart scheduling in integrated first mile and last mile solutions still need more attention and research. According to that, the main focus of this research is the modelling and analysis of integrated FMLM supply using smart scheduling based on the use of Industry 4.0 innovations.

The aim of this paper is to investigate the effect of real-time smart scheduling on the efficiency of FMLM supply. The contribution of this paper to the literature is twofold: (1) description of an integrated model of FMLM delivery including the optimization problem of assignment of first mile and last mile delivery tasks to delivery routes and facility location and assignment and (2) development of a black-hole-based algorithm to solve the smart scheduling problem.

3. Model Framework

In traditional less than load shipping (LTLS), most deliveries are performed in the morning and pickup operations are made in the afternoon. It this case, the first mile operations (pickups) and the last mile operations (deliveries) are separated. The model framework of the smart scheduling in FMLM delivery makes it possible to analyze the possibilities of integrated handling of deliveries and pickup operations to optimize the operation costs and the utilization of resources, like package delivery trucks, drivers, and hubs. Hubs have played an important role in traditional delivery solutions, but the application of Industry 4.0 inventions makes it possible to redefine the hub and spoke centralized transport topology optimization paradigm.

There are two different types of deliveries: (1) scheduled deliveries, which are scheduled and assigned to delivery trucks, and (2) open tasks, which are not scheduled. The supply chain includes m scheduled routes with n_i locations, where i is the route ID. There are open tasks which have to be picked up and if possible delivered to the destination. The logistics system includes one hub and p spokes, where picked up packages can be stored if it is not possible to deliver the package with point-to-point transport (Figure 5).

The decision variables of this model are the following: assignment of open tasks (new packages) to scheduled routes, assignment of picked up packages to delivery routes or hubs, and scheduling of pickup operations of new packages. These decision variables include an integrated optimization problem: scheduling and assignment problem.

The decision variables describe the decisions to be made. In this model it must be decided: (a) which open tasks by which package delivery truck in which time is picked up or (b) which package delivery truck delivers the package to the destination through which hubs. These decisions represent the abovementioned assignment and scheduling problem. With this in mind, we define the following positions describing the layout of the interconnected logistics network:

(i) $p_{i,j}^{\mathrm{SC}}$ is the position of the delivery point j of the scheduled delivery route i, where $i \in (1, 2, \ldots, m)$ and $j \in (1, 2, \ldots, n_i)$.

(ii) p_k^{OP} is the position of the pickup point of the open task k, where $k \in (1, 2, \ldots, q)$.

(iii) p_k^{OD} is the position of the destination of the open task k.

(iv) p_l^{SP} is the position of spoke l, where $l \in (1, 2, \ldots, p)$.

(v) p^{HU} is the position of the hub.

The objective function of the problem describes the minimization of the costs of the whole delivery process.

$$\min \quad C = C^{\mathrm{S}} + C^{\mathrm{SF}} + C^{\mathrm{ST}} + C^{\mathrm{OP}} + C^{\mathrm{OD}}, \tag{1}$$

where C^{S} is the costs of scheduled delivery routes without any assigned open task, C^{SF} is the costs of loading and traveling from the hub or spoke to the first destination, C^{ST} is the costs of traveling from the last destination to the hub including unloading, C^{OP} is the pickup costs of assigned open tasks, and C^{OD} is the delivery cost of assigned open tasks.

The first part of the cost function (2) includes the sum of transportation costs of scheduled delivery routes without assignment of open tasks, where the transportation routes are the function of positions of delivery and pickup points and the load influences the specific transportation costs:

$$C^{\mathrm{S}} = \sum_{i=1}^{m} \sum_{j=1}^{n_i - 1} c_i \left(q_{i,j} \right) \cdot l_{i,j} \left(p_{i,j}^{SC} \right), \tag{2}$$

where c_i is the specific transportation cost of delivery truck i, $q_{i,j}$ is the load of delivery truck i passing delivery point j, and l_{ij} is the length of the transportation route i between destination j and destination $j+1$.

The second part of the cost function (2) includes the costs of loading and traveling from the hub or spoke to the first destination:

$$C^{\mathrm{SF}} = \sum_{i=1}^{m} c_i \left(q_{i,j} \right) \cdot l_{i0}. \tag{3}$$

The third part of the cost function (2) includes the costs of traveling from the last destination to the hub including unloading:

$$C^{\mathrm{ST}} = \sum_{i=1}^{m} c_i \left(q_{i,j} \right) \cdot l_{i,n_m}. \tag{4}$$

The fourth part of the cost function (2) includes the pickup costs of assigned open tasks:

$$C^{\mathrm{OP}} = \sum_{k=1}^{q} \sum_{i=1}^{m} \sum_{j=1}^{n_i+\beta_i} \mathbf{x}_{k,i,j} \cdot \left(l_{i,j,k} + l_{i,k,j+1} \right) \cdot c_i \left(q_{i,j} \right), \tag{5}$$

where β_i is the number of assigned open tasks to delivery route i, $\mathbf{x}_{k,i,j}$ is the assignment matrix of pickup operations of open tasks to scheduled delivery routes as the decision variable, $l_{i,j,k}$ is the transportation length between the scheduled destination j and the pickup destination of the open task j, and $l_{i,k,j+1}$ is the transportation length between the pickup destination of the open task and the scheduled destination $j+1$.

The fifth part of the cost function (2) includes the delivery costs of assigned open tasks:

$$C^{\mathrm{OD}} = \sum_{k=1}^{q} \sum_{i=1}^{m} \sum_{j=1}^{n_i+\beta_i} \mathbf{x}_{k,i,j}^{*} \cdot \left(l_{i,j,k} + l_{i,k,j+1} \right) \cdot c_i \left(q_{i,j} \right), \tag{6}$$

where $\mathbf{x}_{k,i,j}^{*}$ is the assignment matrix of delivery operations of open tasks to scheduled delivery routes as the decision variable. If the pickup of open task k is assigned to delivery route

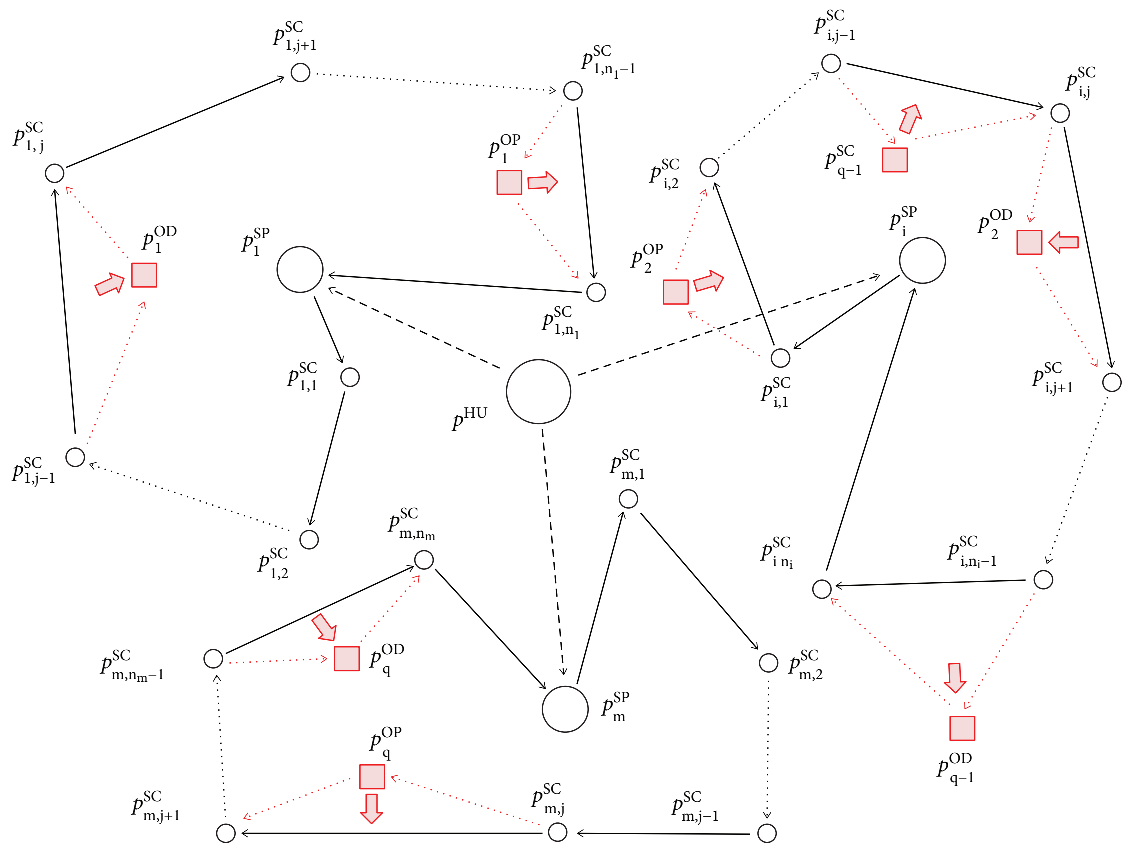


Figure 5: Model framework of FMLM supply without time window.

i after delivery point j, then $\mathbf{x}_{k,i,j} = 1$. If the delivery of open task k is assigned to delivery route i after delivery point j, then $\mathbf{x}^{*}_{k,i,j} = 1$.

The solutions of this integrated assignment and scheduling problem are limited by the following three constraints:

Constraint 1. The capacity of package delivery trucks is not to exceed after the assignment of open tasks. The new loading of delivery truck i passing pickup point j can be calculated by adding the assigned open pickup task and subtracting the value of a previously assigned delivery of an open task as follows:

$$q_{i,j} + \sum_{k=1}^{q} q_k \cdot \mathbf{x}_{k,i,j} - \sum_{k=1}^{q} q_k \cdot \mathbf{x}^{*}_{k,i,j+1} \le Q^{\max}_{i}, \quad \forall i, j, \qquad (7)$$

where $Q^{\max}_{i}$ is the maximum capacity of delivery truck i.

Constraint 2. It is not allowed to exceed the upper and lower limits of pickup and delivery operation time in each scheduled destination within the time frame.

$$\tau^{\min}_{i,j} \le \tau^{s}_{i,j} + \sum_{k=1}^{q} \mathbf{x}_{k,i,j} \cdot \left(\tau^{ao}_{i,j,k} + \tau^{ao}_{i,k,j+1}\right) \le \tau^{\max}_{i,j}, \qquad (8)$$

where $\tau^{s}_{i,j}$ is the scheduled pickup/delivery time at delivery point j of route i without any added open tasks, $\tau^{\min}_{i,j}$ is the lower limit of pickup/delivery time at delivery point j of route i, $\tau^{\max}_{i,j}$ is the upper limit of pickup/delivery time at delivery point j of route i, $\tau^{ao}_{i,j,k}$ is the traveling time between destination j of route i and destination of the open task k, and $\tau^{ao}_{i,k,j+1}$ is the traveling time between the assigned open task k and the succeeding destination of the scheduled delivery route.

Constraint 3. It is not allowed to exceed the upper and lower limits of pickup operation time in each assigned open task destination within the time frame.

$$\tau^{\text{pmin}}_{k} \le \tau^{s}_{i,j} + \sum_{k=1}^{q} \mathbf{x}_{k,i,j} \cdot \left(\tau^{ao}_{i,j,k} + \tau^{ao}_{i,k,j+1}\right) \le \tau^{\text{pmax}}_{k}, \qquad (9)$$

where τ^{pmin}_{k} is the lower limit of pickup time of the assigned open task k and τ^{pmax}_{k} is the upper limit of the pickup time of the assigned open task k.

Constraint 4. It is not allowed to exceed the upper and lower limits of delivery operation time in each assigned open task destination within the time frame.

$$\tau^{\text{dmin}}_{k} \le \tau^{s}_{i,j} + \sum_{k=1}^{q} \mathbf{x}_{k,i,j} \cdot \left(\tau^{ao}_{i,j,k} + \tau^{ao}_{i,k,j+1}\right) \le \tau^{\text{dmax}}_{k}, \qquad (10)$$

where τ_k^{dmin} is the lower limit of delivery time of the assigned open task k and τ_k^{dmax} is the upper limit of the delivery time of the assigned open task k.

It follows directly from (9) and (10) that

$$\tau_k^{\mathrm{ps}} < \tau_k^{\mathrm{ds}} \bigwedge \tau_k^{\mathrm{pmin}} \le \tau_k^{\mathrm{ps}} \le \tau_k^{\mathrm{pmax}} \bigwedge \tau_k^{\mathrm{dmin}} \le \tau_k^{\mathrm{ds}} \le \tau_k^{\mathrm{dmax}}, \tag{11}$$

where τ_k^{ps} is the scheduled pickup time of open task k and τ_k^{ds} is the scheduled delivery time of open task k.

The decision variables can only assume binary values, so we associate restrictions with the abovementioned decision variables.

$$\mathbf{x}_{k,i,j}^{*} \in (0,1) \bigwedge \mathbf{x}_{k,i,j} \in (0,1). \tag{12}$$

Figure 2 demonstrates the model framework including time window and capacity constraints. As the figure shows, the assigned open tasks have a great impact on the scheduled pickup and delivery time and can be calculated as follows:

$$\tau_{i,\xi+1}^{\mathrm{S}} = \tau_{i,\xi}^{\mathrm{S}} + \tau_{i,\xi,k}^{\mathrm{ao}} + \tau_{i,k,\xi+1}^{\mathrm{ao}}, \tag{13}$$

and in the same way, the assigned open task increases the capacity utilization of the delivery truck after pickup operation and decreases after delivery operation:

$$q_{i,\xi+1} = q_{i,\xi+1} + q_k \wedge q_{i,\mathrm{Q}+1} = q_{i,\mathrm{Q}+1} - q_k. \tag{14}$$

The time window and the capacity of delivery trucks make the optimization problem more complicated. Without constrained time window and capacity, the open tasks can be assigned to the nearest delivery point of the delivery route, as seen in Figure 6.

4. Discrete Black Hole Algorithm

Black holes are exotic and powerful places in the outer space where the gravitation forces are so high that it can trap not only particles, planets, and stars but also light. Black holes are born when stars die. Dying stars reach a point with zero volume and infinite density and become a singularity. The environment of black holes can be analyzed, but the black holes are invisible. The Schwarzschild radius is the radius of the event horizon. The distance between particles and the black hole has a great impact on the behavior of the particles. If the distance between a star and a planet is much higher than the Schwarzschild radius, then the particle can move in any directions. If this distance is larger than the Schwarzschild radius but this difference is not too much, the spacetime is deformed, and more particles are moving towards the center of the black hole than in other directions. If a particle reaches the Schwarzschild radius, then it can move only towards the center of the black hole (Figure 7). The black hole optimization is based on this phenomenon of black holes [26].

The first phase of the black hole optimization is the so-called big bang, when an initial population of stars is generated in the search space. Each star represents one candidate solution of the optimization problem. The coordinates of the star in the n-dimensional search space represent the decision variables of the n-dimensional optimization problem.

$$\overrightarrow{x}^{\mathrm{S}_i} = \left(x_1^{\mathrm{S}_i}, x_2^{\mathrm{S}_i}, \ldots, x_n^{\mathrm{S}_i}\right), \quad x_j^{\mathrm{S}_i} \in \mathbb{N}. \tag{15}$$

The second phase of the algorithm is the evaluation of the stars with an objective function.

$$\nu^{\mathrm{S}_i} = \nu^{\mathrm{S}_i}\left(x_1^{\mathrm{S}_i}, x_2^{\mathrm{S}_i}, \ldots, x_n^{\mathrm{S}_i}\right). \tag{16}$$

The third phase is to choose one black hole. It is possible to choose more than one star as the black hole, but in this case, the algorithm is almost similar as gravity force optimization. Black holes are the stars with the highest gravity force, with the highest value of objective function.

$$\nu^{\mathrm{BH}} = \max_i \left(\nu^{\mathrm{S}_i}\right). \tag{17}$$

The fourth phase of the algorithm is to move the stars towards the black hole in the search space. The operator to calculate the new position of stars takes only the gravity force between stars and the black hole into account, and the gravity force among stars is neglected. The movement of the stars is a discrete process because of the integer decision variables.

$$x_j^{\mathrm{S}_i}(t+\Delta t) = x_j^{\mathrm{S}_i}(t) + \mathrm{round}\left\{\mathrm{Rnd} \cdot \left(x_j^{\mathrm{BH}}(t) - x_j^{\mathrm{S}_i}(t)\right)\right\}. \tag{18}$$

The movement of stars towards the black hole changes the decision variables of the solution represented by the moving star so that the decision variables will move to the decision variables of the best solution represented by the black hole. Figure 8 shows the movement of stars in the case of the two-dimensional Ackley function (23).

Stars reaching the photon sphere are forced to travel in orbits. The radius of the photon sphere is written as

$$r^{\mathrm{PS}} = \frac{3 \cdot g \cdot m}{c^2} = \frac{3}{2} \cdot r^{\mathrm{SS}}, \tag{20}$$

where g is the gravitational constant, m is the mass of the black hole, and c is the speed of light in vacuum. Stars reaching the event horizon will be absorbed, and a new star, representing a new candidate solution of the optimization problem, is generated in the search space. The radius of the event horizon (the Schwarzschild radius) is calculated as follows:

$$r^{\mathrm{SS}} = \frac{f^{\mathrm{BH}}}{\sum_{i=1}^{n} f^{\mathrm{S}_i}}, \tag{21}$$

where r^{SS} is the radius of the event horizon, f^{BH} is the gravity force of the black hole, and f^{S_i} is the gravity force of the ith star.

The generation of the new stars is based on the big bang phase of the big bang big crunch (BBBC) algorithm,

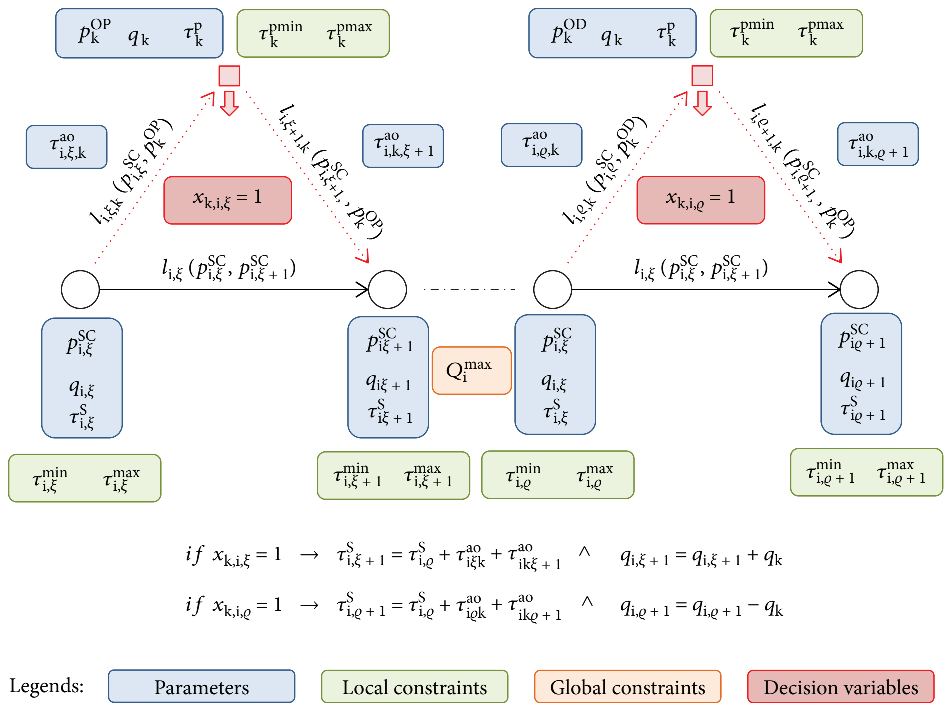


FIGURE 6: Model framework of FMLM supply with time window and capacity constraints.

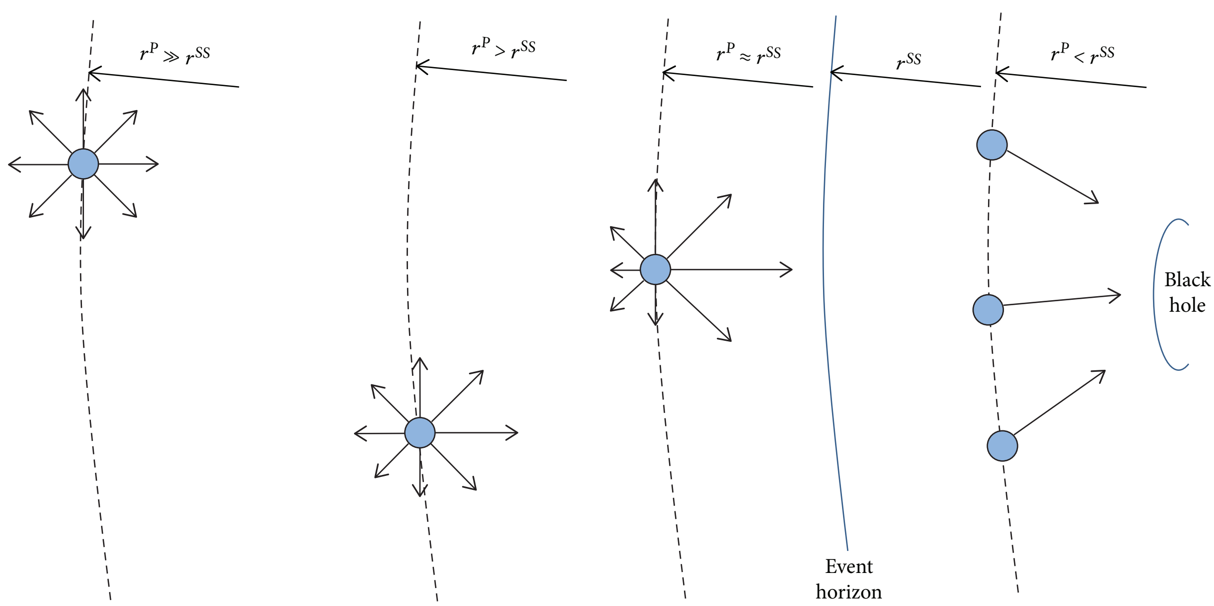


FIGURE 7: Impact of the distance between particles and event horizon (Schwarzschild radius) on the behavior of particles.

where new individuals (stars) are generated around the center of mass.

$$x_j^{S_i} = x_j^{BH} + \text{round}\left(\theta \cdot N(0,1) \cdot \frac{x_j^{S_i,\max} - x_j^{S_i,\min}}{\varepsilon}\right), \quad (22)$$

where θ is a constant, $N(0,1)$ is a random number according to a standard normal distribution, and ε is the iteration number [48].

