IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 4 March 2024, accepted 17 March 2024, date of publication 21 March 2024, date of current version 3 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3380317

RESEARCH ARTICLE

A Network Modeling and Analysis Approach for Pharma Industry Regulatory Assessment

THEODOSIA CHARITOU[1,2], EFTHIMIOS LALLAS[2], VASSILIS C. GEROGIANNIS[3], AND ANTHONY KARAGEORGOS[2]

[1]Department of Computer Science and Biomedical Informatics, University of Thessaly, 38221 Larissa, Greece
[2]School of Technology, University of Thessaly, 38221 Larissa, Greece
[3]Department of Digital Systems, University of Thessaly, 38221 Larissa, Greece

Corresponding authors: Theodosia Charitou (tcharitou@uth.gr) and Anthony Karageorgos (karageorgos@uth.gr)

This work was supported in part by the Auspice of the "Smart Pharmaceutical Manufacturing (SPuMoNI)" Research Project funded by CHIST-ERA, from 2019 to 2022; and in part by the 352 Horizon 2020 Future and Emerging Technologies Programme of the European Union 353 through the ERA-NET Co-Fund Funding Scheme under Grant CHIST-ERA BDSI Call 2017.

ABSTRACT Regulatory compliance in the pharmaceutical industry is challenging, requiring dedicated resources and meticulous control over production processes to ensure adherence to established regulatory guidelines, specifically ALCOA+ (Attributable, Legible, Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, and Available) principles. This paper introduces an innovative approach to assess pharma regulatory compliance, utilizing a network model of the production process. The model dynamically configures production line characteristics based on manufacturing process data, overcoming complexity and scalability challenges. Purpose: The main purpose is to address the challenges of regulatory compliance in the pharmaceutical industry by introducing a novel approach using a network model. The research question involves assessing the effectiveness of this model in ensuring compliance with ALCOA+ principles. Methods: The approach involves dynamic configuration of the network model parameters based on manufacturing process data. Network analysis methods are then applied to evaluate the conformity of manufacturing process data to ALCOA+ principles. Results: Testing the proposed approach on a real dataset from a representative pharma production line demonstrates its effectiveness in assessing pharma regulatory compliance. The results highlight the potential of network modelling in managing data quality and integrity within the regulatory framework. Conclusions: The study concludes that the network model offers a strategic solution for evaluating and ensuring regulatory compliance in pharmaceutical manufacturing. The approach shows promise in addressing the complexities of data management within the stringent regulatory framework of the industry.

INDEX TERMS ALCOA+, betweenness centrality, graph network modeling, network analysis, pharma industry, regulatory compliance.

I. INTRODUCTION

The complex nature of modern industrial manufacturing infrastructures and relevant business processes urge for a set of accurate, compact, and well-defined rules, so as to be managed in ultimate transparency and ensure corporation integrity. Those rules are normally emanated from various types of standards, specifications and even government laws, so as to compose a complete framework, to be applied in various sectors such as health, pharmaceutical, banking, data security, manufacturing, ecological sectors, to name but a few. The real challenge lies in the fact that those rules tend to increase rapidly in numbers and complexity rate, and therefore require a dynamic regulatory framework along with efficient regulatory management to keep up with them. Regarding pharma manufacturing sector, those regulatory rules are in perfect alignment with the 9 ALCOA/ ALCOA+ principles (Attributable, Legible,

The associate editor coordinating the review of this manuscript and approving it for publication was Hongwei Du.

Contemporaneous, Original, Accurate, Complete, Consistent, Enduring, and Available) [1], that emphasize on ensuring data integrity, and product quality, as well [2]. Numerous factors have contributed to the growing importance of data integrity requirements and their efficient management, such as i) the need to maintain the quality and safety of medicines, ii) the ever increasing stringent requirements concerning data traceability and audit trails, iii) the evolving consumer and end-user expectations that boost competition within the pharmaceutical industry, and iv) the dependence of Industry 4.0 technologies and systems on enhanced equipment integration and digital data transfer. Hence, dealing effectively with pharma regulatory compliance is a difficult and challenging task, as pharma manufacturing environments tend to grow up rapidly, merging heterogeneous production lines, along with complex raw data originated from production line infrastructures.