The fifth phase is the evaluation of stars. Stars with the best gravity force become the new black holes, and the old black holes become stars. This phase of the black hole algorithm avoids trapping into local optimum. Termination criteria of the algorithm can be the number of iteration steps, computational time, or the measure of convergence.

5. Sensitivity Analysis

Within the frame of this chapter, the sensitivity analysis of the black hole algorithm is described. In continuous cases,

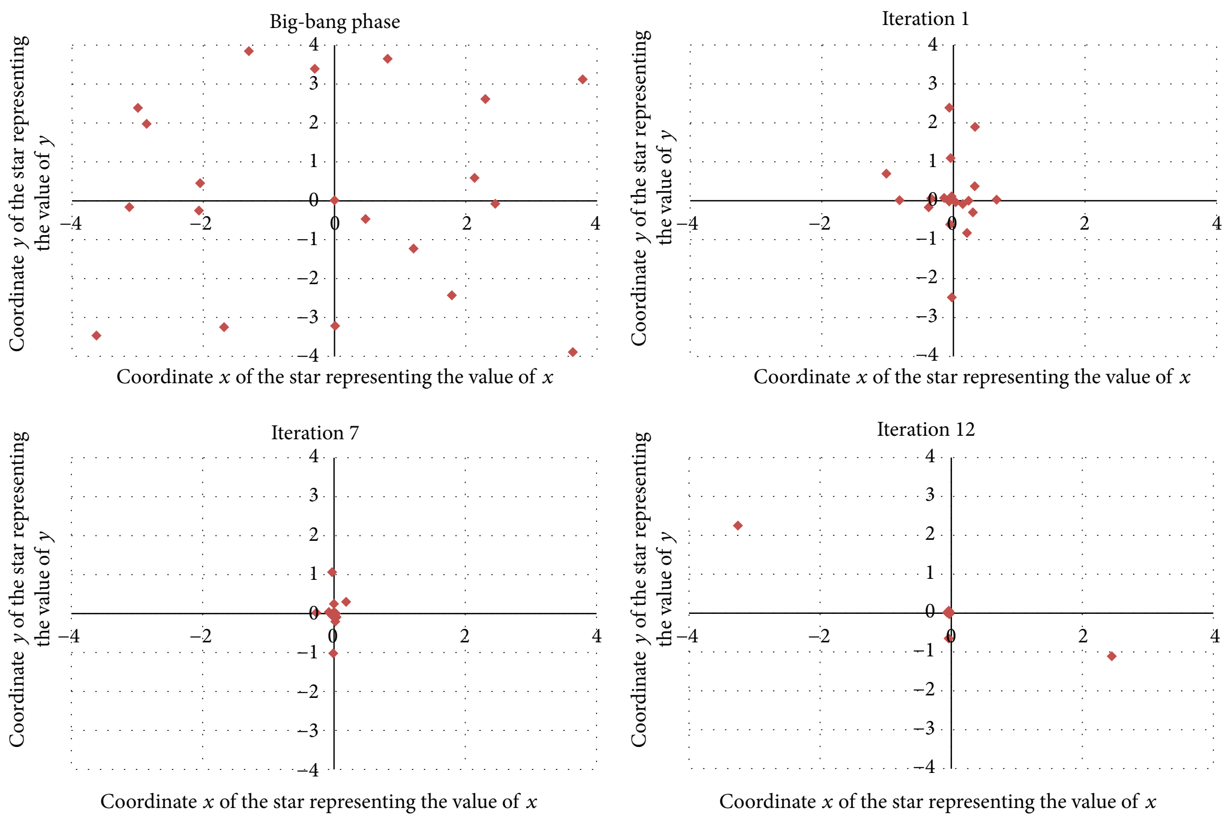


Figure 8: Movement of particles in BHO in the case of Ackley function (23).

there is a wide range of benchmarking functions to evaluate heuristics and metaheuristics [49], but in the case of discrete algorithms, these benchmarking functions can be used with constraints. Within the frame of this section, we evaluate the sensitivity of the algorithm with benchmark functions in the case of continuous problems:

(i) Ackley function:

$$v(x,y) = -20 \cdot e^{-0.2\cdot\sqrt{0.5\cdot(x^2+y^2)}} - e^{0.5\cdot(\cos 2\cdot\pi\cdot x+\cos 2\cdot\pi\cdot y)} + e + 20, \tag{23}$$

with a search domain of $x_i \in [-5, 5]$ and a global minimum of $v(0, 0) = 0$.

(ii) Bukin function:

$$v(x,y) = 100 \cdot \sqrt{\left|\frac{y-x^2}{100}\right|} + \frac{|x+10|}{100}, \tag{24}$$

with a search domain of $x \in [-15, -5] \bigwedge y \in [-3, 3]$ and a global minimum of $v(-10, 1) = 0$.

(iii) Cross-in-tray function:

$$v(x,y) = \frac{\left|\sin x \cdot \sin y \cdot e^{\left|100-\sqrt{x^2+y^2}/\pi\right|+1}\right|^{0.1}}{100}, \tag{25}$$

with a search domain of $x, y \in [-10, 10]$ and four global minimums.

(iv) Easom function:

$$v(x,y) = -\cos x \cdot \cos y \cdot e^{(x-\pi)^2+(y-\pi)^2}, \tag{26}$$

with a search domain of $x, y \in [-100,100]$ and a global minimum of $v(\pi, \pi) = -1$.

(v) Eggholder function:

$$v(x,y) = -(y+47) \cdot \sin\sqrt{\left|\frac{x}{2}+(y+47)\right|} + x \cdot \sin\sqrt{|x-(y+47)|}, \tag{27}$$

with a search domain of $x, y \in [-512, 512]$ and a global minimum of $v(512{,}404.23) = -959.64$.

(vi) Himmelblau's function:

$$v(x,y) = \left(x^2+y+11\right)^2 + \left(x+y^2-7\right)^2, \tag{28}$$

with a search domain of $x, y \in [-5, 5]$ and four global minimums.

(vii) Lévi function:

$$v(x,y) = \sin^2 3\cdot\pi\cdot x + (x-1)^2 \cdot \left(1+\sin^2 3\cdot\pi\cdot y\right) + (y-1)^2 \cdot \left(1+\sin^2 2\cdot\pi\cdot y\right), \tag{29}$$

with a search domain of $x, y \in [-10, 10]$ and a global minimum of $v(1, 1) = 0$.

(viii) Matyas function:

$$v(x, y) = 0.26 \cdot \left(x^2 + y^2\right) - 0.48 \cdot x \cdot y, \quad (30)$$

with a search domain of $x_i \in [-10, 10]$ and a global minimum of $v(0, 0) = 0$.

(ix) Modified sphere function:

$$v(\overrightarrow{x}) = \sum_{i=1}^{n} x_i^n, \quad (31)$$

with a search domain of $x_i \in [-\infty, \infty]$ and a global minimum of $v(0, 0) = 0$.

(x) Three-hump camel function:

$$v(x, y) = 2 \cdot x^2 - 1.05 \cdot x^4 + \frac{x^6}{6} + x \cdot y + y^2, \quad (32)$$

with a search domain $x, y \in [-5, 5]$ and a global minimum of $v(0, 0) = 0$.

The purpose of this evaluation is to analyze the effect of the combination of the standard BHO with the BBBC algorithm (22) and to compare with the genetic algorithm (GA) and the harmony search algorithm (HSA).

As Table 1 demonstrates, the combination of the BHO and BBBC algorithms decreases the error value after 50 iteration steps. As Figure 9 shows, the results of BHBBBC algorithm are comparable with the other three algorithms.

Table 2 demonstrates the impact of the numbers of decision variables on the required iteration steps to reach the defined accuracy. We added some n-dimensional benchmark function as follows:

(xi) Elliptic function:

$$v(\overrightarrow{x}) = \sum_{i=1}^{n} \left(10^6\right)^{(i-1)/(n-1)} \cdot x_i^2, \quad (33)$$

with a search domain of $x_i \in [-100,100]$ and a global minimum of $v(0, \ldots, 0) = 0$.

(xii) Rosenbrock function:

$$v(\overrightarrow{x}) = \sum_{i=1}^{n-1} \left(100 \cdot \left(x_{i+1} - x_i^2\right)^2 + (x_i - 1)^2\right), \quad (34)$$

with a search domain of $x_i \in [-\infty, \infty]$ and a global minimum of $v(1, \ldots, 1) = 0$.

(xiii) Styblinski-Tang function:

$$v(\overrightarrow{x}) = \frac{1}{2} \cdot \sum_{i=1}^{n} \left(x_i^4 - 16 \cdot x_i^2 + 5 \cdot x_i\right), \quad (35)$$

with a search domain of $x_i \in [-5, 5]$ and $v(2.903534, \ldots, 2.903534) = -39.16599$.

As Table 2 shows, the increased size of the problem led to the increase of the required iteration steps to reach the defined accuracy of $a = 5 \cdot 10^{-6}$.

6. Smart Scheduling of FMLM Delivery

Within the frame of this chapter, case studies are analyzed. The purpose of this chapter is to analyze the smart scheduling possibilities of smart scheduling to validate the usage of Industry 4.0 inventions in FMLM delivery. Scenario 1 is described in Figure 10.

Scenario 1 is a simple model with time window and constrained loading capacity, where the usability of the above-mentioned algorithm is demonstrated. Scenario 1 includes three different package delivery routes of three different companies. The purpose of this demo problem is to assign the open task (pickup as first mile operation) to a possible route and deliver the product to the destination (delivery as last mile operation) to minimize the costs of the delivery (1), while constraint, like time window and capacity, is taken into consideration ((7), (8), (9), and (10)). Table 3 shows the time window and the scheduled delivery and pickup times.

The loading capacity of each delivery trucks is 180 large brown postal boxes (430 mm × 300 mm × 180 mm), and the initial loadings at the first destinations are $q_{1,1} = q_{2,1} = q_{3,1} = 100$. The pickup time of the open task is between 9:00 and 12:00, and the delivery time must be between 12:00 and 15:00. Table 4 demonstrates the quantity of scheduled pickup and delivery operations.

This simple optimization problem can be solved with the 1-dimensional version of the above-described BHBBBC algorithm. The optimization algorithm resulted the following: the open task can be picked up after the 2nd destination of the 2nd route, and the delivery is after the 4th destination (Figure 11).

Scenario 2 shows the optimization results of a larger system. There are four routes from four different package delivery services. The initial routes were scheduled without any cooperation. The scenario demonstrates the real-time smart scheduling possibility of four open tasks using floating car data captured from triangulation, vehicle reidentification, GPS-based methods, or smartphone-based monitoring (Figure 12).

As Figure 7 shows, the assignment of open tasks takes into consideration not only the time window but also the constrained capacity of delivery trucks. In the case of open task 2, the pickup operation is scheduled to the nearest scheduled delivery destination $p_{2,4}^{SC}$, but the delivery operation cannot be scheduled after the nearest scheduled delivery $p_{2,12}^{SC}$, because of the constrained truck capacity:

$$l_{2,10,2}\left(p_{2,10,2}^{SC}, p_{2,11,2}^{SC}, p_2^{OP}\right) > l_{2,12,2}\left(p_{2,13,2}^{SC}, p_{2,14,2}^{SC}, p_2^{OP}\right). \quad (36)$$

As Figure 13 shows, in the case of Scenario 2, the nearest pickup and delivery destination would lead to the overload of the delivery truck, while in the case of a shifted delivery destination, the loading capacity is not exceeded.

Table 1: Error values of BHO in the case of 10 benchmark functions after 50 iteration steps.

Evaluation function	Standard BHO	BHO & BBBC	Genetic algorithm	Harmony search
Ackley (23)	$3.66E-07$	$4.05E-11$	$4.67E-06$	$1.28E-07$
Bukin (24)	$2.45E-06$	$3.58E-12$	$5.45E-07$	$9.08E-07$
Cross-in-tray (25)	$8.55E-09$	$9.24E-11$	$7.32E-09$	$6.98E-08$
Easom (26)	$1.18E-05$	$1.05E-10$	$2.09E-04$	$8.18E-09$
Eggholder (27)	$5.50E-07$	$8.88E-14$	$3.12E-07$	$1.98E-08$
Himmelblau (28)	$5.79E-08$	$9.14E-15$	$2.25E-06$	$1.05E-08$
Lévi (29)	$1.20E-06$	$7.46E-09$	$7.34E-08$	$3.12E-08$
Matyas (30)	$9.12E-08$	$7.59E-11$	$1.78E-07$	$6.70E-09$
Modified sphere (31)	$2.21E-08$	$4.22E-10$	$1.93E-06$	$2.40E-08$
Three hump camel (32)	$1.51E-06$	$8.06E-13$	$4.17E-08$	$7.79E-010$

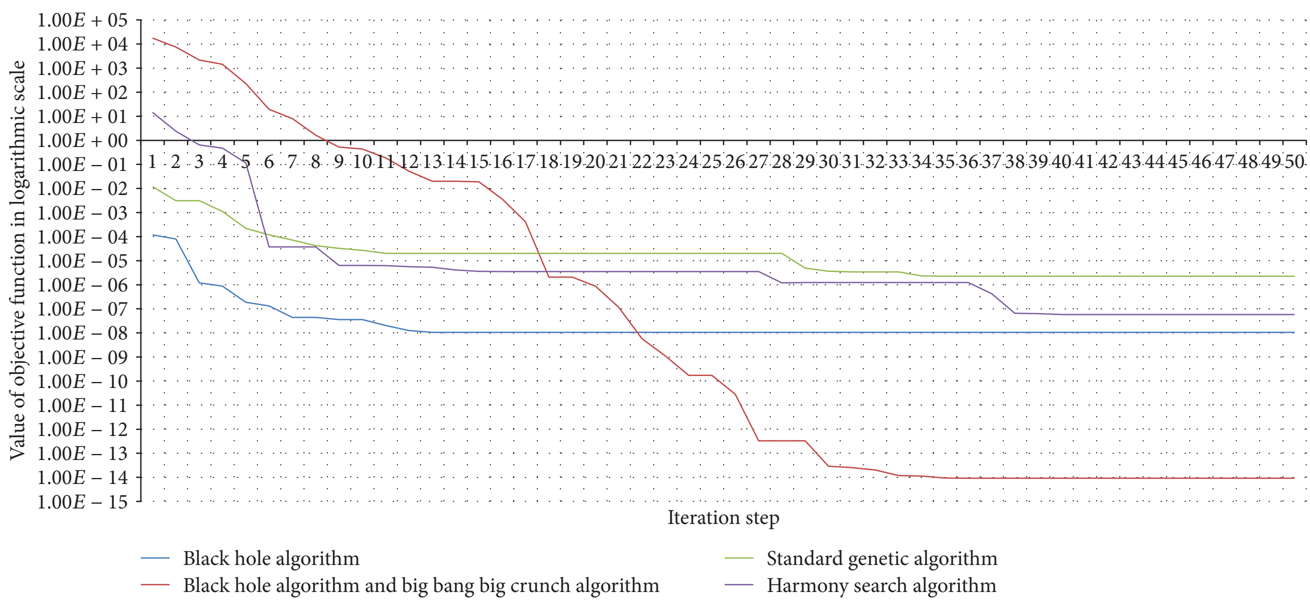


Figure 9: Convergence of algorithms in the case of Himmelblau's function (28).

Table 2: Number of required iteration steps to reach the predefined accuracy.

Evaluation function	$n=2$	$n=10$	$n=20$	$n=100$
Elliptic (33)	25	56	119	525
Rosenbrock (34)	34	47	155	712
Styblinski-Tang (35)	42	102	211	1232

The above-described optimization of the first mile last mile supply chain can lead to increased efficiency, flexibility, and availability, while the value of each environmental indicators and operation costs are decreased.

7. Conclusions and Further Research Directions

Industry 4.0 solutions make it possible to improve traditional supply chain solution in hyperconnected logistics systems. This study developed a methodological approach for real-time smart scheduling of the first mile last mile delivery of cooperating delivery companies. In this paper, firstly, we review and systematically categorized the recent works presented for the design of FMLM supply. Then, motivated from the gaps in the literature, a model for cooperating FMLM supply is developed. We proposed a general model. The described model includes different delivery routes of different companies, where the cooperation is based on Industry 4.0 solution including vehicle reidentification, GPS-based methods, and smartphone-based monitoring. The smart scheduling means the real-time optimization of the assignment of open tasks to the scheduled routes depending on the captured information from the running processes. The smart scheduling problem was solved with a newly developed metaheuristic combining the BHO and BBBC algorithms. The sensitivity analysis showed the efficiency of the integration of both swarming heuristics.

The scientific contributions of this paper are the following: a model for the integrated real-time scheduling of first mile and last mile operations in a package delivery

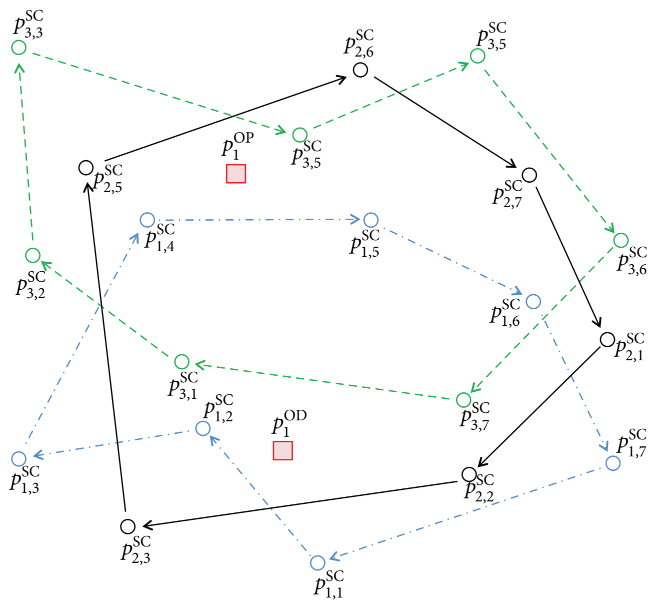


FIGURE 10: Demo problem of FMLM delivery with one open task, time window, and constrained capacity.

TABLE 3: Time window and scheduled delivery time of delivery routes (lower limit–scheduled time–upper limit).

	Destination 1	Destination 2	Destination 3	Destination 4	Destination 5	Destination 6	Destination 7
Route 1	9:30–**9:50**–10:10	9:50–**10:40**–10:45	10:30–**11:15**–12:00	11:30–**12:20**–13:00	12:00–**12:55**–13:30	10:00–**13:30**–14:00	—
Route 2	9:00–**9:50**–10:00	9:00–**10:20**–12:45	10:00–**11:00**–13:00	11:30–**12:00**–14:00	12:00–**12:35**–14:30	10:00–**13:30**–15:00	13:00–**14:00**–15:20
Route 3	10:00–**10:00**–11:10	9:50–**10:15**–11:45	11:30–**11:05**–12:50	12:10–**12:20**–15:00	13:00–**13:25**–16:00	10:00–**13:50**–16:00	12:00–**14:30**–16:30

TABLE 4: Scheduled number of boxes to be picked up and delivered (pickup/delivery).

	Destination 1	Destination 2	Destination 3	Destination 4	Destination 5	Destination 6	Destination 7
Route 1	30/—	—/20	20/—	—/30	80/—	—/30	—
Route 2	20/—	—/30	40/—	—/10	10/—	20/—	30/—
Route 3	10/—	20/—	30/—	—/30	40/—	—/60	30/—

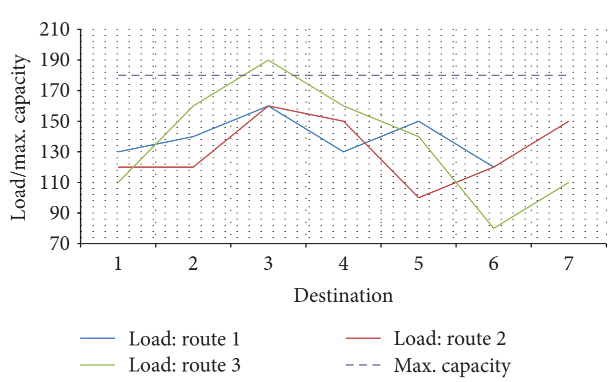


FIGURE 11: Load of each route after scheduling the open task for the route (route 3 is not available for the open task, because the load exceed the maximum loading capacity at destination 3).

environment, where the hyperconnected operation is based on Industry 4.0 solutions, and a new metaheuristic combining the black hole optimization and the big bang big crunch algorithm. The results can be generalized, because the model can be applied to different supply chain applications, especially in the case of a multitier supply chain for the automotive industry. The described methods make it possible to support managerial decisions; the operation strategy of the package delivery companies and the cooperation contract among them can be influenced by the results of the above-described contribution.

However, there are also directions for further research. In further studies, the model can be extended to a more complex model including additional constraints, like availability of human resources or the stochasticity of the parameters. This study only considered the BHBBBC optimization as the possible solution algorithm for the described NP-hard problem.

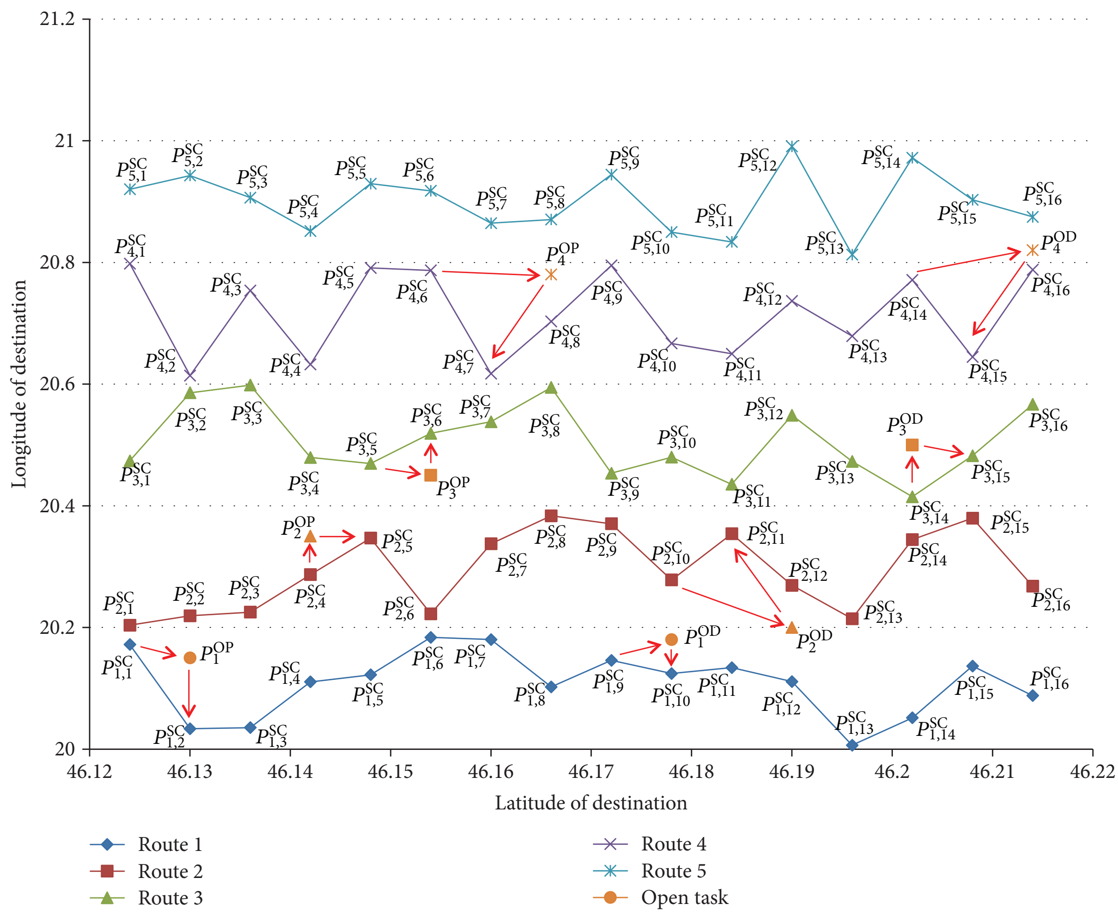
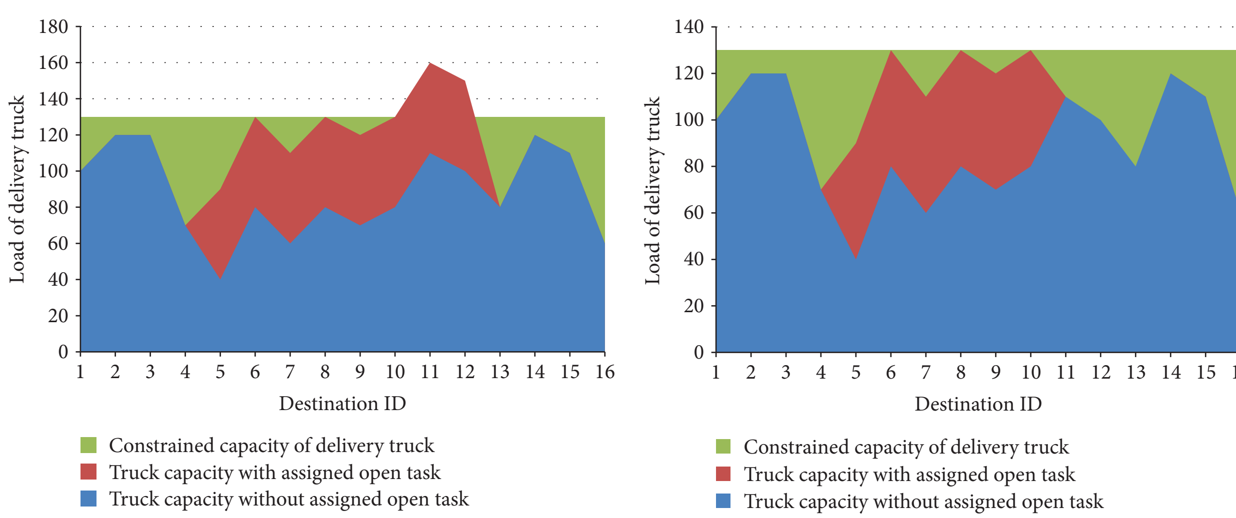


FIGURE 12: Scenario 2 with five scheduled delivery routes and four open tasks for real-time smart scheduling.



(a)

(b)

FIGURE 13: The load of the delivery truck depending on the assigned pickup and delivery destination. (a) Delivery from the nearest destination. (b) Delivery of open task from a shifted destination.

In reality, other heuristic methods can be also suitable for the solution of this problem. The convergence of the described algorithm can be improved using enhanced operators, and the behavior of the described metaheuristics to other optimization approaches can be tested. This should be also considered in the future research.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement no. 691942. This research was partially carried out in the framework of the Center of Excellence of Mechatronics and Logistics at the University of Miskolc.

References

[1] M. Lage Junior and M. Godinho Filho, "Variations of the Kanban system: literature review and classification," *International Journal of Production Economics*, vol. 125, no. 1, pp. 13–21, 2010.

[2] S. Thirumalai and K. K. Sinha, "Customer satisfaction with order fulfillment in retail supply chains: implications of product type in electronic B2C transactions," *Journal of Operations Management*, vol. 23, no. 3-4, pp. 291–303, 2005.

[3] Y. Ji, S. Qu, and Z. Yu, "Bi-level multi-objective optimization model for last mile delivery using a discrete approach," *Journal of Difference Equations and Applications*, vol. 23, no. 1-2, pp. 179–190, 2017.

[4] N. Noyan, B. Balcik, and S. Atakan, "A stochastic optimization model for designing last mile relief networks," *Transportation Science*, vol. 50, no. 3, pp. 1092–1113, 2016.

[5] K. Herrel, "A visual interactive simulation application for minimizing risk and improving outbound logistical efficiency in time-sensitive attended home deliveries and services," *SIMULATION*, vol. 90, no. 4, pp. 377–404, 2014.

[6] T. J. Kull, K. Boyer, and R. Calantone, "Last-mile supply chain efficiency: an analysis of learning curves in online ordering," *International Journal of Operations & Production Management*, vol. 27, no. 4, pp. 409–434, 2007.

[7] H. Pålsson, F. Pettersson, and L. Winslott Hiselius, "Energy consumption in e-commerce versus conventional trade channels - insights into packaging, the last mile, unsold products and product returns," *Journal of Cleaner Production*, vol. 164, pp. 765–778, 2017.

[8] J. R. Brown and A. L. Guiffrida, "Carbon emissions comparison of last mile delivery versus customer pickup," *International Journal of Logistics Research and Applications*, vol. 17, no. 6, pp. 503–521, 2014.

[9] T. Vanelslander, L. Deketele, and D. van Hove, "Commonly used e-commerce supply chains for fast moving consumer goods: comparison and suggestions for improvement," *International Journal of Logistics Research and Applications*, vol. 16, no. 3, pp. 243–256, 2013.

[10] S. F. W. T. Lim, E. Rabinovich, D. S. Rogers, and T. M. Lasester, "Last-mile supply network distribution in omnichannel retailing: a configuration-based typology," *Foundations and Trends® in Technology, Information and Operations Management*, vol. 10, no. 1, pp. 1–87, 2016.

[11] H. Wang and A. Odoni, "Approximating the performance of a "last mile" transportation system," *Transportation Science*, vol. 50, no. 2, pp. 659–675, 2016.

[12] M. G. Hilgers, C. C. Elrod, and J. B. Kampunzu, "Supply chain dynamics relief of sudden-onset disasters," *International Journal of Emergency Management*, vol. 9, no. 2, pp. 93–112, 2013.

[13] B. Balcik, B. M. Beamon, and K. Smilowitz, "Last mile distribution in humanitarian relief," *Journal of Intelligent Transportation Systems*, vol. 12, no. 2, pp. 51–63, 2008.

[14] M. Vanajakumari, S. Kumar, and S. Gupta, "An integrated logistic model for predictable disasters," *Production and Operations Management*, vol. 25, no. 5, pp. 791–811, 2016.

[15] R. A. Cook and E. J. Lodree, "Dispatching policies for last-mile distribution with stochastic supply and demand," *Transportation Research Part E: Logistics and Transportation Review*, vol. 106, pp. 353–371, 2017.

[16] M. S. Sodhi and C. S. Tang, "Buttressing supply chains against floods in Asia for humanitarian relief and economic recovery," *Production and Operations Management*, vol. 23, no. 6, pp. 938–950, 2014.

[17] O. Ergun, L. Gui, J. L. Heier Stamm, P. Keskinocak, and J. Swann, "Improving humanitarian operations through technology-enabled collaboration," *Production and Operations Management*, vol. 23, no. 6, pp. 1002–1014, 2014.

[18] K. J. Poppe, S. Wolfert, C. Verdouw, and T. Verwaart, "Information and communication technology as a driver for change in agri-food chains," *EuroChoices*, vol. 12, no. 1, pp. 60–65, 2013.

[19] Q. Lu, M. Goh, and R. De Souza, "A SCOR framework to measure logistics performance of humanitarian organizations," *Journal of Humanitarian Logistics and Supply Chain Management*, vol. 6, no. 2, pp. 222–239, 2016.

[20] T. Bányai, P. Veres, and B. Illés, "Heuristic supply chain optimization of networked maintenance companies," *Procedia Engineering*, vol. 100, pp. 46–55, 2015.

[21] P. Veres, T. Bányai, and B. Illés, "Optimization of in-plant production supply with black hole algorithm," *Solid State Phenomena*, vol. 261, pp. 503–508, 2017.

[22] J. Korponai, Á. B. Tóth, and B. Illés, "Context of the inventory management expenses in the case of planned shortages," *Engineering Management in Production and Services*, vol. 9, no. 1, pp. 26–35, 2017.

[23] T. M. Pinho, J. P. Coelho, G. Veiga, A. P. Moreira, and J. Boaventura-Cunha, "A multilayer model predictive control methodology applied to a biomass supply chain operational level," *Complexity*, vol. 2017, Article ID 5402896, 10 pages, 2017.

[24] R. Elhabyan, W. Shi, and M. St-Hilaire, "A Pareto optimization-based approach to clustering and routing in wireless sensor networks," *Journal of Network and Computer Applications*, vol. 114, pp. 57–69, 2018.

[25] V. Modrak and Z. Soltysova, "Novel complexity Indicator of manufacturing process chains and its relations to indirect complexity indicators," *Complexity*, vol. 2017, Article ID 9102824, 15 pages, 2017.

[26] T. Bányai, "Supply chain optimization of outsourced blending technologies," *Journal of Applied Economic Sciences*, vol. 12, no. 4(50), pp. 960–976, 2017.

[27] I. D. Cardenas, W. Dewulf, J. Beckers, C. Smet, and T. Vanelslander, "The e-commerce parcel delivery market and the implications of home B2C deliveries vs pick-up points," *International Journal of Transport Economics*, vol. 44, no. 2, pp. 235–256, 2017.

[28] K. H. Leung, K. L. Choy, P. K. Y. Siu, G. T. S. Ho, H. Y. Lam, and C. K. M. Lee, “A B2C e-commerce intelligent system for re-engineering the e-order fulfilment process,” *Expert Systems with Applications*, vol. 91, pp. 386–401, 2018.

[29] D. Weatherspoon and A. Ross, “Designing the last mile of the supply chain in Africa: firm expansion and managerial inferences from a grocer model of location decisions,” *International Food and Agribusiness Management Review*, vol. 11, no. 1, pp. 1–16, 2008.

[30] Y. Chandani, A. Pomeroy, S. Andersson, M. Noel, M. K. Pahl, and T. Williams, “Factors affecting availability of essential medicines among community health workers in Ethiopia, Malawi, and Rwanda: solving the last mile puzzle,” *The American Journal of Tropical Medicine and Hygiene*, vol. 87, 5_Supplement, pp. 120–126, 2012.

[31] L. Malik, I. Sánchez-Díaz, G. Tiwari, and J. Woxenius, “Urban freight-parking practices: the cases of Gothenburg (Sweden) and Delhi (India),” *Research in Transportation Business & Management*, vol. 24, pp. 37–48, 2017.

[32] M. M. Dome and S. Prusty, “Determination of vegetable post-harvest loss in the last-mile supply chain in Tanzania: a lean perspective,” *International Journal of Logistics Systems and Management*, vol. 27, no. 2, pp. 133–150, 2017.

[33] T. Aized and J. S. Srai, “Hierarchical modelling of last mile logistic distribution system,” *The International Journal of Advanced Manufacturing Technology*, vol. 70, no. 5-8, pp. 1053–1061, 2014.

[34] L. Zhou, X. Wang, L. Ni, and Y. Lin, “Location-routing problem with simultaneous home delivery and customer's pickup for city distribution of online shopping purchases,” *Sustainability*, vol. 8, no. 8, p. 828, 2016.

[35] A. Hübner, H. Kuhn, and J. Wollenburg, “Last mile fulfilment and distribution in omni-channel grocery retailing,” *International Journal of Retail & Distribution Management: A strategic planning framework*, vol. 44, no. 3, pp. 228–247, 2016.

[36] H. W. T. So, A. Gunasekaran, and W. W. C. Chung, “Last mile fulfilment strategy for competitive advantage,” *International Journal of Logistics Systems and Management*, vol. 2, no. 4, pp. 404–418, 2006.

[37] S. J. Rennemo, K. F. Rø, L. M. Hvattum, and G. Tirado, “A three-stage stochastic facility routing model for disaster response planning,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 62, pp. 116–135, 2014.

[38] P. F. Publow, “Consider third party logistics to address your company’s challenges,” *Canadian Apparel*, vol. 31, no. 3, pp. 8-9+37, 2007.

[39] J. Edwards, A. McKinnon, and S. Cullinane, “Comparative carbon auditing of conventional and online retail supply chains: a review of methodological issues,” *Supply Chain Management: An International Journal*, vol. 16, no. 1, pp. 57–63, 2011.

[40] C. G. Hoehne and M. V. Chester, “Greenhouse gas and air quality effects of auto first-last mile use with transit,” *Transportation Research Part D: Transport and Environment*, vol. 53, pp. 306–320, 2017.

[41] R. Scarinci, I. Markov, and M. Bierlaire, “Network design of a transport system based on accelerating moving walkways,” *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 310–328, 2017.

[42] L. Kota and K. Jármai, “Mathematical modeling of multiple tour multiple traveling salesman problem using evolutionary programming,” *Applied Mathematical Modelling*, vol. 39, no. 12, pp. 3410–3433, 2015.

[43] A. P. Piotrowski, J. J. Napiorkowski, and P. M. Rowinski, “How novel is the “novel” black hole optimization approach?,” *Information Sciences*, vol. 267, pp. 191–200, 2014.

[44] E. Pashaei and N. Aydin, “Binary black hole algorithm for feature selection and classification on biological data,” *Applied Soft Computing*, vol. 56, pp. 94–106, 2017.

[45] W. Gao, “Investigating the critical slip surface of soil slope based on an improved black hole algorithm,” *Soils and Foundations*, vol. 57, no. 6, pp. 988–1001, 2017.

[46] M. K. Smail, H. R. E. H. Bouchekara, L. Pichon, H. Boudjefdjouf, A. Amloune, and Z. Lacheheb, “Non-destructive diagnosis of wiring networks using time domain reflectometry and an improved black hole algorithm,” *Nondestructive Testing and Evaluation*, vol. 32, no. 3, pp. 286–300, 2017.

[47] R. Soto, B. Crawford, R. Olivares et al., “Solving the non-unicost set covering problem by using cuckoo search and black hole optimization,” *Natural Computing*, vol. 16, no. 2, pp. 213–229, 2017.

[48] O. Hasançebi and S. Kazemzadeh Azad, “An exponential big bang-big crunch algorithm for discrete design optimization of steel frames,” *Computers & Structures*, vol. 110-111, pp. 167–179, 2012.

[49] S.-Z. Zhao, P. N. Suganthan, and S. Das, “Self-adaptive differential evolution with modified multi-trajectory search for CEC’2010 large scale optimization,” in *Swarm, Evolutionary, and Memetic Computing. SEMCCO 2010*, B. K. Panigrahi, S. Das, P. N. Suganthan, and S. S. Dash, Eds., vol. 6466 of Lecture Notes in Computer Science, pp. 1–10, Springer, Berlin, Heidelberg, 2010.

Hindawi
Complexity
Volume 2018, Article ID 6203754, 15 pages
https://doi.org/10.1155/2018/6203754

Research Article

Multilayer Network-Based Production Flow Analysis

Tamás Ruppert, Gergely Honti, and János Abonyi

MTA-PE "Lendület" Complex Systems Monitoring Research Group, Department of Process Engineering, University of Pannonia, Veszprém, Hungary

Correspondence should be addressed to János Abonyi; janos@abonyilab.com

Received 24 February 2018; Revised 18 May 2018; Accepted 27 May 2018; Published 16 July 2018

Academic Editor: Dimitris Mourtzis

A multilayer network model for the exploratory analysis of production technologies is proposed. To represent the relationship between products, parts, machines, resources, operators, and skills, standardized production and product-relevant data are transformed into a set of bi- and multipartite networks. This representation is beneficial in production flow analysis (PFA) that is used to identify improvement opportunities by grouping similar groups of products, components, and machines. It is demonstrated that the goal-oriented mapping and modularity-based clustering of multilayer networks can serve as a readily applicable and interpretable decision support tool for PFA, and the analysis of the degrees and correlations of a node can identify critically important skills and resources. The applicability of the proposed methodology is demonstrated by a well-documented benchmark problem of a wire-harness production process. The results confirm that the proposed multilayer network can support the standardized integration of production-relevant data and exploratory analysis of strongly interconnected production systems.

1. Introduction

Industry 4.0 is a strategic approach to design optimal production flows by integrating flexible and agile manufacturing systems with Industrial Internet of Things (IIoT) technology [1] enabling communication between people, products, and complex systems [2–4]. The integration of manufacturing and information systems is, however, a challenging task [5]. Horizontal and intercompany integration should connect the elements of the supply chain [6], while vertical integration should connect information related to the entire product life cycle [7]. According to this new concept, the improvement and optimization of production technologies based on cyber-physical systems (CPS) are realized by the simultaneous utilization of information related to production systems [8], products, models [9], simulators, and process data [10, 11].

CPS- and Industry 4.0-type solutions also enable the compositions of smaller cells providing more flexibility with regard to production [12]. This idea leads to decentralized manufacturing [13] and emerging next generation machine systems [14]. This trend highlights the importance of the relationship between flexibility and complexity [15].

The complexity of production systems can be divided into the physical and functional domains [16]. To analyze this aspect, our focus is on the production flow analysis of production systems as production analysis has multiple perspectives according to the hierarchical decomposition of the production system: (1) production flow analysis studies the activities needed to make each part and machines to be used to simplify the material flow. (2) Company flow analysis studies the flow of materials between different factories to develop an efficient system in which each facility completes all the parts it makes. (3) Factory flow analysis plans the division of the factory into groups or departments each of which manufactures all the parts it makes and plans a simple unidirectional flow system by joining these departments. (4) Group analysis divides each department into groups, each of which completes all the parts it makes—groups which complete parts with no backflow, crossflow (between groups), and no need to buy any additional equipment. (5) Line analysis analyzes the flow of materials between the machines in each group to identify shortcuts in the plant layout, and (6) tooling analysis tries to minimize setup time by finding sequences that minimize the required additional tooling for the following job [17].

Production flow analysis (PFA) is a technique to identify both groups and their associated "families" by analyzing the information in component process routes which show the activities (often referred as operations) needed to make each part and the machines to be used for each activity [18, 19]. Every production flow analysis begins with data gathering during which nonvalue adding activity should be optimized [20]. When dealing with large quantities of manufacturing data, a representational schema that can efficiently represent structurally diverse and dynamical system have to be taken into consideration. Standards like ISO 18629, 10303 (STEP), and 15531 (MANDATE) support information flow by standardizing the description of production processes [21]. Based on these standards and web semantics, a manufacturing system engineering (MSE) knowledge representation scheme, called an MSE ontology model, was developed as a modeling tool for production [22]. The MSE ontology model by its very nature can be interpreted as a labeled network.

A simple multidimensional representation is proposed that can unfold the complex relationships of production systems. Network models are ideal to represent connections between objects and properties [23]. However, as a multidimensional problem that requires flexibility due to the continuously growing amount of information is in question and a new multidimensional approach in the form of a multilayer network [24] is presented.

For the analysis of the resultant ontology-driven labeled multilayer network, techniques to facilitate cell formation and competency assignment for operators were developed.

Manufacturing cell formation aims to create manufacturing cells from a given number of machines and products by partitioning similar machines which produce similar products. Standard cell formation problems handle products and machines while their connections are represented by two-layered bipartite graphs or machines-products incidence matrices. Classical algorithms are based on clustering and seriation of the incidence matrices. Recently, various alternative algorithms have been developed, for example, self-organizing maps [25] of fuzzy clustering-based methods [26]. What is common in most of these approaches is that they only take two variables into account [27]. However, complex manufacturing processes should be characterized by numerous properties, like the type of products and resources, and the required skills of operators should be also taken into account at successful line balancing since the skills of the operators are influencing the speed of the conveyor belt [28]. Dynamic job rotation [29] also requires efficient allocation of the assembly tasks while taking into account the constraints related to the available skills of the operators.

To handle these elements of the production line, the traditional cell formation problem was extended into a multidimensional one. The main idea is to represent these problems by multilayered graphs and apply modularity analysis to identify the groups of items that could be handled together to improve the production process.

An entirely reproducible benchmark problem was designed to demonstrate our methodology. As an example, the problem of process flow analysis of wire-harness production was selected as this product is complex and varies significantly [30] as the geometries and components of the harness vary depending on the final products [31]. Since there are challenges in the selection of the cost-effective design [32] and the demand for flexibility and a short delivery time urge the definition of product families produced from the submodules [33], the problem requires the advanced integration of process- and product-relevant information.

The remaining part of the paper is structured as follows. In Section 2, a multilayer network model is formalized that was developed to represent production systems. In Section 3, how production flow analysis problems can be interpreted as network analysis tasks is discussed. Section 3.1 describes the applicability of network science in PFA. Section 3.2 formalizes the projection of the multilayer networks and studies how conditional connections can be defined, while Section 3.3 applies this projection to calculate the node similarities. The group formation task is described in Section 3.4, where the results of this approach on benchmark examples are also presented. The detailed case study starts in Section 4 with the definition of the wire-harness production use case. The details of the problem are given in the Appendix. Section 4.1 demonstrates the applicability of similarity and modularity analysis. The workload analysis is given in Section 4.2, while interesting applications related to the evaluation of the flexibility of operator-task assignment problems are discussed in Section 4.3. Finally, conclusions are drawn in Section 5.