II. MOTIVATION-BACKGROUND

Various research attempts have been focused in modelling that complex, ever evolving pharma environment, along with an appropriate regulatory compliance framework. There is a plethora of approaches which have been successfully applied on other sector business process modelling, and they are mainly based on combining well established knowledge management semantic techniques (Resource Description Framework-RDF, natural language processing-NLP, etc), with semantic web based representations, such as ontologies coupled with Web Ontology Language (OWL) [3], or NLP accordingly [4]. The successful results of the combination of conceptual modelling with semantic representation have driven the research community to adopt relevant techniques in various industrial environments, including the pharma manufacturing sector. There are many NLP [5], or ontology based conceptual representations which have been attempted in pharma industry regulatory sector, such as, [6], [7], [8], [9], [10], [11]. Artificial Intelligence (AI) and Machine Learning (ML) methods have already shown great results in supporting those pharma manufacturing systems in multiple ways, such as process monitoring auditing, and control, data mining and processing, digital transformation integration with other technologies (e.g. blockchain), [12], [13], and prediction potential [14]. Digital Twin (DT) solutions have also been successfully applied, facilitating the transformation of pharma manufacturing environments [15]. However, as already mentioned, in those dynamically evolved and scalable pharma manufacturing environments, being characterized by large heterogeneous data volumes and continuously changing bodies of rules, it is hard to tackle efficiently with data and compliance management. Optimized data driven approaches in knowledge models may be a promising solution [16]. The existing conceptual semantic based representation approaches sooner or later will have to face bottlenecks of computation processing performance due to their inherent polynomial dependence in the size of data input batch reports, especially as they tend to scale up fast and merge more and more heterogeneous production lines [17]. At this point, the idea of managing graph network structured data, instead of data word semantics, along with their multiple relationships, may lead to a promising and lightly processing burden model solutions.

III. CONTRIBUTION AND STRUCTURE OF THE PAPER

Inspired by the above motivations, in this research work we propose a network analysis method that applies pharma regulatory compliance by assessing ALCOA principle violation, in a network modelling of a pharmaceutical line process which are analysed using graph theory. This network modelling is a simulation of the pharma production manufacturing process via adoption of a graph network structure, whereas its nodes represent production line data and infrastructure resources, interconnected with each other via edges accordingly, when there is a pharma domain relationship and interaction. ALCOA+ principles have been also represented by graph network nodes allowing for interconnection via edges, with other node types whenever their predefined criteria are met. On the contrary, any nodes not meeting ALCOA+ criteria result in reduced connectivity, which means violation. The network analysis approach can be applied to check for any reduced connectivity across ALCOA+ nodes, indicating the significant edges decrease, and providing ALCOA+ compliance assessment accordingly. This approach simplifies the compliance procedure, as it only checks for network connectivity changes along ALCOA+ nodes, and hence significantly reduces the computation time required, especially when dealing with heavy topologies that represent scalable and complex pharma manufacturing production lines, with multiple batches and recipe combinations. This paper is structured as follows. A related work about network modelling section follows up introduction section. Section III in Material and Methods is dedicated to an analytical method description, including graph network design and construction details, the batch report data and ALCOA+ principle inclusion within network structure, and the network analysis scope and metrics selection. Section IV describes and discusses over the results of network analysis, associating them with ALCOA+ violation percentages, and elaborating thus on a regulatory compliance assessment. Finally, section V concludes the paper and indicates future goals and prospects in a next work version.

IV. BACKGROUND ON NETWORK MODELING

The study of network models and their characteristics is inspired by empirical analysis of real networks. A large part of the research work dealing with complex networks in manufacturing either discovers structural characteristics of those networks or investigates the relation between structural network characteristics and the performance of the material flow [18]. Assessing the actual material flow in a network provides more reliable information than considering network structural properties. However, the advantage of using network measures is that they provide a systematic