2. Multilayer-Network Representation of Production Systems

Essential information about the products to be assembled, parts to be manufactured, materials to be used, methods and techniques to convert the material to the required finished components, and manpower to operate the plant is usually available to a company, but rarely in an appropriate form for ease of digestion by the manager [34]. In this section, we propose a network-based model to study the relationship between these elements.

As can be seen in Figure 1, the proposed network consists of a set of bipartite graphs representing connections between the sets of products $\mathbf{p} = \{p_1, \dots, p_{N_p}\}$, machines/workstations $\mathbf{w} = \{w_1, \dots, w_{N_w}\}$, parts/components $\mathbf{c} = \{C_1, \dots, C_{N_C}\}$, activities (operations) $\mathbf{a} = \{a_1, \dots, a_{N_a}\}$, and their categorical properties (referred as activity types) $\mathbf{t} = \{t_1, \dots, t_{N_t}\}$ and skills of the operators needed to perform the given activity $\mathbf{s} = \{s_1, \dots, s_{N_s}\}$.

The relationships among these sets are defined by bipartite graphs $G_{i,j} = (O_i, O_j, E_{i,j})$ represented by $\mathbf{A}[O_i, O_j]$ biadjacency matrices, where O_i and O_j are used as a general representation of a sets of objects, as $O_i, O_j, \in \{\mathbf{p}, \mathbf{w}, \mathbf{c}, \mathbf{a}, \mathbf{t}, \mathbf{s}\}$.

The edges of these bipartite networks can represent material, energy or information flows, structural relationships, assignments, attributes, and preferences, and the edge weights

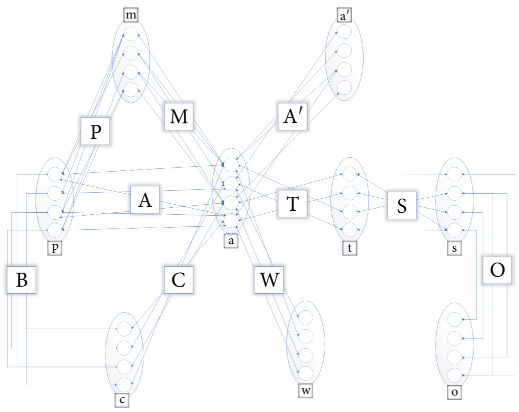


Figure 1: Illustrative network representation of a production system. The definitions of the symbols are given in Table 1.

can be proportional to the number of shared components/ resources or time/cost needed to produce a given product (see Table 1).

The proposed model can be considered as an *interacting or interconnected network* [24], where the family of bipartite networks defines crossed layers. Since different types of connections between the nodes can be defined, the model can also be handled as a *multidimensional network*. Both of these models are the special cases of multilayer networks, which representation is beneficial, since the layers represent the direct connections defined by the bipartite graphs, while the interlayer connections help in term of the visualization of the complex system by arranging the corresponding nodes at the same place within the layers (as it is illustrated in Figure 2).

The previously presented example serves only as an illustration. For real-life applications, the model should be extended and standardized. Manufacturing systems and their information can be organized by following the 5Ms and 5Cs concepts. The 5Ms stand for materials (properties and functions), machines (precision and capabilities), methods (efficiency and productivity), measurements (sensing and improvement), and modeling (prediction, optimization, and prevention). The 5Cs stand for connection (sensors and networks), cloud (data on demand and at anytime), content (correlation and purpose), community (sharing and social), and customization (personalization and value) [8]. Based on the characteristic elements and connections of production systems, the type of nodes and edges of their network [35] can be defined, and the relevant information is summarized in Tables 2–4. Although these concepts are already useful in structuring information, as a standardized solution, the applications of the ADACOR predicates that established relationships among the essential concepts of production management are recommended [36] (see Table 5).

Thanks to the recent standardization and integration of enterprise resource planning (ERP), manufacturing execution systems (MES), shop floor control (SFC), and product lifecycle management (PLM), it is straightforward to identify the connections of the standardized variables of production management and transform them into a multidimensional network model. The model is capable of representing information at different levels, so it can support factory flow analysis and departmental flow analysis, or, according to the concept of Industry 4.0, it can also integrate interorganizational supply chains. The development of organizational models is also supported, for this purpose, solutions following the standard of UN/EDIFACT (the United Nations rules for Electronic Data Interchange for Administration, Commerce and Transport) could be used.

The extracted models lend themselves to be handled in the databases of graphs [37, 38] or RDF-based ontologies [39]. In our work, the related technical details of building and storing graph-based decision systems are not the focus; rather, how information from this model can be extracted to support production flow analysis is of concern. In the next section, such techniques are presented.

3. Production Flow Analysis Relevant Operations on Networks

3.1. From Problems of Production Analysis to Tools of Network Science. The main benefit of the multidimensional network model is that it provides a transparent and easily interpretable integration of process- and product-relevant information and as well as facilitating the tools of network science for production flow analysis.

The aim of production flow analysis (PFA) is to identify bottlenecks and groups in products, components, and machines to highlight possible improvements by redesigning the layout, forming manufacturing cells, scheduling the activities, or identifying line families of products based on clustering the sequences of machine usage.

Modules/part families are sets of machines and parts that are highly likely to work together in one group or be processed in a similar order. Since this definition is similar to the concept of modules in networks, it is assumed that fining modules in (multidimensional) networks can be considered as a useful heuristical approach of PFA.

The application of heuristics in PFA is a well-accepted approach since in most cases, the economic benefits are complicated and time-consuming to calculate, and the resultant complex optimization problems are not easy to solve with classical optimization algorithms/operation research tools. In this paper, we suggest that the following network analysis tools should serve as a good heuristic solutions for specific PFA problems:

(1) Calculation of the loads and usage frequencies—identification of the bottlenecks

   (i) Calculation of unknown dependencies

   (ii) Analysis of node and edge centralities

(2) Group formation—clustering nodes and identifying communities

   (i) Rank-order-based clustering

Table 1: Definition of the biadjacency matrices of the bipartite networks used to illustrate how a production system can be represented by a multidimensional network.

Notation	Nodes	Description	Size
A	Product (**p**)-activity (**a**)	Activity required to produce a product	$N_p \times N_a$
W	Activity (**a**)-workstation/machine (**w**)	Workstation assigned for the activity	$N_a \times N_w$
A $'$	Activity (**a**)-activity (**a**$'$)	Precedence constraint between activities	$N_a \times N_a$
B	Product (**p**)-component/part (**c**)	Component/part required to produce a product	$N_p \times N_c$
P	Product (**p**)-module (**m**)	Module/part family required to produce a product	$N_p \times N_p$
C	Activity (**a**)-component (**c**)	Component/part built in or processed in an activity	$N_a \times N_c$
M	Activity (**a**)-module (**m**)	Activity required to produce a module	$N_a \times N_m$
T	Activity (**a**)-activity type (**t**)	Category of the activity	$N_a \times N_t$
S	Activity type (**t**)-skill (**s**)	Skill/education required for an activity category	$N_t \times N_s$
O	Skill (**s**)-operator (**o**)	Skills of the operators	$N_s \times N_o$

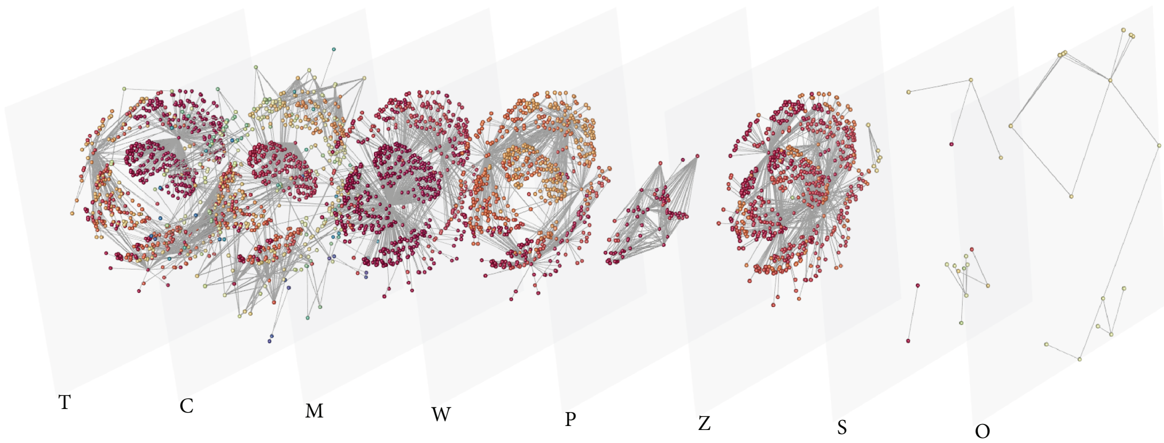


Figure 2: Visualization of the illustrative network as a multilayer/multiplex network highlights how the complex production system can be grouped into modules based on the “viewpoints” of the layers.

Table 2: The edge types of the proposed multilayer network.

	Flow type	Attribute type
Definition	Material, energy, or information flow between the nodes	Representation of the property of the node
Edge weight	Physical attributes of the flow, like quantity, or during discrete events, the frequency of the flow, like the number of hours between events	Similarity measure, meaning the quantity of equal attributes or the similarity of an attribute based on a scale
Self-loop	Inner activities	Not interpreted, as self-similarities are trivial
Parallel edges	Multiple flows can be represented by multilayer/multidimensional networks	Multiaspect similarities can be converted in to edge weights
Serial connections	Paths of the flow of different entities	Interpreted in terms of the time-varying case; shows spreading of a property
Modularity	Highly cooperative nodes	Highly similar nodes

(ii) Similarity-based clustering

(a) Calculation of node similarities of (projected) networks

(b) Clustering nodes and edges based on the calculated similarities

(c) Joining of clusters of different objects to form modules

(iii) Finding modules in the (multilayer) network

(3) Line formation—ordering modules to minimize sequential transfers

(i) Ordering based on the ratio of in/out degrees—Hollier’s method [40]

(ii) Application of graph layout techniques

Table 3: Node types of the proposed network.

	Event type	Resource type	Competency type
Fundamental properties	Occurrence probability, failure rate, cycle time, etc.	Physical properties, quality parameters (capacity, idle state, etc.)	Not generalizable, concept-dependent quantity and quality parameters
Node degree	Event frequency	Resource usage metric	Spreading competency
Modularity	Example: event sequence	Example: resources with the same usage parameters	Example: competencies possessed by the same resources/operators

Table 4: Node edge matchings in the proposed network.

	Flow type (edges)	Attribute type (edges)
Event type (nodes)	Process steps (nodes) and their input-output connections (edges)	Independent variables (nodes) and their settings (edges)
Resource type (nodes)	Information exchange (edges) between information systems (nodes)	Colleges working (nodes) on the same workstations (edges)
Competency type (nodes)	Commitment reporting between (edges) and jobs (nodes)	Same competency demanding (edges) jobs (nodes)

Table 5: The ADACOR predicates can be directly applied to define layers of the network [36] (please note that we use the term activity to refer to operations).

Predicates	Description
ComponentOf(x,y)	Product x is a component of product y
Allocated(x,y,t)	Operation x is allocated to resource y at time t
Available(x,y,t)	Resource x is available at time t for operation y
RequiresTool(x,y)	Execution of operation x requires tool y
HasTool(x,y,t)	Resource x has tool y available in its magazine at t
HasSkill(x,y)	Resource x has property (skill) y
HasFailure(x,y,t)	A disturbance x occurred in resource y at time t
Precedence(x,y)	Operation x requires previous execution of y
UsesRawMaterial(x,y)	Production order x uses raw material y
RequestSetup(x,y)	Operation x needs the execution of setup y
HasProcessPlan(x,y)	Production of x requires process plan y
OrderExecution(u,x,w,y)	Operation u is listed in process plan w (describing production of y) for production order x
HasRequirement(x,y)	Operation x requires property y
HasGripper(x,y,t)	Resource x has gripper y in its magazine at time t
ExecutesOperation(x,y)	Work order x includes operation y

3.2. Projections of the Multilayer Network and Calculation of Undefined Connections. As Figure 3 illustrates, when relationships among the O_i and O_j sets are not directly defined, it is possible to evaluate the relationship between its $o_{i,k}$ and $o_{j,l}$ elements as the number of possible paths or the length of the shortest path between these nodes.

In the case of connected unweighted multipartite graphs, the number of paths intersecting the O_0 set can be easily calculated based on the connected pairs of bipartite graphs as

$$\mathbf{A}_{O_0}\left[O_i, O_j\right] = \mathbf{A}\left[O_0, O_i\right]^{\mathrm{T}} \times \mathbf{A}\left[O_0, O_j\right]. \quad (1)$$

Conditional connections could also provide useful information in terms of PFA. To demonstrate the problem, let us have a look at Figure 4 which shows the network defined in (2). In this example, although operators o_1 and o_3 do not share any machines, the fact that machines m_1 and m_2 produce identical products results in the $\mathbf{A}[O_2|_{O_1}(O_0, O_0)]$ projection operators defining a connection between these operators.

$$\mathbf{A}\left[O_0, O_1\right] = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$
$$\mathbf{A}\left[O_0, O_2\right] = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}, \quad (2)$$
$$\mathbf{A}\left[O_2|_{O_1}(O_0, O_0)\right] = \begin{bmatrix} 2 & 4 & 2 & 0 & 0 \\ 4 & 9 & 5 & 0 & 0 \\ 2 & 5 & 4 & 2 & 1 \\ 0 & 0 & 2 & 4 & 2 \\ 0 & 0 & 1 & 2 & 1 \end{bmatrix}.$$

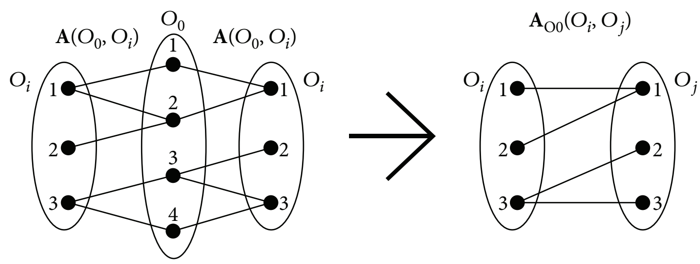


Figure 3: Projection of a property connection.

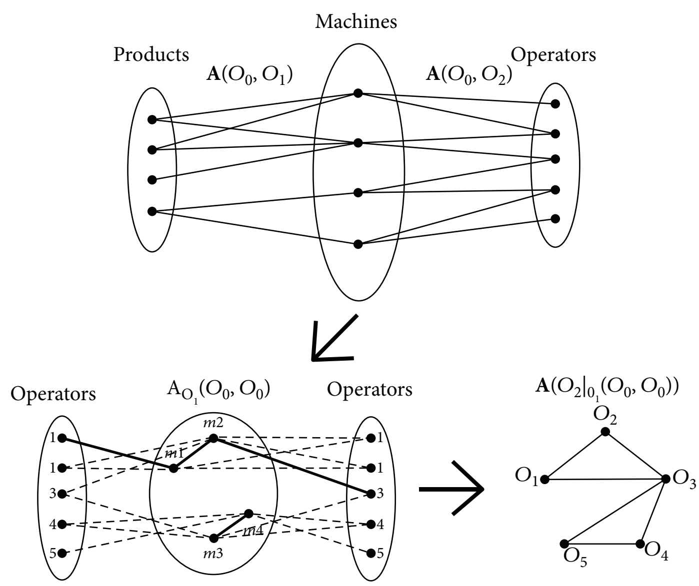


Figure 4: The advantage of complex conditional analysis using inner network.

Formally, in some cases, the $\mathbf{A}[O_i|_{O_k}(O_j, O_j)]$ conditional projections might be of interest defined by

$$\mathbf{A}\left[O_i|_{O_k}(O_j, O_j)\right] = \mathbf{A}\left[(O_j, O_i)\right]^{\mathrm{T}} \times \left(\mathbf{A}\left[(O_j, O_k)\right] \times \mathbf{A}\left[(O_j, O_k)\right]^{\mathrm{T}}\right) \times \mathbf{A}\left[(O_j, O_i)\right], \tag{3}$$

where the resultant $\mathbf{A}[O_i|_{O_k}(O_j, O_j)]$ network states that the ith property set is analyzed based on the $\mathbf{A}_{O_k}[(O_j, O_j)]$ inner network defined by the inner projection of the objects to the jth set.

The projections are not applicable for all types of edges (e.g., the projection with precedence constraints does not result in interpretable networks). Generally, the projections calculate the number of paths between the nodes which number is directly interpretable (e.g., it can reflect the number of assignable operators for a given workstation).

To support these calculations, it is beneficial to utilise the adjacency matrix of the whole multiplex network obtained by *flattening or matricization*:

$$\mathbf{A}_M = \begin{bmatrix} 0_1 & \mathbf{A}_{1,2} & \cdots & \mathbf{A}_{1,N} \\ \mathbf{A}_{2,1} & 0_1 & \cdots & \mathbf{A}_{2,N} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{A}_{N,1} & \mathbf{A}_{N,2} & \cdots & 0_N \end{bmatrix}, \tag{4}$$

where $\mathbf{A}_{i,j}$ is used to represent the $\mathbf{A}[O_i, O_j]$ biadjacency matrices of the $G_{i,j}$ bipartite graphs.

3.3. Calculation of Node Similarities. Node similarities can reveal useful information with regard to PFA, for example, if the similarities of the machines need to be defined based on how many common parts they are processing. When the machines are denoted as k and j, and S_k and S_j as the sets of parts that are connected to these machines, the similarities of the machines can be evaluated according to the Jaccard similarity index [41]:

$$\mathrm{sim}(k, j) = \frac{\left|S_k \cap S_j\right|}{\left|S_k\right| + \left|S_j\right| - \left|S_k \cap S_j\right|}. \tag{5}$$

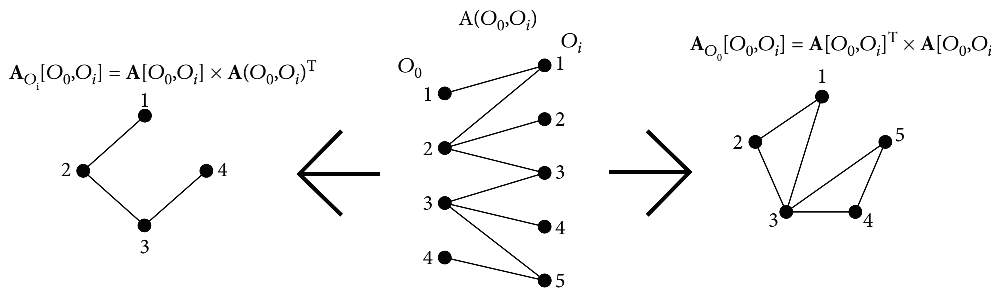


Figure 5: Two different projections can measure how the neighboring node set generates connections among the objects.

The proposed network-based representation is also beneficial in similarity analysis. When $O_0 = \mathbf{w}$ represents the set of machines/workstations and $O_i = \mathbf{c}$ represents the set of components, the $\mathbf{a}_{j,i} = 1$ edge weight stored at the intersection of the jth row and ith column of the $\mathbf{A}[O_0, O_i]$ biadjacency matrix represents that the ith type of component is built in at the jth workstation and the degree of the jth node, $k_j = \sum_i \mathbf{a}_{j,i}$ is identical to the cardinality of the $|S_j|$ set, which means how numbers of component types are built in at the jth workstation.

We can generate two projections for each bipartite network. The first projection connects two O_o nodes (in our case, two workstations) by a link if they are linked to the same O_i node (same components). As Figure 5 illustrates, the $|S_k \cap S_j|$ cardinality is identical to the $j-k$ edge weight of the projected network which represents how many identical components are built in at the kth and jth workstation:

$$\mathbf{A}_{O_0}[O_0, O_i] = \mathbf{A}[O_0, O_i]^{\mathrm{T}} \times \mathbf{A}[O_0, O_i]. \tag{6}$$

The second projection connects the O_i nodes (in our case, two components/parts) by a link if they connect to the same O_o node (workstations), which projection represents how parts are connected by the machines:

$$\mathbf{A}_{O_0}[O_0, O_i] = \mathbf{A}[O_0, O_i] \times \mathbf{A}[O_0, O_i]^{\mathrm{T}}. \tag{7}$$

When the similarities of more layers are taken into account, multiple projections on the same machines can be defined by the weighted sum of their projections:

$$\mathbf{A}[O_0, O_0] = \sum_i w_i \mathbf{A}[O_0, O_i] \times \mathbf{A}[O_0, O_i]^{\mathrm{T}}. \tag{8}$$

3.4. Identifying Modules for Group Formation. Communities are locally dense connected subgraphs in a network, so nodes that belong to a community have a higher probability to link to the other members of that community than to nodes that do not belong to the same community. Our key idea is that finding communities in (multilayer) networks of the proposed models can be used to solve group/cell formation problems of PFA. To formalize the cell formation problem, we utilized the modularity measure introduced by Newman [42] and improved for bipartite graphs by Barber [43].

A module of the network consists of a subgraph whose vertices are more likely to be connected to one another than to the vertices outside the subgraph. Modularity reflects the extent, relative to a random configuration network, to which edges are formed within modules instead of between modules. The modularity can be determined for each community of a network (in PFA, this means the modularity of each production cell can be calculated). For a network with n_c communities, the following modularity value is used to determine the modularity value of community Q_c in terms of each C_c community with N_c nodes connected by L_c links, $c = 1, \ldots, n_c$:

$$Q_c = \frac{1}{L} \sum_{(i,j)\in C_c} \left(\mathbf{a}_{i,j} - \frac{k_i k_j}{L} \right) = \frac{L_c}{L} - \frac{k_i k_j}{L^2}. \tag{9}$$

If the Q_c modularity value of a cluster is a positive value, then the subgraph C_c tends to be a community. The modularity of the full network can be evaluated by summing Q_c over all n_c communities, $Q = \sum_c Q_c$.

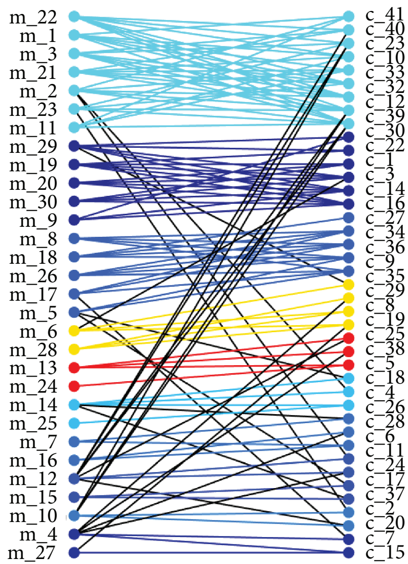
As can be seen, the definition of modularity perfectly fits the problem of manufacturing cell formation. Therefore, we propose a graph modularity maximization-based approach for this purpose. In this study, we adapt the Newman [42], LP-BRIM [44], and adaptive BRIM [43] algorithms available in the BiMAT MATLAB toolbox [45].

To illustrate the applicability of this approach, Figure 6 visualizes a cell formation problem and how the extracted modules can be assigned as manufacturing cells.

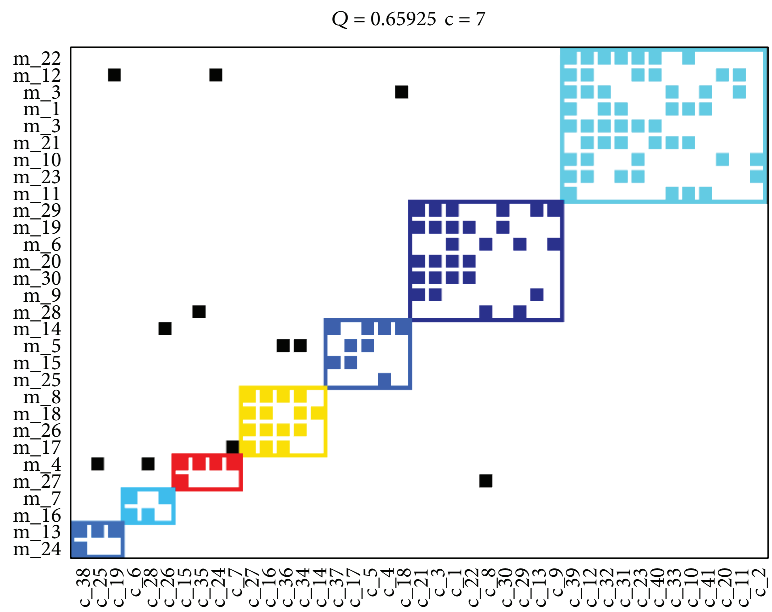
The efficiency of the formation of the cell can be evaluated based on e, the total number of activities; e_0, the number of exceptional elements that are excluded from the cells; and e_v, the number of zeros in the cells [46]:

$$\Gamma = \frac{e - e_0}{e + e_v}. \tag{10}$$

Table 6 compares the efficiencies of cell formation achieved by the proposed clustering and the modularity-based algorithms of cell formation with recently developed advanced goal-oriented optimization results in several benchmark problems of [46]. As can be seen, modularity-based algorithms perform surprisingly well, the Γ values (given as rounded parentages) are near to the optimized performances, and most importantly, the number of machine-part matchings outside of the modules (e_0 values) and the number of modules are much smaller in almost all cases than the optimized reference solutions.



(a) The rows and columns of the biadjacency matrix of the bipartite graph can be reordered to visualize the similarities of the modular graph layout



(b) After reordering/serialization of the biadjacency matrix, the modular structure of the problem is revealed

Figure 6: Modularity analysis of the 30×41 machine-part benchmark example.

Table 6: Cell formation efficiency of bipartite modularity optimization algorithms. The Γ values are given as rounded percentages.

Problem size	Optimization [46]			Newman			LP-BRIM			Adaptive BRIM		
	Number of c	Γ [%]	e_0	Number of c	Γ [%]	e_0	Number of c	Γ [%]	e_0	Number of c	Γ [%]	e_0
14×24	7	72	10	4	67	2	4	67	2	8	62	19
20×20	5	43	50	4	41	48	4	40	48	4	41	50
24×40	11	53	50	7	41	51	7	40	48	8	43	50
28×46	10	45	60	4	37	58	3	33	49	5	39	63
30×41	10	59	40	6	45	11	7	51	11	8	52	12
30×50	12	60	75	9	44	59	10	47	66	9	44	63
37×53	3	59	337	4	49	391	3	53	338	2	53	301

Based on this success, several modularity optimization algorithms were applied. As will be demonstrated in the following section, the approach is also applicable when searching for modules in multiple layers by the multilayer InfoMap algorithm [47, 48].

4. Application to the Analysis of Wire-Harness Production

To provide a detailed and reproducible case study for production flow analysis, an open-source benchmark model of modular wire-harness production was developed. The details of the model are given in the Appendix. The multilayer network model of the production flow analysis problem is formed and analyzed in the MuxViz framework developed for the interactive visualization and exploration of multilayer networks [49]. The established network is depicted in Figure 2.

4.1. Similarity and Modularity Analysis. Analysis of the reducibility of a multilayer network provides useful information about the similarities of the layers [50, 51]. To demonstrate the applicability of this metric, the C, Z, S, O, and T layers were analyzed (see Figure 7).

As can be seen in Figure 8, based on the reducibility of the network two clusters were formed. The first cluster is related to product-process (Z-T-C) layers, while the second collects the operator-skills- (O-S-) relevant information. The importance of the definition of the activity types (layer T) is also highlighted.

Although our network defines part families indirectly in layer M and also groups of these activities (in layer T), it is interesting to observe how the multilayer network is structured and how the analysis of the modularity of the network can form part and activity groups. For this purpose, a multilayer InfoMap algorithm was applied [47, 48].

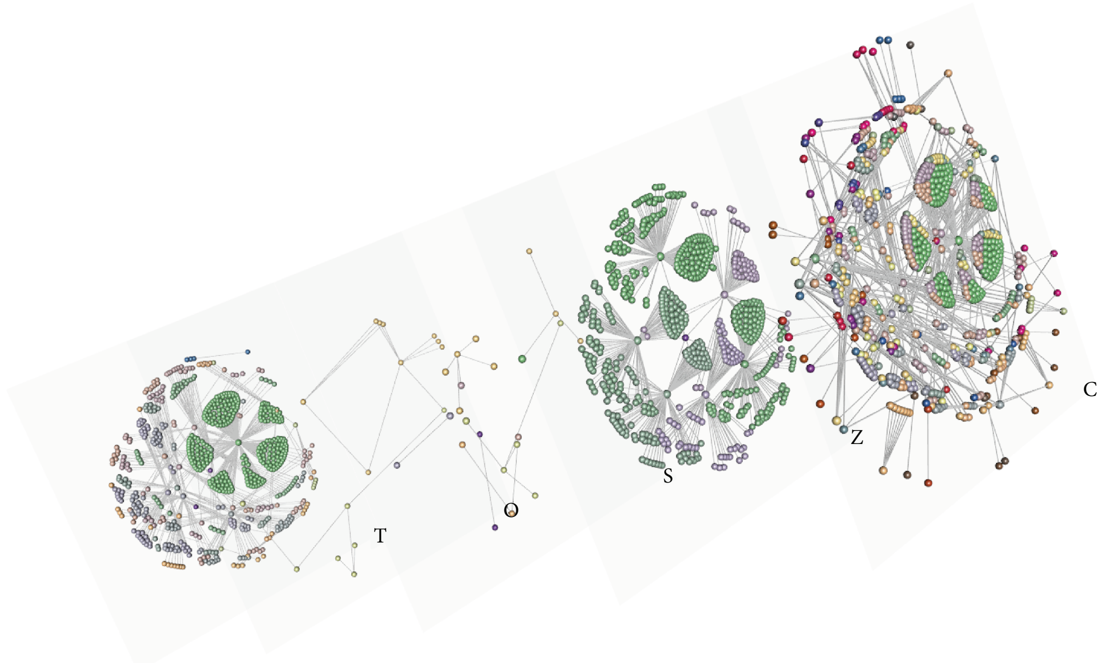


Figure 7: Multilayer network representing the details of the work of the operators (built-in components, C; zones of the activities, Z; skills, S; assignment of the operators to the workstations, O; and activity types, T) (see Table 1 for the detailed definition of the layers).

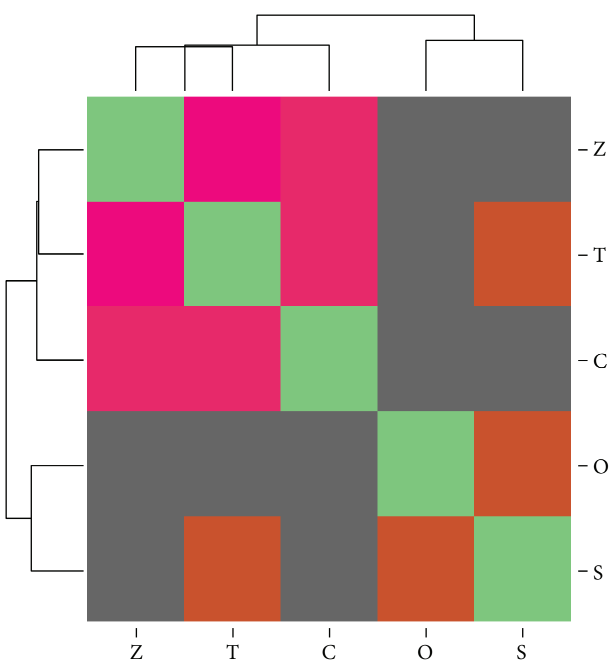


Figure 8: Analysis of the reducibility of the model provides useful information about the similarities of the layers. In our case, the two clusters related to product-process (Z-T-C) and operator-skills (O-S) were revealed. The importance of the definition of the activity types (layer T) is also highlighted.

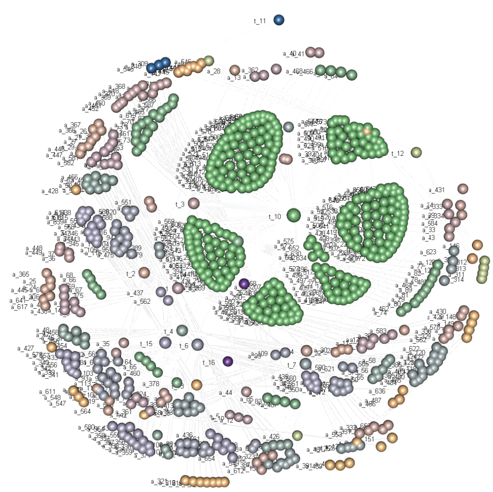
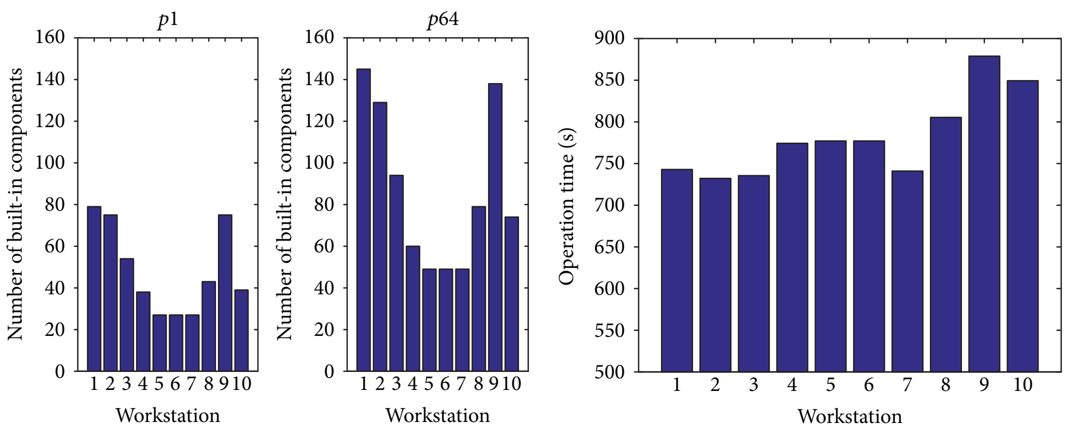


Figure 9: Layer T of the network defines the types of activities. The six clusters formed in this layer reflect the effects of how the activities are distributed among the zones (defined by layer Z), which illustrates the benefit of the multidimensional network-based visual exploration of the production data.

The analysis yielded useful and informative results. 26 modules were identified. Although layer M which represents how the activities are grouped according to different products, this analysis was able to detect the modules of the products $(m_1, \ldots, m_7)$ in terms of the types of the activities $(t_1, \ldots, t_{16})$. This result confirms that the analysis of the modularity of the proposed multilayer network model is useful in fine-tuning the existing part families based on multiple aspects representing the layers of the model.

To demonstrate how such information is useful in the early process-design phase to define technical modules, layer T of the C-Z-S-O-T multilayer network is shown in Figure 9. As can be seen, the most significant module is separated into



(a) Number of built-in components at a given workstation. The figure shows how the workload differs during the production of the base module (p_1) and the most complex product (p_{64})

(b) Total station times during the production of the 16th product

Figure 10: The workloads (number of activities, built-in components, and total activity times) can be easily calculated based on the biadjacency matrices of the proposed model, which supports the balancing of the conveyor belt.

six smaller groups by following the structure of layer Z that defines in which zone the activities occur. The central role of the most frequent and widely distributed t_{10} type of activity (wire-terminal attachment) is also highlighted.

4.2. Workload Analysis. The balancing of modular production is challenging due to the great diversity of products [52]. Besides group formation, the analysis of the workloads is also an important task in production flow analysis. The proposed bipartite network-based model can be directly applied for this purpose as the biadjacency matrices of the layers result in simple calculations. To illustrate this applicability, let us consider the analysis of how well the production line is balanced. The equation $\mathbf{L}_a = \mathbf{M}\mathbf{P}_p'$ represents the activities of the production of the pth product (where $\mathbf{P}_p$ represents the pth column of the $\mathbf{P}$ product-module matrix). As these activities are assigned to the workstations as $\mathbf{L}_w = \mathrm{diag}(\mathbf{L}_a)\mathbf{W}$ and $\mathbf{T}'\mathbf{L}_w$ represents the number of activities grouped by activity types and $\mathbf{T}'\mathbf{C}\mathbf{C}'\mathbf{L}_w$ is the number of built-in components at the workstations, the total activity time at the workstations can be calculated by the following equation, where θ_t represents the elementary activity times given in the appendix:

$$\mathbf{I}_{\text{time}} = \left[\mathbf{T}'\mathbf{L}_w, \mathbf{T}'\mathbf{C}\mathbf{C}'\mathbf{L}_w\right]\theta_t. \tag{11}$$

As Figure 10 illustrates, the calculations above can be used to check how the process is balanced and how the complexity of the product influences the workloads of the workstations.

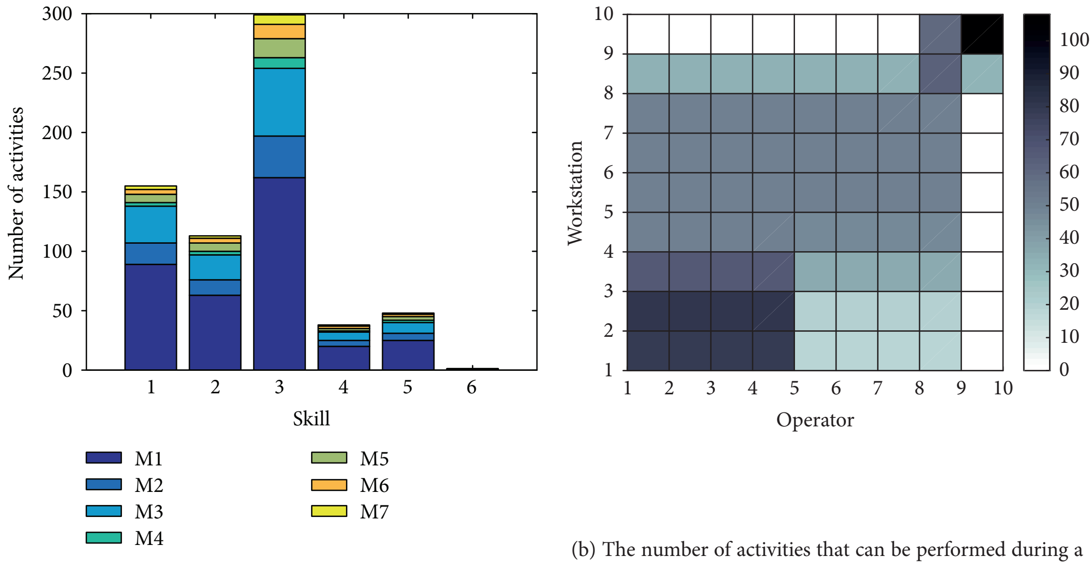
Although the presented workload analysis is not unique to the proposed model, we believe that the results demonstrated the rich information content and broad applicability of multilayer networks which can also be interpreted as a linear algebraic approach model of the system.

4.3. Analysis of the Flexibility of Operator Assignment. In the early 80s, [53] suggested that organisational research should incorporate network perspective. In the early 90s, six themes (turnover/absenteeism, power, work attitudes, job design, leadership, and motivation) dominated the research of microorganisational behaviour [54]. Recently, multilayer networks are becoming widely used in the analysis of social networks where people interact with each other in multiple ways like via mobile phone and emails [55–59]. In this paper, we make the first attempt to integrate such analysis to the modelling and optimisation of production process.

For successful line balancing of wire-harness production, the skills of the operators influencing the speed of the conveyor belt should also be studied [28] and handled [60]. Dynamic job rotation [29] requires efficient allocation of the assembly tasks while taking into account the constraints related to the available skills of the operators. Figure 11 shows the distribution of the required skills as a function of different product modules, $\mathbf{M}'\mathbf{T}\mathbf{S}$. As can be seen, the most in demand is the s_3 terminal-attaching skill, while s_6 is the visual testing skill which is required only once during production. The abilities of the operators can also be calculated, for example, $\mathbf{W}'\mathbf{T}\mathbf{S}\mathbf{O}'$ yields how many activities can be performed at a given operator-workstation assignment (see Figure 11(a)).

The presented analysis can be useful in designing the sessions of the operators by determining the components of critical skills and knowledge. Figure 12 shows the layers S and O of the network. Five groups of activity, skill, and operator nodes were identified with the help of multilayer modularity analysis. The smallest module contains the t_{15} clip installation activity type which requires specialist skills.

As can be seen, the skill s_4 can be considered a key piece of knowledge, because it is related to five types of activities. Operators o_9 and o_{10} possess specialist knowledge, while s_3 consists of group-wise knowledge because it is the most related to the operators.



(a) Distribution of the required skills as a function of the modules of the product

(b) The number of activities that can be performed during a given operator-workstation assignment. The skills of the operators constrain the flexibility of line balancing

FIGURE 11: Analysis of the demand of skills and the flexibility of the operator-workstation assignment.

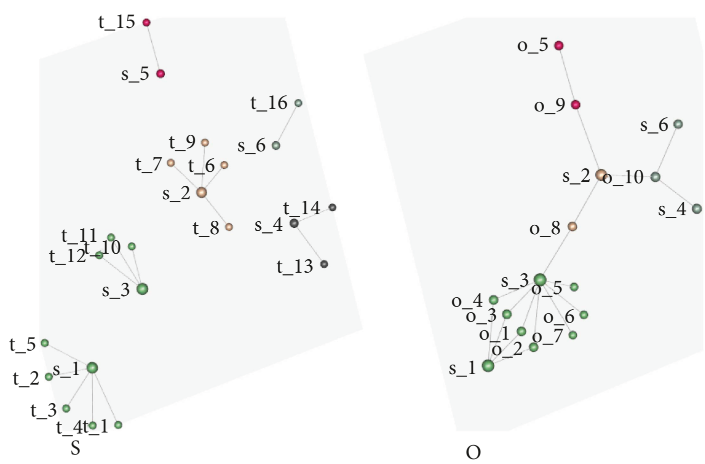


FIGURE 12: Skill (S) and operator (O) layers define the network that can be used to determine elements of critical knowledge which is useful in terms of the design of training programs for the operators.

The presented analysis demonstrated that the analysis of the node degrees can identify the critically essential skills and resources. Skills that have small degrees in the O layer can be considered as the knowledge of specialist, while skills with large degrees are quantified as group-wise knowledge. Skills that have no links at the S layer are useless, while skills that have a small degree at the O layer and high degree at the S layer are critical, as this reflects that a small number of operators can be assigned to a large number of tasks which requires this knowledge.

5. Conclusions

A multilayer network model was developed for production flow analysis to represent the physical and functional domains of production systems by taking into account the aspects of the structure of the system, the variety of machines, products, components, and operators and their interdependencies.

Most of the layers of the model are represented by a bipartite graph, where edges represent material, energy, or information flows and attributes of the objects represented by the nodes of the graph. It was highlighted that the nodes and connections could be easily defined based on standards of process management. As the layers of the network represent different aspects of the production system, the proposed model is flexible and easily extendable.

Following the introduction of the new modeling concept, it was demonstrated how the tools of network science should be used to support production flow analysis. Firstly, it was shown that the analysis of the paths in the network provides useful information about hidden, previously undefined connections. It was recognized that modularity analysis of the network is a promising tool for forming groups in PFA, and the performances of advanced (bipartite and multilayer) network modularity algorithms (like InfoMap) are comparable to the most advanced optimization algorithm tailored to the problem of cell formation.

A detailed benchmark problem was developed to make the research of multivariable algorithms of production flow analysis reproducible. With the help of the studied wire-harness process, the benefits of the modularity analysis of problem-specific sets of layers were demonstrated. The results confirm that the detected groups of activities are useful in terms of fine-tuning of modules (part families). Workload and capability-related network measures were developed. Along with analysis of the node degrees and their correlations, individual-, key-, and group-wise skills could be identified. The biadjacency matrices of the network lead to the calculation of workloads, and the investigation of how the production line is balanced. Besides the numerical analysis, visualizations were presented to demonstrate how multilayer networks provide insights into the critical factors of interconnected production systems, and the results of which confirm that multilayer networks can support the integration of production-relevant data and decision-making related to complex production systems.

Since the handling of the time-varying behaviour of process systems is becoming ever more critical in the field of cyber-physical systems, our future work will focus on the integration of historical process data to define networks of sequential procedures and temporal connections.

Appendix

Details of the Wire-Harness Production Technology

To support the reproducible development of production flow analysis and optimization algorithms, an open-source benchmark problem of a modular wire-harness production system was developed. The core of the system is a paced conveyor shown in Figure 13. Based on data published in [32, 61], N_p was based on 64 products and defined N_m as a combination of 7 modules: m_1 base module, m_2 as left- or right-hand drive, m_3 normal/hybrid, m_4 halogen/LED lights, m_5 petrol/diesel engine, m_6 4 doors/5 doors, and m_7 manual or automatic gearbox. N_a was defined 654 activities/tasks categorized into N_t which consisted of 16 activity types with well-modeled activity times (see Table 7). In these activities, N_c was equal to 64 different built-in part families (component types) (among these are $C_t = 180$ terminals, $C_b = 63$ bandages, $C_c = 25$ clips, and $C_w = 90$ wires). The conveyor N_w consisted of 10 workstations (tables). For every table (workstation), one operator is assigned, $N_o = 10$. The required N_s was also defined as 6 skills of the operators, namely, s_1—laying cable, s_2—spot-tying, s_3—terminal attaching, s_4—connector installing, s_5—clip installing, and s_6—visual testing. N_z was also defined as 6 zones for the workstations (see Figure 14) to study the distribution of

Figure 13: The wire-harness assembly pace conveyor [62]. The conveyor (often referred to as rotary) contains assembly tables consisting of connector and clip fixtures.