and quantitative way to analyse and understand the properties and behaviour of networks, making them a valuable tool in a wide range of fields where network analysis is applicable. Even when actual material data is not yet available and can be only acquired using computer simulations, as is the case in early planning stages for instance, assessment about the manufacturing system can still be done using network measures [19], [20]. Examples are networks formed by human social and professional relationships, and networks found in natural sciences such as physics and biology. Many network-related insights can be transferred to various technological application domains, such as manufacturing and engineering, by considering networks formed by the flow of material, goods, and information. These artificial networks, although different from networks found in nature or society, have similar structure and properties because they are governed by the same principles. Therefore, the same set of mathematical and computational tools can be used to analyse them. Graph network modelling and network analysis has been considered as a state-of-the-art modelling approach for a long time in many application and scientific sectors, such as Network science, Network biology, Network medicine, or application sectors such as manufacturing and logistics industries. Network science, as an interdisciplinary field of study, plays a role in the utilization of the growing availability of network data to investigate a wide range of complex phenomena, such as collective social behaviors, technological progress, financial stability, and biological interactions [21], [22], [23]. Network biology is a rapidly advancing field of research that recognizes biological processes as intricate systems governed not solely by individual proteins or isolated linear pathways but rather by complex, interconnected networks of molecular interactions, often referred to as the interactome [24], [25]. Network medicine theory extends this notion, proposing that disease-associated characteristics do not stem from single gene mutations acting in isolation but rather result from disruptions within a gene's network context [26], [27], [28]. Researchers have access to a multitude of bioinformatics tools and molecular interaction data for constructing networks from gene/protein lists and exploring novel systems-level insights into the target phenotype [29]. Moreover, manufacturing and logistics industries benefit from this development since material flow systems are inherently suited to be modelled as networks. Selected advances in network modelling and analysis within the manufacturing and logistics industries that have been achieved in the past decade are provided in a review paper that highlights the fundamental modelling concept, and several examples of how network models can be utilized to contribute to the resolution of planning and control issues within logistic systems. With the use of network models, it is possible to gain insights into the functioning of logistic systems and identify areas where improvements can be made, leading to increased efficiency and cost savings [30]. A further development in network modelling in manufacturing was the introduction of stochastic models. In particular, the Stochastic Block Model (SBM) allows for a prediction of future states in a manufacturing system [18] and testing different variants and classical machine learning methods for prediction. An SBM is a network model in which groups of similar nodes (such as clusters) are seen as structural equivalent. Instead of explicitly modelling the links between nodes, the general probability of two nodes from two groups being connected is given. In another approach network modelling has been used to study phenomena occurring in interacting economical agents, such as bank bankruptcies, as well as relationships between shareholders, board directors, and stock prices [31]. In business ecosystems in particular, networks have been used to represent relationships and interactions between stakeholders, processes, products, and financial entities to identify bottlenecks, inefficiencies, and opportunities for improvement [19], [20], [32], [33], [34], [35], [36], [37]. Recent research indicates that the simulated network structure flow can undergo changes over a specific period due to various events or changing circumstances, and hence, can be applied in dynamic environments as well (e.g. job transfers or workstation malfunctions). This has been demonstrated in recent studies by [38], [39]. Therefore, it may be advantageous to consider such dynamic processes in specific applications. In general, such network modelled systems are termed complex because it is generally not possible to exactly predict their collective behaviour based on the individual behaviours of their components. However, understanding the mathematical description of these systems enables predicting certain system properties and subsequently taking appropriate control actions. Along this line many dynamical processes, from biological until technological contexts such as epidemic dynamics [40], [41], vector-borne or livestock diseases [42], [43], spreading rumors [44], [45], and synchronization [46], [47] have been investigated. Furthermore, the inclusion of Artificial Intelligence (AI) methods in the processing of network data can further increase the quality of network models. Massive amounts of system data will be collected in the context of Industry 4.0 and enabling detailed analysis of network models. This will significantly increase the transfer of existing network analysis methodologies into practical applications for decision-making in planning, operation, and control of industrial and business systems. The common approach is to model certain parts of an application domain as an artificial network, analyse the network using existing methods and tools, and based on the network analysis results draw conclusions and take control decisions concerning that application domain. Examples include machine learning models [47], semantic representations of language notations [48], [49], systems biology investigations, and public sector organisations and policy networks [50].

V. PROPOSED METHODOLOGY

As mentioned, the method proposed in the current paper is a network modelling approach that simulates pharma production manufacturing process as a graph, typically consisting of nodes and edges. Particularly, the manufacturing line process