Table 7: Types of activities and the related activity times [61]. The activity times are calculated based on fixed and proportional values, for example, when an operator is laying four wires over one foot, according to the t_4 model, the activity time will be $1 \times 6.9\,\mathrm{s} + 4 \times 4.2 = 23.7\,\mathrm{s}$.

ID	Activity	Remark	Unit	Time (s)
t_1	Point-to-point wiring on chassis	Direct wiring	Number of wires	4.6
t_2	Laying in U-channel			4.4
t_3	Laying flat cable			7.7
t_4	Laying wire(s) onto harness jig	Laying flat cable	Base time	6.9
			Per wire	4.2
t_5	Laying cable connector (one end) onto harness jig	To the same breakout	Base time	7.4
			Per wire	2.3
t_6	Spot-tying onto cable and cutting it with a pair of scissors			16.6
t_7	Lacing activity		Base time	1.5
			Per additional stitch	3.6
t_8	Taping activity		Base time	1.8
			Per stitch	5.0
t_9	Inserting into tube or sleeve		Base time	3.0
			Per inch	2.4
t_{10}	Attachment of wire terminal	Terminal-block fastening (fork lug)		22.8
t_{11}	Screw fastening of terminal			17.1
t_{12}	Screw-and-nut fastening of terminal			24.7
t_{13}	Circular connector	Installation only		11.3
t_{14}	Rectangular connector	Latch or snap-on		24.0
t_{15}	Clip installation			8.0
t_{16}	Visual testing			120.0

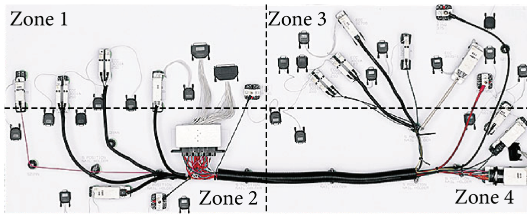


Figure 14: Zones were defined in the workstations to analyze the distribution of the fixtures and the related workload. The figure has been edited based on [63].

the fixtures on the tables. The related **Z** matrix is defined based on the layout of the table and shows the relationship between the activities and zones of the workstation, which facilitates a detailed analysis of the workload in the workstations. All of this information is represented by a set of bipartite graphs defined in Table 1 and depicted in Figure 2. The related dataset is freely and fully available on the website of the authors: https://www.abonyilab.com.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that no conflict of interest exists with regard to the publication of this paper.

Acknowledgments

This research was supported by the National Research, Development and Innovation Office (NKFIH), through Project OTKA-116674 (process mining and deep learning in the natural sciences and process development) and the EFOP-3.6.116-2016-00015 Smart Specialization Strategy (S3) Comprehensive Institutional Development Program.

References

[1] N. Papakostas, J. O'Connor, and G. Byrne, "Internet of Things technologies in manufacturing: application areas, challenges and outlook," in *2016 International Conference on Information Society (i-Society)*, pp. 126–131, Dublin, Ireland, 2016, IEEE.

[2] M. Rüßmann, M. Lorenz, P. Gerbert et al., *Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries*, Boston Consulting Group, 2015.

[3] S. Wang, J. Wan, D. Li, and C. Zhang, "Implementing smart factory of industrie 4.0: an outlook," *International Journal of Distributed Sensor Networks*, vol. 12, no. 1, Article ID 3159805, 2016.

[4] R. Anderl, "Industrie 4.0 - advanced engineering of smart products and smart production," in *19th International Seminar on High Technology, Technological Innovations in the Product Development*, pp. 1–14, Piracicaba, SP, Brazil, 2014.

[5] Z. Irani, A. M. Sharif, and P. E. D. Love, "Transforming failure into success through organisational learning: an analysis of a manufacturing information system," *European Journal of Information Systems*, vol. 10, no. 1, pp. 55–66, 2001.

[6] D. M. Lambert and M. C. Cooper, "Issues in supply chain management," *Industrial Marketing Management*, vol. 29, no. 1, pp. 65–83, 2000.

[7] V. Gurbaxani and S. Whang, "The impact of information systems on organizations and markets," *Communications of the ACM*, vol. 34, no. 1, pp. 59–73, 1991.

[8] J. Lee, E. Lapira, B. Bagheri, and H. A. Kao, "Recent advances and trends in predictive manufacturing systems in big data environment," *Manufacturing Letters*, vol. 1, no. 1, pp. 38–41, 2013.

[9] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *The International Journal of Advanced Manufacturing Technology*, vol. 94, no. 9–12, pp. 3563–3576, 2018.

[10] H. J. Shin, K. W. Cho, and C. H. Oh, "SVM-based dynamic reconfiguration CPS for manufacturing system in Industry 4.0," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 5795037, 13 pages, 2018.

[11] M. Karaköse and H. Yetis, "A cyberphysical system based mass-customization approach with integration of Industry 4.0 and smart city," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 1058081, 9 pages, 2017.

[12] A. P. Hameri, "Production flow analysis—cases from manufacturing and service industry," *International Journal of Production Economics*, vol. 129, no. 2, pp. 233–241, 2011.

[13] D. Mourtzis and M. Doukas, "Decentralized manufacturing systems review: challenges and outlook," *Logistics Research*, vol. 5, no. 3-4, pp. 113–121, 2012.

[14] J. Prince and J. M. Kay, "Combining lean and agile characteristics: creation of virtual groups by enhanced production flow analysis," *International Journal of Production Economics*, vol. 85, no. 3, pp. 305–318, 2003.

[15] G. Chryssolouris, K. Efthymiou, N. Papakostas, D. Mourtzis, and A. Pagoropoulos, "Flexibility and complexity: is it a trade-off?," *International Journal of Production Research*, vol. 51, no. 23-24, pp. 6788–6802, 2013.

[16] K. Efthymiou, D. Mourtzis, A. Pagoropoulos, N. Papakostas, and G. Chryssolouris, "Manufacturing systems complexity analysis methods review," *International Journal of Computer Integrated Manufacturing*, vol. 29, no. 9, pp. 1025–1044, 2016.

[17] J. L. Burbidge, "Production flow analysis for planning group technology," *Journal of Operations Management*, vol. 10, no. 1, pp. 5–27, 1991.

[18] J. L. Burbidge, "Production flow analysis," *Production Engineering*, vol. 50, no. 4-5, pp. 139–152, 1971.

[19] J. Burbidge, "A manual method of production flow analysis," *Production Engineering*, vol. 56, no. 10, pp. 34–38, 1977.

[20] A. R. Rahani and M. Al-Ashraf, "Production flow analysis through value stream mapping: a lean manufacturing process case study," *Procedia Engineering*, vol. 41, pp. 1727–1734, 2012.

[21] A. F. Cutting-Decelle, L. Pouchard, B. Das, R. Young, and J. Michel, "Utilising standards based approaches to information sharing and interoperability in manufacturing decision support," in *International Conference FAIM04*, pp. 1–12, Toronto, ON, Canada, 2004.

[22] H. K. Lin and J. A. Harding, "A manufacturing system engineering ontology model on the semantic web for inter-enterprise collaboration," *Computers in Industry*, vol. 58, no. 5, pp. 428–437, 2007.

[23] A. L. Barabási, *Linked: The New Science of Networks*, AAPT, 2003.

[24] S. Boccaletti, G. Bianconi, R. Criado et al., "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014.

[25] M. Chattopadhyay, P. K. Dan, and S. Mazumdar, "Application of visual clustering properties of self organizing map in machine–part cell formation," *Applied Soft Computing*, vol. 12, no. 2, pp. 600–610, 2012.

[26] O. Torkul, I. H. Cedimoglu, and A. Geyik, "An application of fuzzy clustering to manufacturing cell design," *Journal of Intelligent & Fuzzy Systems*, vol. 17, no. 2, pp. 173–181, 2006.

[27] S. Li and H. Mehrabadi, "Generation of block diagonal forms using hierarchical clustering for cell formation problems," *Procedia CIRP*, vol. 17, pp. 44–49, 2014.

[28] E. Aguirre and B. Raucent, "Performances of wire harness assembly systems," in *Proceedings of 1994 IEEE International Symposium on Industrial Electronics (ISIE'94)*, pp. 292–297, Santiago, Chile, 1994, IEEE.

[29] G. Michalos, S. Makris, L. Rentzos, and G. Chryssolouris, "Dynamic job rotation for workload balancing in human based assembly systems," *CIRP Journal of Manufacturing Science and Technology*, vol. 2, no. 3, pp. 153–160, 2010.

[30] H. Park, S. H. Lee, and M. R. Cutkosky, "Computational support for concurrent engineering of cable harnesses," *Computer Engineering*, vol. 1, 1992.

[31] A. B. Conru, "A genetic approach to the cable harness routing problem," in *Proceedings of the First IEEE Conference on Evolutionary Computation. IEEE World Congress on Computational Intelligence*, pp. 200–205, Orlando, FL, USA, 1994.

[32] N. S. Ong, "Activity-based cost tables to support wire harness design," *International Journal of Production Economics*, vol. 29, no. 3, pp. 271–289, 1993.

[33] B. Agard and M. Tollenaere, "Design of wire harnesses for mass customization," in *Recent Advances in Integrated Design and Manufacturing in Mechanical Engineering*, pp. 53–62, Springer, 2003.

[34] I. G. K. Ei-Essawy and J. Torrance, "Component flow analysis — an effective approach to production systems' design," *Production Engineering*, vol. 51, no. 5, pp. 165–170, 1972.

[35] T. Csiszér, *Use of Network Approach in the Improvement of Operation Efficiency on the Field of Quality Development, [Ph.D. Thesis]*, University of West Hungary, Cziráki József Doctoral School, 2014.

[36] S. Borgo and P. Leitão, "Foundations for a core ontology of manufacturing," in *Ontologies*, pp. 751–775, Springer, 2007.

[37] I. Robinson, J. Webber, and E. Eifrem, *Graph Databases*, O'Reilly Media, Inc., 2013.

[38] I. Robinson, J. Webber, and E. Eifrem, *Graph Databases: New Opportunities for Connected Data*, O'Reilly Media, Inc., 2015.

[39] S. Bechhofer, "OWL: Web ontology language," in *Encyclopedia of Database Systems*, pp. 2008-2009, Springer, 2009.

[40] A. Aherwar, "An implementation of group technology using 'C' on a problem of machine sequencing," in *Recent Trends in Information Technology (RTIT)*, Gwalior, India, February 26-27, 2010.

[41] Y. Li, P. Luo, and C. Wu, "A new network node similarity measure method and its applications," 2014, https://arxiv.org/abs/1403.4303.

[42] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[43] M. J. Barber, "Modularity and community detection in bipartite networks," *Physical Review E*, vol. 76, no. 6, article 066102, 2007.

[44] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Physical Review E*, vol. 76, no. 3, article 036106, 2007.

[45] C. O. Flores, T. Poisot, S. Valverde, and J. S. Weitz, "BiMat: a MATLAB package to facilitate the analysis of bipartite networks," *Methods in Ecology and Evolution*, vol. 7, no. 1, pp. 127–132, 2016.

[46] T. H. Wu, C. C. Chang, and J. Y. Yeh, "A hybrid heuristic algorithm adopting both Boltzmann function and mutation operator for manufacturing cell formation problems," *International Journal of Production Economics*, vol. 120, no. 2, pp. 669–688, 2009.

[47] M. Rosvall and C. Bergstrom, "Maps of information flow reveal community structure in complex networks," 2007, https://arxiv.org/abs/0707.0609v1.

[48] M. Rosvall, D. Axelsson, and C. T. Bergstrom, "The map equation," *The European Physical Journal Special Topics*, vol. 178, no. 1, pp. 13–23, 2009.

[49] M. De Domenico, M. A. Porter, and A. Arenas, "MuxViz: a tool for multilayer analysis and visualization of networks," *Journal of Complex Networks*, vol. 3, no. 2, pp. 159–176, 2015.

[50] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Layer aggregation and reducibility of multilayer interconnected networks," 2014, https://arxiv.org/abs/1405.0425.

[51] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3–5, pp. 75–174, 2010.

[52] F. K. Pil and M. Holweg, "Linking product variety to order-fulfillment strategies," *Interfaces*, vol. 34, no. 5, pp. 394–403, 2004.

[53] N. M. Tichy, M. L. Tushman, and C. Fombrun, "Social network analysis for organizations," *The Academy of Management Review*, vol. 4, no. 4, pp. 507–519, 1979.

[54] D. Krackhardt and D. J. Brass, *Intraorganizational Networks: The Micro Side*, Sage Publications, Inc., 1994.

[55] M. Magnani and L. Rossi, "The ML-model for multi-layer social networks," in *2011 International Conference on Advances in Social Networks Analysis and Mining*, pp. 5–12, Kaohsiung, Taiwan, 2011.

[56] A. Socievole, F. De Rango, and A. Caputo, "Wireless contacts, Facebook friendships and interests: analysis of a multi-layer social network in an academic environment," in *2014 IFIP Wireless Days (WD)*, pp. 1–7, Rio de Janeiro, RJ, Brazil, 2014, IEEE.

[57] L. Weiyi, C. Lingli, and H. Guangmin, "Mining essential relationships under multiplex networks," 2015, https://arxiv.org/abs/1511.09134.

[58] M. E. Dickison, M. Magnani, and L. Rossi, *Multilayer Social Networks*, Cambridge University Press, 2016.

[59] Y. Murase, J. Török, H. H. Jo, K. Kaski, and J. Kertész, "Multilayer weighted social network model," *Physical Review E*, vol. 90, no. 5, 2014.

[60] G. Lodewijks, “Two decades dynamics of belt conveyor systems,” *Bulk Solids Handling*, vol. 22, no. 2, pp. 124–132, 2002.

[61] N. S. Ong and G. Boothroyd, “Assembly times for electrical connections and wire harnesses,” *The International Journal of Advanced Manufacturing Technology*, vol. 6, no. 2, pp. 155–179, 1991.

[62] “Assembly line conveyor systems,” 2015, https://www.pacline.com/photos/photos-by-solution/assembly-line-conveyors/.

[63] “Panduit introduces new technology for harness assembly board layout,” 2015, https://www.assemblymag.com/articles/93034-panduit-introduces-new-technology-for-harness-assembly-board-layout.

Hindawi
Complexity
Volume 2018, Article ID 4632195, 13 pages
https://doi.org/10.1155/2018/4632195

Research Article

On the Design Complexity of Cyberphysical Production Systems

Luis Ribeiro and Martin Hochwallner

Department of Management and Engineering, Division of Manufacturing Engineering, Linköping University, 58183 Linköping, Sweden

Correspondence should be addressed to Luis Ribeiro; luis.ribeiro@liu.se

Received 23 February 2018; Accepted 9 May 2018; Published 10 June 2018

Academic Editor: Dimitris Mourtzis

Establishing mass-customization practices, in a sustainable way, at a time of increased market uncertainty, is a pressing challenge for modern producing companies and one that traditional automation solutions cannot cope with. Industry 4.0 seeks to mitigate current practice's limitations. It promotes a vision of a fully interconnected ecosystem of systems, machines, products, and many different stakeholders. In this environment, dynamically interconnected autonomous systems support humans in multifaceted decision-making. Industrial Internet of Things and cyberphysical systems (CPSs) are just two of the emerging concepts that embody the design and behavioral principles of these highly complex technical systems. The research within multiagent systems in manufacturing, by embodying most of the defining principles of industrial CPSs (ICPSs), is often regarded as a precursor for many of today's emerging ICPS architectures. However, the domain has been fuzzy in specifying clear-cut design objectives and rules. Designs have been proposed with different positioning, creating confusion in concepts and supporting technologies. This paper contributes by providing clear definitions and interpretations of the main functional traits spread across the literature. A characterization of the defining functional requirements of ICPSs follows, in the form of a scale, rating systems according to the degree of implementation of the different functions.

1. Introduction

The need and pursuit for highly adaptable systems are not an exclusive endeavour of Industry 4.0 and more generally of what is understood as the 4th Industrial Revolution. The motivation can be at least traced back to the aftermath of the 1970s oil shock when mainly European and American companies strongly invested in technology and particularly in the so-called flexible manufacturing systems (FMS) [1], at the same time Japanese companies were developing lean manufacturing principles. Even if the early incarnations of FMS where relatively unsuccessful, the need went on as mass customization became the excellence paradigm in production [2].

Mass customization works generally in economies of scale with relative stable markets. However, at a time when unpredictable market demands are accompanied by almost continuous and fast-paced innovation processes and production sustainability is the new excellence paradigm, the existing automation practices that had for some years supported mass customization processes are starting to subside.

For more than two decades now, researchers have turned to distributed computation practices and artificial intelligence in the quest for new metaphors for developing, designing, and implementing more adaptive production systems [3, 4]. The idea of reconfigurable manufacturing systems (RMS) [5], as a manufacturing system that is designed at the outset for rapid change in structure, as well as in hardware and software components, in order to quickly adjust production capacity and functionality within a part family in sudden changes in market or in regulatory requirements, emerged almost in parallel. Along also came the concepts of holonic manufacturing systems (HMS) [6] and bionic manufacturing systems [7].

Collectively, these paradigms share quite a few important design principles and ideology. Among them is the notion that systems should be modular. Modularity is interpreted therein in both physical and logical terms. Standardized interfaces between the system's modules are of paramount importance to ensure functional composability and scalability. The modules themselves should denote a higher degree of autonomy, and their functionalities are generally self-contained, a principle that in some branches of robotics became known as embodiment [8] and that has very close ties to the current cyberphysical conceptualization.

The quick expansion of information technologies (IT), especially those related to computer networks, in the late 90s to early 2000s has paved the way for the Industry 4.0 (I4.0) as web service-related technologies and service-oriented architecture started to penetrate and dominate industrial IT infrastructures [9, 10]. Even if they did not fundamentally change the ongoing enterprise processes, and most notably left nearly untouched the field level control layers of production facilities, they have opened up for a more interoperable use of information.

From there, several relatively competing paradigms have developed, namely, industrial Internet of Things (IIoT), cloud automation/manufacturing (CA), cyberphysical production systems (CPPS), and industrial cyberphysical systems (ICPS). It is difficult to pinpoint exactly which properties pertain exclusively to which. It is also unclear if some are manifestations of others. The so-called 4th Industrial Revolution, or Industry 4.0, seems to be the overarching umbrella for a lot of scientific and technical activities that are generally anticipated to lead to the next generation of automation systems. So much so that this anticipation has motivated several research efforts globally [11]: the H2020 Framework Programme for Research and Innovation under the Factories of the Future Public-Private Partnership (EU); the Industrial Internet Consortium, created by AT&T, Cisco Systems, General Electric, IBM, and Intel in 2014 (US); the Made in China 2025 initiative (CN); and globally many other initiatives.

The authors position this paper within the CPPS/ICPS thematic, detailed in greater extent in the subsequent section, which they perceive as the involving paradigm for most of the activities occurring within the factory of the future. As with all the other paradigms, there is resounding fuzziness in concepts and technologies and their relative roles in the creation of next-generation automation systems. There are various ideas regarding the functions and reference architectures of such infrastructures, and these are spread and presented in different contexts.

In the previous context, the present paper develops along several contributing lines. Drawing from reference literature in the fields of multiagent systems (MAS) applied to manufacturing, CPPS/ICPS, and to a lesser extent from cloud automation and more general literature in CPSs, the paper synthetizes and functionally formulates the zero-tier requirements that have been attributed to CPPSs in the context of I4.0.

The requirement formulation is accompanied by a corresponding structural model positioned at the same abstraction level.

Both the requirements and the structural model are then discussed in greater detail and are the base for developing a scale that enables the classification of CPPSs. Such scale ranges from a system not fulfilling the minimum requirements to be convertible to a minimal CPPS implementation up to how a full featured CPPS would fulfil the synthetized requirements.

The authors trust that such a scale can be used as an indicator of preparedness for tackling the challenges that are believed to be workable for a system implemented under the premises of I4.0. Simultaneously, the proposed scale can be used as a tool to inform subsequent system design and implementation directions towards target levels or functional goals.

In the remainder of this paper, the methodology considered during the research presented herein is detailed in the next section, followed by the main results and discussion. The paper finishes with the main concluding thoughts and pertinent research questions worth of a follow-up.

1.1. Materials and Methods. This work's methodological approach is based on traditional requirements engineering (RE), sometimes also called requirement management (RM). RE/RM is a systematic and disciplined approach to the specification and management of requirements [12] and in part a discipline of systems engineering [12, 13].

A requirement is therefore understood as [12]

(i) a condition or capability needed by a user to solve a problem or achieve an objective;

(ii) a condition or capability that must be met or possessed by a system or system component to satisfy a contract, standard, specification, or other formally imposed documents;

(iii) a documented representation of a condition or capability as in (i) or (ii).

Requirements can be elicited from many different sources [13]: documents, systems in operation, and stakeholders. Stakeholders are here understood as a person or organization that has a direct or indirect influence on a system's requirement [12].

The stakeholders are here addressed by their roles representing relevant concerns. General stakeholders for CPPSs are the law, the product user, the plant customer including the sales channel, the procurement channel for parts/material (material inflow), the equipment operator/worker, the production engineer (process, quality), the plant operator, the plant owner, the equipment maintenance/installation, and the equipment manufacturer.

Another source of requirements is documents. In the context of this paper, relevant documents are laws, standards, and publications. This work contains a literature review and consolidation based on publications in the area. The selected literature was obtained from a systematic analysis of publication arising mainly from members of several technical communities, namely, the following IEEE technical committees: industrial agents and industrial cyberphysical systems. The selection was guided by a tool developed to analyse vast networks of coauthorships partially described in [14]. The authors assume that to an important extent, the surveyed publications are also reflecting the interests of the many different stakeholders. The authors also accept that this vision is debatable; however, given the novelty of the domain and the degree of speculation around it, previously peer-validated information entails an important level of consensus around the views expressed that otherwise would be harder to attain.

At the same time and on the very same line, this contribution is intended to serve as source of requirements upon which other contributions can be developed which will eventually consider more matured views from other stockholders.

One particular important differentiating factor from conventional RE/RM lies in the fact that the present paper does not concentrate on one specific system but rather on an entire class of systems. Scope is normally defined as [12] the range of things that can be shaped and designed when developing a system. As such, it should be evident that the present contribution has a metalevel positioning due to its wider scope.

Further on this, potential CPPSs inherit the requirements for existing systems in operation, and any particular design derived from it will have to comply with, at least, similar requirements applicable to its predecessor.

This contribution assumes these previous requirements but focuses on the novel aspects that are perceived to pertain particularly to the CPPS domain. When there was an alignment of requirements but the CPPS formulation supposed an important reinterpretation, then the requirement was included in the discussion.

The starting point for the subsequent discussion is the adapted CPPS structural models presented in [15, 16]. Of particular interest is the definition of CPPS (Figure 1) understood here as an aggregation of human resources, cyberphysical production modules (CPPM) (defined next), subsystems, and aggregated products towards which it establishes one or several cyberphysically formulated interaction interfaces. These interfaces are used for monitoring and control of the CPPS operations as well as to tap into the knowledge generated both by the human resources, the CPPMs, and the subsystems during the production process as well as knowledge generated by its aggregated products throughout their life cycle. This internal knowledge is used in different time scales to continuously improve operations and to inform the strategic consumptions of capital, raw materials and energy in their many forms. Cyberphysically formulated products will also generate value for external systems, as part of networks of things and services, towards which they maintain interaction interfaces. The outcome of such interactions is external knowledge. Access to it may be offered as a value-adding service that can potentially help to further improve CPPS operations.

The subsequent CPPM definition adheres to the notion of module as detailed in [17] which states that a module is tightly coupled within and loosely connected to the rest of the system.

The CPPM definition (Figure 2) is adapted from and envisions a CPPM as [16] a module consisting of three logically aggregated entities: an equipment, a controller or computing platform, and a cyberrepresentation of this whole. The computing platform may be shared between several cyberrepresentations if it provides access to the equipment that they represent. The cyberrepresentation contains both the interface and the algorithms that enable the module to interact with other modules, human resources, or subsystems without the need for reprogramming it and implements a hardware abstraction layer that decouples the interaction and execution logic from the equipment details.

The authors will consider these starting structural formulations throughout the paper to support the discussion of the design complexity of CPPS across the many different stepwise implementations.

2. Results and Discussion

2.1. Engineering Requirements from the Specialized Literature. The specialized literature offers many important insights into the characteristics that belong to a CPPS. However, these many different characteristics have been defined at equally many abstraction levels. In general and as previously discussed in [16, 18, 19], scientific contributions have therefore ranged from hardware in the loop implementations up to reference architectures. This means that the expectations that are posed upon system realization also vary substantially. There is also normally a disconnection between reference architectures, where desirable characteristics are embedded in a conceptual design, and their subsequent realization that often does not adhere to the conceptual principles.

The distance between conceptual design and concrete implementations is also aggravated by the fact that many different characteristics are normally specified without a degree of obligation other than a mention that the characteristic is desired.

In the following analysis, the authors, sustained by a survey of selected literature, proceed with formally specifying requirements accordingly to [13] and also, but to a much lesser degree, following a top-down decomposition approach proper of axiomatic design processes [20].

The practical implications of this approach is that all the synthetized CPPS characteristics are expressed in the form object + degree of obligation + function and that each identified characteristic is at least expanded by one level in depth. They are also formulated, to best effort degree, according to independence axiom that postulates that an optimal design always maintains the independence of functional requirements [20]. The objects of interest are drawn from the CPPS/CPPM definitions presented before. The degree of obligation is an enumeration with three possible values with the following semantics [13]:

(i) Shall: the object must implement the function attributed to it.

(ii) Should: the object will, if possible, implement the function attributed to it.

(iii) Will: the object may come to implement the function attributed to it at some point in the future.

The selected literature reflects an extremely wide set of "desirable" characteristics of a CPPS. A considerable number of these are using different nomenclatures for similar

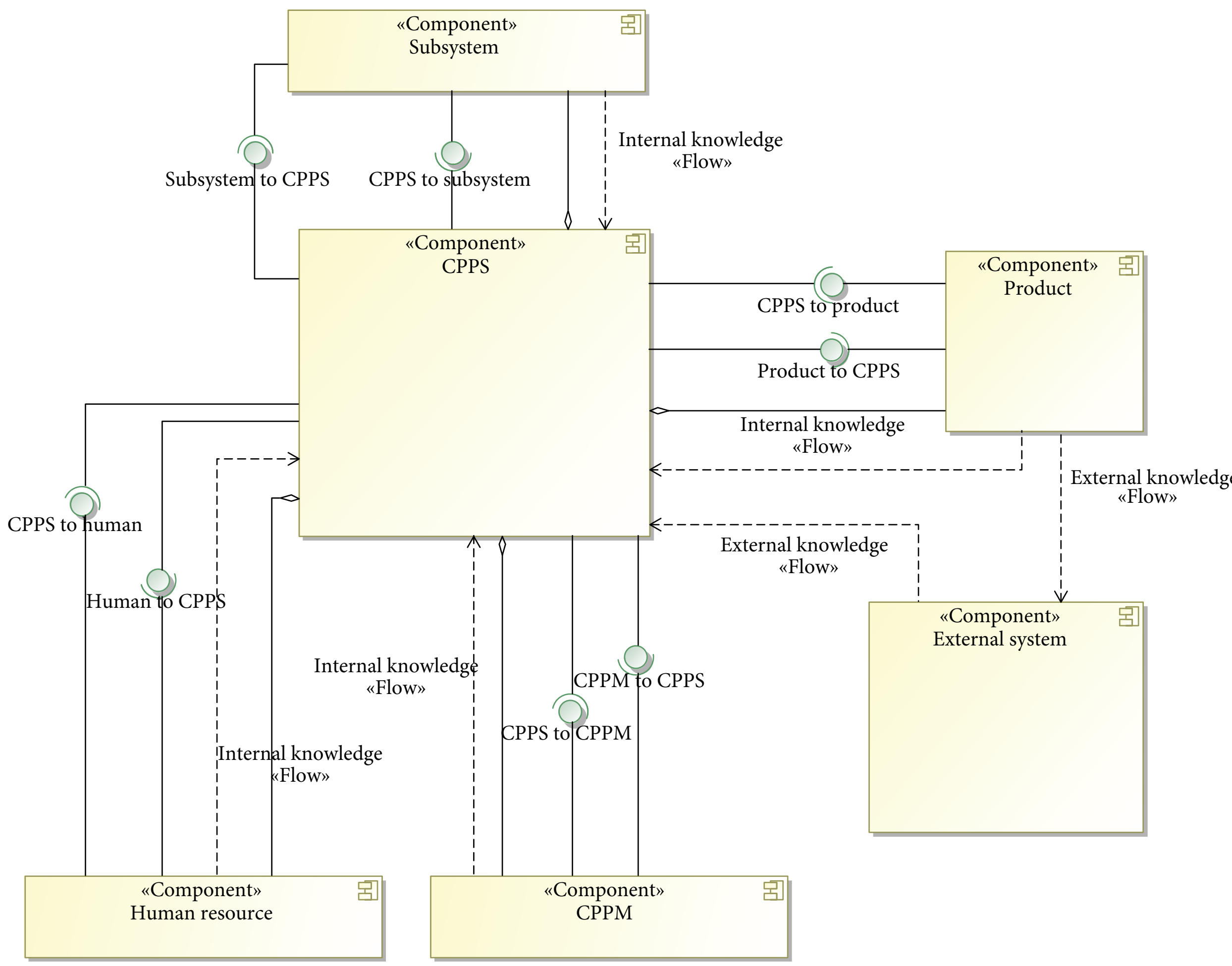


Figure 1: CPPS structural model (adapted from [15]).

functions. In the scope of this work, the authors have clustered sufficiently close attributes into the following top-level requirements: adaptability, convertibility, integrability, predictability, usability, diagnosability, safety, and security. Their top-level definition (level 1) provides typically a (re)formulation that is closer to the normal interpretation of similar concepts but is already functionally specified. The subrequirements (level 1.X) provide further details that conduce to and inform about the top level. A few requirements are very close or even taken from other authors. These are properly identified. Some of the nomenclatures were also taken from previous works on reconfigurable manufacturing systems, in particular [5, 21, 22], but unless otherwise stated has been reinterpreted and formulated in a functional way.

From the top-level requirements, the authors would like to discuss first the three that they believe are the main differentiating requirements within a CPPS: adaptability, convertibility, and integrability (Figure 3). At the core of almost all research agendas in I4.0 lies the idea of constantly adapting systems. Adaptation can happen in structure, function, or both. As such, adaptation can only be implemented if the system components can be integrated with each other (integrability). It additionally requires a relative modular physical structure (convertibility) to support a wide scope of adaptive solutions beyond simple functional adaptation.

The extent to which different adaptability, convertibility, and integrability functions are implemented is a direct measurement of the degree of how cyberphysical, at the light of I4.0, a production system really is.

The tables below (Tables 1–3) reflect the results of the literature survey and the reformulation of many of the spread and sparsely defined desired characteristics for these three main requirements. The related literature sources include only the ones that have more strongly stated or otherwise contributed to the proposed RE/RM exercise. Similar concepts and definitions will of course be found in other papers in the specialized literature but they are either contained by the cited sources or otherwise had generally weaker requirements.

The adaptability requirements (Table 1) focus mainly on the expectations on system behavior. It has been accepted for many years now that the ability to adapt to changing conditions is of paramount importance for production systems. The notion of CPPS seems to encompass also the possibility of structural adaptation whereby mobile equipment can even change, in a more or less autonomous way, the factory layout. The authors set a strong emphasis in this ability to adapt.

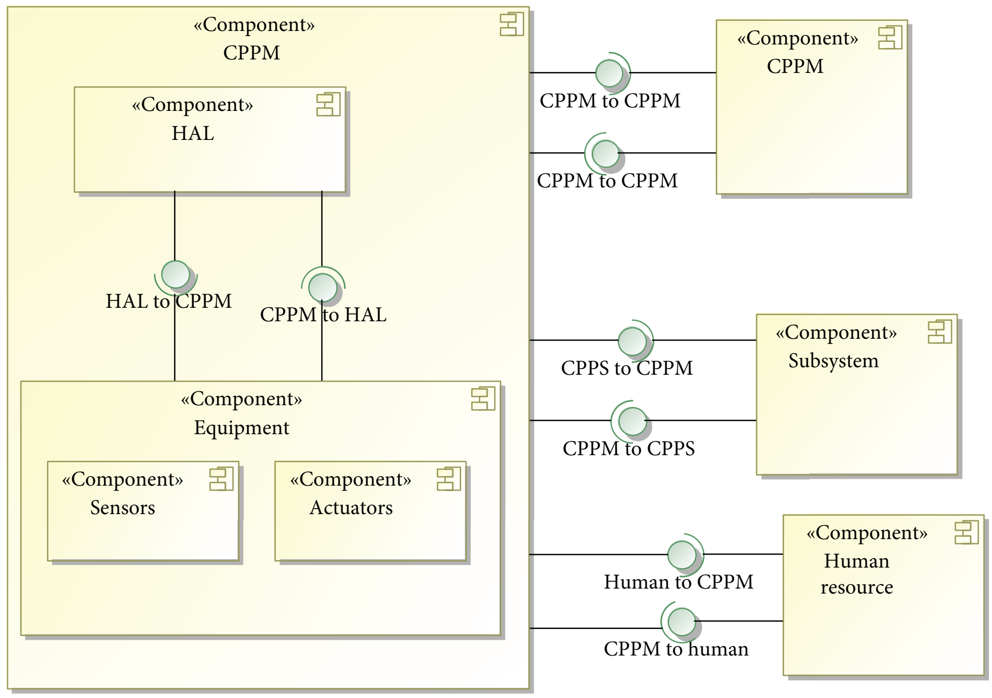


Figure 2: CPPM structural model (adapted from [16]).

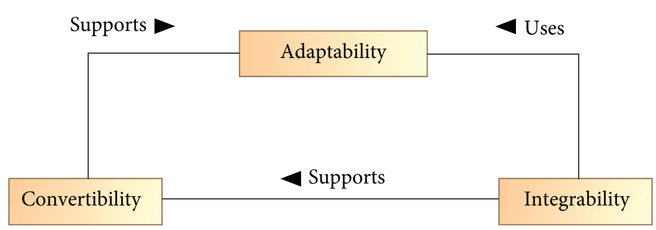


Figure 3: Associations between the core requirements in CPPSs.

Table 1: Synthesis and reformulation of requirements on adaptability.

Requirement/ subrequirement	(Re)formulation	Related literature sources
Adaptability 1	The CPPS shall adapt its behavior and structure to changing conditions.	[24–33]
Adaptability 1.1	The CPPS shall have the perception of environmental elements and events with respect to time or space, the comprehension of their meaning, and the projection of their status after some variables have changed, such as time, or some other variables, such as a predetermined event [34].	[24, 34]
Adaptability 1.2	The CPPS should learn its future behavior from past experiences.	[24]
Adaptability 1.3	The CPPS shall adapt in the most optimal way possible.	[35]
Adaptability 1.4	The CPPMs and subsystems should collaborate with other CPPMs, subsystems, and human resources in the same system during operation.	[24, 36, 37]
Adaptability 1.5	The CPPS shall autonomously attempt to overcome faults or otherwise support users in this process.	[26, 29, 38]

Conditions that would cause a system to adapt could vary in their emergency and implementation time frame and naturally include faults and failures, surges, or drops in production as well as the ramp up of new products. The literature suggests a high degree of obligation in the implementation of such functions. This reason has justified the use of the obligation operator “shall” in most requirements in Table 1. The two exceptions would be adaptability 1.2 where learning is mentioned and adaptability 1.4 where collaboration is stated.

The justification for the lower degree of obligation in adaptability 1.2 relates to the fact that this is something that the specialized literature mentions quite frequently; however, learning implies the generation and availability of a substantial amount of data when applied at a system level. On

Table 2: Synthesis and reformulation of requirements on convertibility.

Requirement/ subrequirement	(Re)formulation	Related literature sources
Convertibility 1	The CPPS shall have a modular physical structure that enables the rapid adjustment of its production functionalities.	[21, 22]
Convertibility 1.1	The CPPM shall provide reusable physical interfaces that enable it to participate in distinct production contexts.	[21, 22, 33, 35, 37]
Convertibility 1.2	The CPPS should provide interfaces that enable the physical composition of its CPPMs.	[6]

Table 3: Synthesis and reformulation of requirements on integrability.

Requirement/ subrequirement	(Re)formulation	Related literature sources
Integrability 1	The CPPS shall provide mechanisms to integrate new heterogeneous components and systems that are unknown to it.	[24, 36, 37]
Integrability 1.1	The CPPS will connect to other CPSs and extract information from them that is used to support its operations.	[24, 26, 29, 36]
Integrability 1.2	The CPPM and subsystems shall provide reusable logical interfaces that enable the integration of new equipment.	[6, 24, 29, 31, 38]
Integrability 1.3	The CPPS shall use as the basis of its operation a domain specific language that in a uniform way enables CPPMs, subsystems, and human resources to describe their capabilities, operational requirements, and status to the CPPS.	

systems designed on the onset for frequent change (both physical and logic), acquiring significant data may prove to be a challenge. In this context, adaptability 1.2 is in principle important but may not be implementable in the general case.

Collaboration (adaptability 1.4) is frequently cited as the preferred mechanism to support adaptation. However, many other forms of interaction [23] are possible, and collaboration may not be desirable at all times. In some cases, competitive interaction strategies may, for example, attain better results. This is normally context dependent.

Table 2 shows the convertibility requirements. If adaptation is reflected on how the system behaves, then convertibility is about the physical characteristics of the system that ultimately allow it to make use of its adaptive behavior. Modularity is greatly recognized as the prevailing characteristic, and a CPPS should ensure that its components can be combined in different ways to adapt and generate new functions when required. It therefore requires from its CPPMs a minimal level of mechanical interfacing and compatibility.

Convertibility alone is only providing guarantees of mechanical conformity. Surfaces, cabling, power lines, and so on should match. Integrability covers therefore the logical interoperability of the system. Cyberphysical system must be put together along these two dimensions. The degree of obligation is very high and supposes the existence of logical abstractions that enable a CPPS to interpret and use the functions of equipment that were previously unknown to it (integrability 1). These include the usage of machine interpretable languages and compatible logical interfaces (integrability 1.2 and integrability 1.3).

In the future, if the vision of I4.0 fully consolidates and becomes a reality, the scope of action of CPPSs will extend beyond just a cluster of CPPSs and may come to encompass other CPSs that relate to it (integrability 1.1).

As mentioned before, this contribution assumes that the existing requirements applicable to today's production system will also to a certain extent apply to CPPSs, and the authors do not inspect or address them in their full range. However, from these, there are a few that may require a fundamentally different approach or otherwise are challenges, in particular predictability, usability, diagnosability, safety, and security.

Predictability (Table 4) has been traditionally interpreted in two ways: the behavior of the system is transparent and explainable and the system behaves in a time-predictable way. The authors believe that the first should be inherent to any adaptive process in the industry and therefore interpret predictability more towards the second direction (time predictably as in predictability). The obligation degree here is maximum, and the general contract is that the CPPS delivers a performance that is compatible with the processes under its control (predictability 1.1). This does not necessarily mean hard real-time performance but means a behavioral delivery that enables planning and asserts production quality (predictability 1.2 and predictability 1.3).

Predictability is generally related to usability (Table 5). No matter how adaptable the system is, without a proper set of control tools that enable the straightforward operation of the system (usability 1), CPPSs cannot be accepted. The novelty in CPPS tools lies in that they must exploit and support the adaptive capabilities of the system and hence enable the design of products along the entire range of possibilities covered by the convertibility and integrability possibilities of a given setup (usability 1.1).

Table 4: Synthesis and reformulation of requirements on predictability.

Requirement/ subrequirement	(Re)formulation	Related literature sources
Predictability 1	The CPPS shall operate in a time-predictable way.	
Predictability 1.1	The CPPS shall operate at a time frame that is compatible to that of the physical processes it seeks to control.	[24, 36, 37]
Predictability 1.2	The CPPS shall operate in a way that enables short-, medium-, and long-term planning of production activities.	
Predictability 1.3	The CPPS shall produce with predictable quality.	[38]

Table 5: Synthesis and reformulation of requirements on usability.

Requirement/ subrequirement	(Re)formulation	Related literature sources
Usability 1	The CPPS shall provide tools for efficient development of diverse production processes.	[29, 35]
Usability 1.1	The CPPS shall provide tools that enable the production of the full number of variants that could be enacted considering the available resources and their potential configurations.	[21, 22]
Usability 1.2	The CPPS shall provide information that is tailored to the needs of specific users' roles.	[24, 36, 39]

In the scope of supporting operations, CPPS must also ensure that operators are being informed at the right time about the relevant events taking place in the system. But the amount of events in a CPPS is, due to its adaptive behavior, overwhelming. As such, the system must have the ability to filter and direct only the relevant information required to support the different decision-making processes (usability 1.2).

The ability to understand what a complex system is doing has ramifications not only to its normal operation but also, at least as importantly if not more, to its behavior under disturbances. CPPS shall attempt to detect and diagnose faults and failures within. Given their complex and compositional structure, this may be a challenge depending on the size and the intricacy of the underlying interactions. The obligation here is to do a best effort in diagnosability (Table 6) and while operating in a fully traceable way. In a CPPS context, traceability means to a great extent make a set of complex causality flows understandable.

Safety (Table 7) is one of the emerging concerns within the CPPS-related literature. It is an absolute obligation that has so many different dimensions at the moment that make it very difficult to characterize. Today's safety regulations apply mainly to fixed installations and conventional automation solutions. The regulations are scarcer when considering autonomous machines and even more when considering mobile autonomous machines that may interact with humans. Most human-machine interactions are, in an industrial context, assessed on an individual case.

The success of CPPS and I4.0 in general requires that safety regulations are extended and come to include, in the general case, situations where autonomous systems, performing structural adjustments in a production floor, can be accepted and validated.

Security (Table 8) is a more tangible endeavour in CPPSs. The obligation is to execute in a way that protects the CPPS from harmful influence from its components or external systems. Most of the readily available supporting technologies are generally poor when it comes to fulfilling security requirements. While this problem has already been proven a challenge in running systems, it is aggravated in the context of CPPSs due to their general low maturity.

2.2. Grading CPPSs according to Their Degree of Implementation. Different system designs will adhere to the requirements specified before to a different extent. Some systems may not adhere at all. This creates a range of possibilities that collectively inform about the ability of evolving a specific system or system design in order to tackle the emerging challenges envisioned by I4.0.

The authors therefore propose a scale that distinguishes between the different stages at which a system or system design may be. The proposed scale ranges from −1 to 5, where −1 would apply to a system that, for different reasons, cannot be transformed to reach a CPPS level of functional delivery and a 5 would apply to a full-featured CPPS. Such scales are relatively common in the software development domain, for example, the Capability Maturity Model Integration (CMMI) rates software development organizations and guides their development process in respect to the way that software development process is conducted.

While the authors think that such a fully developed model would be extremely complex to develop in the scope of CPPSs due to the level of maturity of the entire field, they believe that similar ideas can be usefully developed considering the requirements earlier detailed. The discussion of the different levels is therefore presented next. A conceptual exemplary system is discussed along each level to better illustrate evolve-ability challenges and potential.

Level 0—at this level, substantial engineering effort would be required to evolve the system to CPPS-grade functionalities.

(i) Adaptability: the production system does not support adaptation. There are practical ways to revamp the system to become adaptable but substantial engineering effort is necessary.

Table 6: Synthesis and reformulation of requirements on diagnosability.

Requirement/ subrequirement	(Re)formulation	Related literature sources
Diagnosability 1	The CPPS shall whenever possible autonomously detect and diagnose the root causes of product defects [21] or otherwise actively support users in their identification.	[21, 22]
Diagnosability 1.1	The CPPS shall operate in a traceable way.	[35]

Table 7: Synthesis and reformulation of requirements on safety.

Requirement/subrequirement	(Re)formulation	Related literature sources
Safety 1	The CPPS shall operate in a safe way.	[24, 40]

Table 8: Synthesis and reformulation of requirements on security.

Requirement/ subrequirement	(Re)formulation	Related literature sources
Security 1	The CPPS shall operate in a secure way.	[24, 36, 38, 40]
Security 1.1	The CPPS shall not allow harmful interference from other CPSs it interacts with.	
Security 1.2	The CPPS shall provide secure access to its functions and modules.	

(ii) Convertibility: the equipment is generally not modular and/or follows essentially a mechanical monolithic structure. Parts within the equipment can be replaced, maintained, or serviced but replacing does not lead to change in function.

(iii) Integrability: the equipment is not connected through standardized interfaces in a consistent way and many different interfaces and interaction protocols are in place. Integration is handled in a case by case basis.

(iv) Other requirements: the system fulfils the minimum level of requirements for operating according to the production objectives and applicable regulations.

A semiautomated workshop with several workstations and/or machine tools is a typical scenario for a level 0 system. In such a scenario, equipment is generally not integrated with each other. Production is moved from the workstation/machine to workstation/machine by a manual process or otherwise by a fixed and indexed automated transportation solution. The same production processes are always used at the same workstations/machines or otherwise reconfigured by an operator that manually introduces or selects new programs already available from the workstations/machines. More advanced machines may also automatically act upon the presence of workpieces if they are electronically tagged in a way that enables a specific machine to execute a program for that specific workpiece. The immediate steps that should be taken to improve such a system, bringing it to the next level, include improving its integrability. As legacy equipment is frequently found in such environments, it is generally not straightforwardly possible to use more advanced communication protocols, for example, web service-based protocols. Adapter devices will have to be considered in these cases that create a harmonizing layer over legacy components and act as command interpreters and translators. These adapter devices are also usually a prerequisite for improved adaptability. Convertibility issues will need to be addressed over time by replacing the physical equipment with more modular equipment where applicable.

Level −1—at this level, it is not practically possible to evolve the system.

(i) Level −1 is defined after level zero since its starting point is the same; however, there are either legal restrictions (e.g., regulatory or IP protection) that block any further system integration effort. Equipment or systems exist in the form of self-contained isolated islands.

Level 1—at this level, the system is in principle modular and some of these modularity considerations have been incorporated at design time to cater for potential long-term changes.

(i) Adaptability: the production system is programmed to cover all the foreseeable cases where decisions may need to be taken and interactions between components are static and permanent.

(ii) Convertibility: the equipment is generally modular and it is possible to replace equipment with some with similar capabilities. These changes usually require a nonnegligible system stoppage at least in the area where the new equipment is being integrated.

(iii) Integrability: the equipment is in principle connected using standard interfaces and/or communication protocols and it is controllable through standard automation languages.

(iv) Other requirements: the system fulfils the minimum level of requirements for operating according to the production objectives and applicable regulations.

A system at level 1 generally denotes a relatively high level of automation whereby processes and machines are generally connected and influence each other. Communication occurs over standard industrial bus systems, and standard data formats are considered making it possible to replace similar components. Generally, more than one bus communication system is supported which improves the opportunities for integrating new equipment. However, there are different system islands that, although integrated, use information flow and formats that are generally specifically defined in the context of that island and cannot be automatically interpreted or used elsewhere. The main challenge in evolving such system lies in harmonizing the information semantics and ensuring there is a minimal level of equipment semantic interfacing that allows global commands to be issued. These may come to include selectively (de)activating the equipment, selecting specific programs within controllers or machines, influencing and controlling workpiece planning, and scheduling and routing strategies.

Level 2—similar to level 1.

(i) Adaptability: same as the previous level. However, the system is prepared so that similar modules may be activated or deactivated as part of its operations.

(ii) Convertibility: the equipment is generally modular and was designed in a way that with a minimal stoppage can be integrated into the production system.

(iii) Integrability: the equipment is in principle connected using standard interfaces; additional effort has been put in ensuring that modules with the same function and same interfaces can be integrated with a minimal or no reprogramming effort.

(iv) Other requirements: the system fulfils the minimum level of requirements for operating according to the production objectives and applicable regulations.

A level 2 system in a fairly tuned level 1 system where great care has been put at design phase to enable the quick replacement of equipment, a typical scenario is a production line where different products are produced during different shifts and the line requires the addition and/or removal of equipment, or tools, to support the different production processes (i.e., there is a system preparation process during shift changeovers). The system still features static logical links to the different pieces of equipment which is the main improvement point to reach a level 3 system. Here, the strategy is to improve the logical interfaces rendering them interpretable and discoverable to enable dynamic orchestration of components. The adoption of emerging standards such as OPC-UA and other web service-based formats is a proper step in this direction. Such interfaces will enable a substantial reduction of the preparation time as more processes may be easily designed and implemented. At the same time, the integration of new machines becomes easier.

Level 3—at this level, the system's structure and behavior are similar to level 2; however, equipment can be discovered dynamically and functions and processes can be activated and designed by orchestration of the different components.

(i) Adaptability: the production system is programmed to cover all the main cases where decisions may need to be taken; however, new cases can be easily introduced by orchestrating the system's components. The system is reactive to the introduction of new components and can use them as long as there is an orchestration routine including them.

(ii) Convertibility: the equipment is modular and was designed in a way that with a minimal stoppage can be integrated into the production system.

(iii) Integrability: the equipment is in principle connected using standard interfaces described in machine-interpretable formats that support the recognition and operation as previously described.

(iv) Other requirements: the system fulfils the minimum level of requirements for operating according to the production objectives and applicable regulations. Its more dynamic and complex structure may also enable the usage of advanced data collection and processing practices for diagnosability.

A level 3 system is what would be attained if a factory would be created at the present time with the best possible available technology. It assumes the seamless integration and harmonization of equipment functions. This means that within the convertibility limits of the system, many different production processes may be enacted in a practical way. This requires the work of experts for designing these processes and system orchestration routines. This includes both the design for normal operations and recovery actions. These routines are of varying complexity depending on the system's and products' nature and size but once developed, the system is able to follow and react to changes on them seamlessly. Level 3 is particularly aligned with the objectives of the RAMI 4.0 model [41]. However, at this level, system autonomy is still constrained. The system is still executing what it is being told to do. Multilevel autonomous decision-making mechanism is not by default included in the design which focuses mainly in integrability aspects.

Level 4—at this level, the system has the ability to interpret its structure and from it infer and make available possible orchestration routines. The system self-orchestrates within its autonomy boundaries and calls upon human resources to intervene for taking actions beyond its autonomy scope.

(i) Adaptability: the system has a high degree of autonomy which it uses to dynamically allocate different resources during the production process. Newly introduced equipment is immediately recognized and brought into operation. The system can plan in function of the available resources and the current production targets.

(ii) Convertibility: the equipment is modular and was designed in a way it can be integrated without any stoppage.

(iii) Integrability: the equipment is connected using standard interfaces described in machine-interpretable formats that support the recognition and operation as previously described.

(iv) Other requirements: the system fulfils the minimum level of requirements for operating according to the production objectives and applicable regulations. Its more dynamic and complex structure enables the usage of advanced data collection and processing practices for diagnosability. In particular, the system has the ability of understanding and inferring from different production contexts which resources should be allocated to carry out the required tasks.

Level 4 systems have been demonstrated, with limitations, in scientific setups and prototypes. A fairly extensive revision can be found in [42]. They draw upon the interface discoverability and interpretability opportunities to take autonomous decisions concerning the workflow of workpieces under dynamic conditions. They require a very good cyberphysical alignment in respect to form and function that usually is not available in today's systems. The generation of self-orchestration routines requires the generalization of functions. Among the many demonstrated systems and proposed architectures, it is usual to consider high-level functions such as transformation, transportation, and coordination. They generally require a one-to-one mapping between the cybercontrolling entity and its physical counterpart with a high degree of embodiment that is generally complex to attain and partly motivates this paper. A complete discussion on these design issues can be found in [16] and the references therein. The steps to level five are a multidisciplinary open research question that will include a substantial increase of the systems' cyberphysical autonomy.

Level 5—this level reflects the ideal CPPS configuration denoting very high cyberphysical autonomy that enables it to even dynamically change its structure.

(i) Adaptability: the system has the ability to autonomously self-adjust including its physical structure. These changes may come to include acquiring missing functionality of components from a digital marketplace. The system plans and adjusts at different time frames. The scope of self-adjustment includes catering not only for production disturbances but also for faults and failure recovery.

(ii) Convertibility: components have a high mechanical interoperability which otherwise can be adjusted by the system itself. Components are usually mobile or transferable by other components.

(iii) Integrability: the system has the ability to recognize any new component or system that is added to it and can generate new functionality by combining the functions provided by the newly added component or system with the ones existing.

(iv) Other requirements: all to their full extent.

2.3. Integrated Discussion. Even if most expectations concerning what a CPPS should be are somehow dispersed in the literature and detailed in different contexts, there is a set of prevailing requirements that clearly suggest the dominating characteristics. These requirements are, themselves, prone to multiple interpretations. This makes it difficult to draw the line where a system goes from a conventionally automated system to a full-featured CPPS. The distinction is harder since most of today's systems already exhibit characteristics that rightfully belong to CPPS.

Current research directions set an important focus on interfacing and system integration aspects. These are very important since, as suggested by the structural models in Figures 1 and 2, they are the pillars of convertibility and integrability. However, a smaller focus has been set on the advanced behavioral aspects.

Previous research in MAS has partially covered some ground in this direction. The functional requirements uncovered as part of the RE/RM exercise in this contribution are reflecting just that. The grading scale previously presented shows that, on the more advanced levels, the key functions are of behavioral nature and although interfacing aspects are fundamental, they do not cater for the adaptability requirements alone.

Most of today's production systems can be positioned between levels −1 to 2 in the scale. Level three is approaching the research domain and reflects the best of what could nowadays be achieved with advanced technologies, for example, OPC-UA, coupled to mainstream AI. This possibility arises from the fact that, generally, web service stacks are relatively easy to integrate with current automation controllers. Such integration will typically not change the controller's native programming languages, in a practice that is well aligned with the traditional automation pyramid [43] also known as the ISA-95 model.

The ISA-95 model has however been recently deemed less adequate for the sound development for CPPSs, and new models such as the RAMI 4.0 (DIN SPEC 91345 [41]) have emerged. The RAMI 4.0 model provides a better coverage on the nature of the communication between the system's components and introduces the idea of an administrative shell acting as a single point of information retrieval. The administrative shell provides a broad coverage for collecting information related to some of the main requirements early specified but generally excludes the behavioral modelling of the components.

The authors believe that level 4 represents the current research front and it is already making important assumptions regarding the nature of the physical part of the system. The redesigning effort for transitioning from levels 3 to 4 is still substantial, and the maturity of the current supporting technologies is generally low which currently undermines their adoption.

Finally, level 5 takes a look into the future's full-featured CPPSs. Here, the degree of system autonomy and integration is very high and the system complexity is maximal. While level 4 would be attainable with a considerable engineering effort in the direction of applied research, level 5 is still beyond reach without a more multidisciplinary and more fundamental research effort.

There is no doubt, therefore, that the systems of the future will entail a much higher complexity along the dimensions discussed in this paper. Physical complexity arising from improving the system's integrability and convertibility should be quantifiable at large (number of parts, material properties, interface types, etc.).

However, the quantification of behavioral complexity is a more elusive task. Today's production systems are already extremely complex entities combining physical and logical aspects. The fact that very few systems can be brought to an operational status seamlessly is a clear sign that there is already a certain degree of unknown behavior generating unexpected effects. This inherent uncertainty is important to understand the general complexity baseline of production systems. To better substantiate the discussion around increasing degrees of behavioral complexity and associated uncertainty, it is helpful to introduce the idea of "weak emergence" defined as [44] macrostate P of a system S with microdynamic D is weakly emergent if P can only be derived from D and S's external conditions but only by simulation.

At the light of the previous definition and assuming that today's systems could be set up seamlessly without uncertainty, for a relatively small system and a reduced number of fixed component interactions, it could be feasible to fully formally verify the system behavior. A larger system would require the creation of black boxes, individually verified, and a subsequent verification of the whole based on a selection of inputs and outputs in and from these black boxes. In the current state of developments, this would be believable up until level 3 in the scale discussed before. Until level 3, the causality matrix of the system is extremely complex but to a certain extent still possibly manageable with some simplifications. At level 4, autonomous behavior is introduced. This dramatically complicates the causality matrix because it removes operational constraints from the system decision-making processes. Level 5 is the theoretical limit case.

The richer the adaptive process, the bigger the effort in creating traceable system information and the stronger the requirements on the noncore requirements of CPPSs, namely, diagnosability, predictability, safety, and security. In this context, while the fulfilment of the core requirements is what will ensure the anticipated sophistication in production activities, it is the fulfilment of that most likely will dictate their success. This makes them fundamental research topics that need to be understood less at the light of levels −1 to 2 but more at levels 3–5. Seamless operation of CPPS will only be attained by managing and understanding design complexity as a whole.

3. Conclusions

This paper proposes an RE/RM exercise that, drawing from reference literature in supporting concepts for what is now understood as I4.0, provides a synthetized interpretation of the main requirements pending upon CPPS development. Unlike most of the literature definitions, the set of requirements is functionally formulated to better pinpoint the behavioral expectations related to CPPSs. However, even this is prone to multiple interpretations. To partially overcome this challenge, this contribution proposes a scale for classifying CPPSs according to the degree to which they adhere to the different requirements. The paper further discusses the main implications of the different levels and positions of the existing developments.

The authors hope that such contribution will shed a better light on the science of engineering CPPSs by removing partially the conceptual fuzziness and by providing a sound basis for positioning existing systems and understanding the cost of transitioning between different levels.

The discussion is as generalized as possible and targets CPPSs and not their specific instances. The authors believe that this metalevel positioning is very important within the scope of I4.0 since it captures its system of systems nature in a tangible way. It also highlights that the main design challenges require a holistic perspective that promotes and respect the cyberphysical nature of the system components.

Data Availability

The supporting data was collected from the literature cited within the article.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This research was financed by Linköpings Universitet.

References

[1] J. A. B. d. Oliveira, *Coalition Based Approach for Shop Floor Agility – a Multiagent Approach*, FCT-UNL, 2003.

[2] B. J. Pine, *Mass Customization: the New Frontier in Business Competition*, Harvard Business Press, 1999.

[3] W. Shen, L. Wang, and Q. Hao, "Agent-based distributed manufacturing process planning and scheduling: a state-of-the-art survey," *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, vol. 36, no. 4, pp. 563–577, 2006.

[4] W. Shen and D. H. Norrie, "Agent-based systems for intelligent manufacturing: a state-of-the-art survey," *Knowledge and Information Systems*, vol. 1, no. 2, pp. 129–156, 1999.

[5] Y. Koren, U. Heisel, F. Jovane et al., "Reconfigurable manufacturing systems," *CIRP Annals*, vol. 48, no. 2, pp. 527–540, 1999.

[6] H. Van Brussel, J. Wyns, P. Valckenaers, L. Bongaerts, and P. Peeters, "Reference architecture for holonic manufacturing systems: PROSA," *Computers in Industry*, vol. 37, no. 3, pp. 255–274, 1998.

[7] K. Ueda, "A concept for bionic manufacturing systems based on DNA-type information," in *Human Aspects in Computer Integrated Manufacturing*, pp. 853–863, Elsevier, 1992.

[8] R. Pfeifer, M. Lungarella, O. Sporns et al.M. Lungarella, F. Iida, J. Bongard et al., "On the information theoretic implications of embodiment – Principles and methods," in *50 Years of Artificial Intelligence: Essays Dedicated to the 50th Anniversary of Artificial Intelligence*, pp. 76–86, Springer, Berlin, Heidelberg, 2007.

[9] A. W. Colombo, T. Bangemann, S. Karnouskos et al., *Industrial Cloud-Based Cyber-Physical Systems*, The IMC-AESOP Approach, 2014.

[10] F. Jammes and H. Smit, "Service-oriented paradigms in industrial automation," *IEEE Transactions on Industrial Informatics*, vol. 1, no. 1, pp. 62–70, 2005.

[11] P. Leitao, S. Karnouskos, L. Ribeiro, J. Lee, T. Strasser, and A. W. Colombo, "Smart agents in industrial cyber–physical systems," *Proceedings of the IEEE*, vol. 104, no. 5, pp. 1086–1101, 2016.

[12] M. Glinz, "A glossary of requirements engineering terminology," in *Standard Glossary of the Certified Professional for Requirements Engineering (CPRE) Studies and Exam*, IREB - International Requirements Engineering Board, 2011.

[13] C. Rupp, *Requirements-Engineering und-Management: Professionelle, Iterative Anforderungsanalyse für die Praxis*, Hanser Verlag, 2007.

[14] L. Ribeiro, S. Karnouskos, P. Leitão, and T. I. Strasser, "A community analysis of the IEEE IES industrial agents technical committee," in *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*, pp. 6139–6144, Beijing, China, 2017.

[15] L. Ribeiro, "Cyber-physical production systems' design challenges," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*, pp. 1189–1194, Edinburgh, UK, 2017.

[16] L. Ribeiro and M. Björkman, "Transitioning from standard automation solutions to cyber-physical production systems: an assessment of critical conceptual and technical challenges," *IEEE Systems Journal*, pp. 1–13, 2017.

[17] N. Chiriac, K. Hölttä-Otto, D. Lysy, and E. Suk Suh, "Level of modularity and different levels of system granularity," *Journal of Mechanical Design*, vol. 133, no. 10, article 101007, 2011.

[18] A. M. Farid, "Measures of reconfigurability and Its key characteristics in intelligent manufacturing systems," *Journal of Intelligent Manufacturing*, vol. 28, no. 2, pp. 353–369, 2017.

[19] A. M. Farid and L. Ribeiro, "An axiomatic design of a multiagent reconfigurable mechatronic system architecture," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 5, pp. 1142–1155, 2015.

[20] N. P. Suh, *The Principles of Design (No. 6)*, Oxford University Press on Demand, 1990.

[21] Y. Koren and M. Shpitalni, "Design of reconfigurable manufacturing systems," *Journal of Manufacturing Systems*, vol. 29, no. 4, pp. 130–141, 2010.

[22] M. G. Mehrabi, A. G. Ulsoy, and Y. Koren, "Reconfigurable manufacturing systems: key to future manufacturing," *Journal of Intelligent Manufacturing*, vol. 11, no. 4, pp. 403–419, 2000.

[23] G. D. M. Serugendo, M.-P. Irit, and A. Karageorgos, "Self-organisation and emergence in MAS: an overview," *Informatica*, vol. 30, no. 1, pp. 45–54, 2006.

[24] M. Broy, M. V. Cengarle, and E. Geisberger, "Cyber-physical systems: imminent challenges," in *Large-Scale Complex IT Systems. Development, Operation and Management. Monterey Workshop 2012*, pp. 1–28, Springer, Berlin, Heidelberg, 2012.

[25] S. K. Khaitan and J. D. McCalley, "Design techniques and applications of cyberphysical systems: a survey," *IEEE Systems Journal*, vol. 9, no. 2, pp. 350–365, 2015.

[26] S. Balasubramanian, R. W. Brennan, and D. H. Norrie, "An architecture for metamorphic control of holonic manufacturing systems," *Computers in Industry*, vol. 46, no. 1, pp. 13–31, 2001.

[27] R. F. Babiceanu and F. F. Chen, "Development and applications of holonic manufacturing systems: a survey," *Journal of Intelligent Manufacturing*, vol. 17, no. 1, pp. 111–131, 2006.

[28] C. Pach, T. Berger, T. Bonte, and D. Trentesaux, "ORCA-FMS: a dynamic architecture for the optimized and reactive control of flexible manufacturing scheduling," *Computers in Industry*, vol. 65, no. 4, pp. 706–720, 2014.

[29] A. Giret and V. Botti, "Holons and agents," *Journal of Intelligent Manufacturing*, vol. 15, no. 5, pp. 645–659, 2004.

[30] P. Leitão, A. W. Colombo, and F. Restivo, "A formal specification approach for holonic control systems: the ADACOR case," *International Journal of Manufacturing Technology and Management*, vol. 8, no. 1/2/3, pp. 37–57, 2006.

[31] P. Leitão and F. Restivo, "ADACOR: a holonic architecture for agile and adaptive manufacturing control," *Computers in Industry*, vol. 57, no. 2, pp. 121–130, 2006.

[32] P. Leitão, J. Barbosa, and D. Trentesaux, "Bio-inspired multi-agent systems for reconfigurable manufacturing systems," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 5, pp. 934–944, 2012.

[33] V. Marik and D. McFarlane, "Industrial adoption of agent-based technologies," *IEEE Intelligent Systems*, vol. 20, no. 1, pp. 27–35, 2005.

[34] M. R. Endsley, *Designing for Situation Awareness: an Approach to User-Centered Design*, CRC press, 2016.

[35] M. Riedl, H. Zipper, M. Meier, and C. Diedrich, "Cyber-physical systems alter automation architectures," *Annual Reviews in Control*, vol. 38, no. 1, pp. 123–133, 2014.

[36] J.-S. Yoon, S.-J. Shin, and S.-H. Suh, "A conceptual framework for the ubiquitous factory," *International Journal of Production Research*, vol. 50, no. 8, pp. 2174–2189, 2012.

[37] M. Moghaddam and S. Y. Nof, "The collaborative factory of the future," *International Journal of Computer Integrated Manufacturing*, vol. 30, no. 1, pp. 23–43, 2017.

[38] X. Xu, "From cloud computing to cloud manufacturing," *Robotics and Computer-Integrated Manufacturing*, vol. 28, no. 1, pp. 75–86, 2012.

[39] E. A. Lee, "Cyber-physical systems-are computing foundations adequate," in *Position Paper for NSF Workshop On Cyber-Physical Systems: Research Motivation, Techniques and Roadmap*, pp. 1–9, Austin, TX, USA, 2006.

[40] A. Banerjee, K. K. Venkatasubramanian, T. Mukherjee, and S. K. S. Gupta, "Ensuring safety, security, and sustainability of mission-critical cyber–physical systems," *Proceedings of the IEEE*, vol. 100, no. 1, pp. 283–299, 2012.

[41] D. SPEC, "91345: 2016-04," *Reference Architecture Model Industrie 4.0 (RAMI4. 0)*, 2016.

[42] P. Leitão and S. Karnouskos, *Industrial Agents: Emerging Applications of Software Agents in Industry*, Morgan Kaufmann, 2015.

[43] A. ISA, "ISA-95.00. 01-2000: enterprise-control system integration part 1: models and terminology," Technical report, ISA, The Instrumentation, Systems, and Automation Society, 2000.

[44] M. A. Bedau, "Weak emergence," *Noûs*, vol. 31, no. s11, pp. 375–399, 1997.