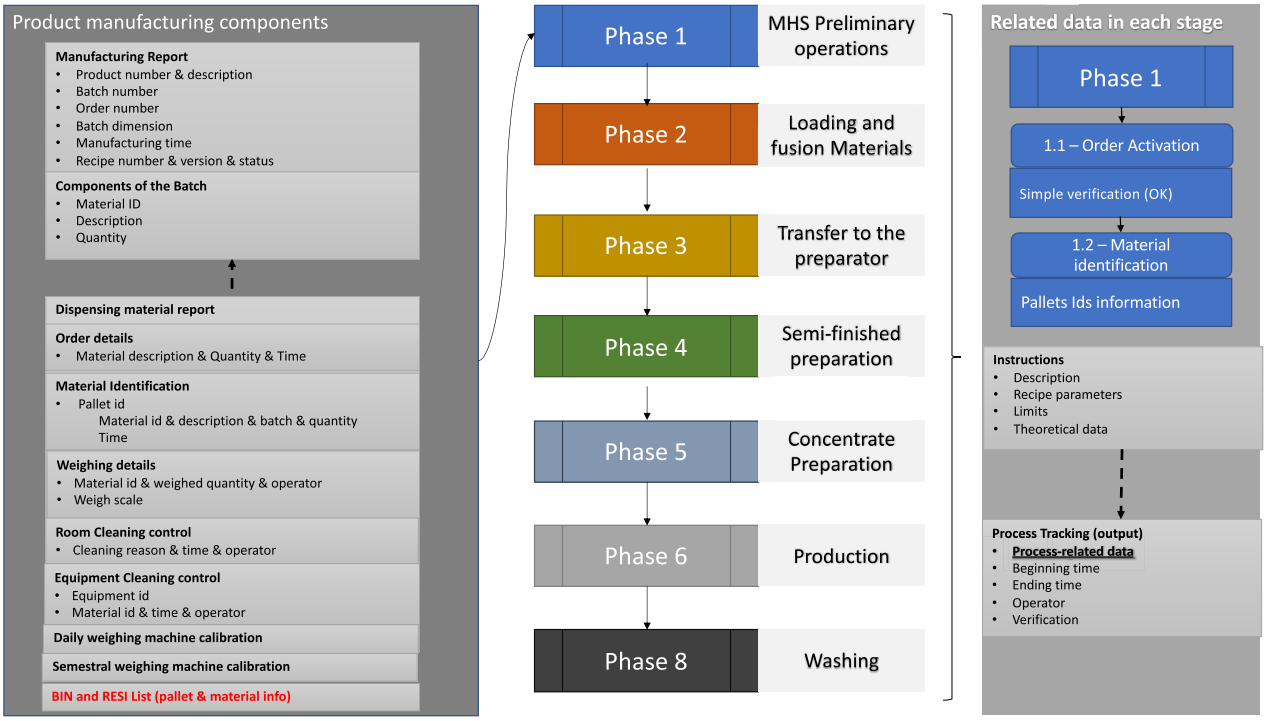


FIGURE 1. Product manufacturing components of each recipe.

has been considered as a network, with its components being related to product materials, recipe procedures and instructions, and machinery equipment, represented as nodes, and their relationships as edge connections, accordingly. Apart from regular pharma component nodes, ALCOA+ principles have been considered and included as nodes within graph structure, as well. ALCOA+ node interconnections with other regular nodes have been set, according to predefined criteria, emanating from corresponding ALCOA+ principles [1]. E.g. for Attributable principle, the corresponding ALCOA+ node has been related and hence be connected to a regular node that represents pharma system personnel.

As soon as the graph network structure has been completely set, representing the overall production line process for a given recipe of a product, then the simulation runs, and the network analysis, which has been based upon many metrics, and mostly on betweenness centrality, has been searching for ALCOA+ node connectivity status, for a given batch recipe report. In case of a batch that experiences ALCOA+ violation, it would result in reduced connectivity, compared to an ideal, fully complied batch. Hence, our novel network analysis has been checked for any reduced connectivity across ALCOA+ nodes, indicating the significant edges decrease, and providing ALCOA+ compliance assessment accordingly. The proposed approach includes the following methodology steps: i) Network design and construction, ii) ALCOA+ node integration and iii) Network connectivity validation – network analysis.

A. NETWORK DESIGN AND CONSTRUCTION

In this study, as an initial step we start by defining the manufacturing line process as a network, identifying its various phases, and the components of each phase. Each component is related to product materials, recipe procedures and instructions, and machinery equipment that can be represented as a node, and the connections between them as edges respectively. Normally in pharma industry, the automated control and management system (Manufacturing Execution System - MES and SCADA/PLC) of a pharma manufacturing product line, are responsible as entitled, for the automated control and management of the manufacturing line process and the corresponding execution of the recipes for each product. At the end of each process for a given recipe of a product (order), these systems provide a comprehensive batch recipe report. Hence, each order recipe is uniquely associated with a specific batch report. The report includes essential details such as product identification, unique identifiers, batch dimensions, processing start and finish dates, recipe identification, version information, approval date, as well as comprehensive information regarding the materials/components utilized in the production line. The materials/components are labelled with their respective identifiers, full names, and characteristics, including quantities measured in grams (gr) or kilograms (kg), as described in Fig 1. These nodes and their relationships (edges) have been defined by batch recipe report data, and identify the distinct phases within the recipe. To achieve this, we established the network structure for the product manufacturing line process using a dataset comprising recipes

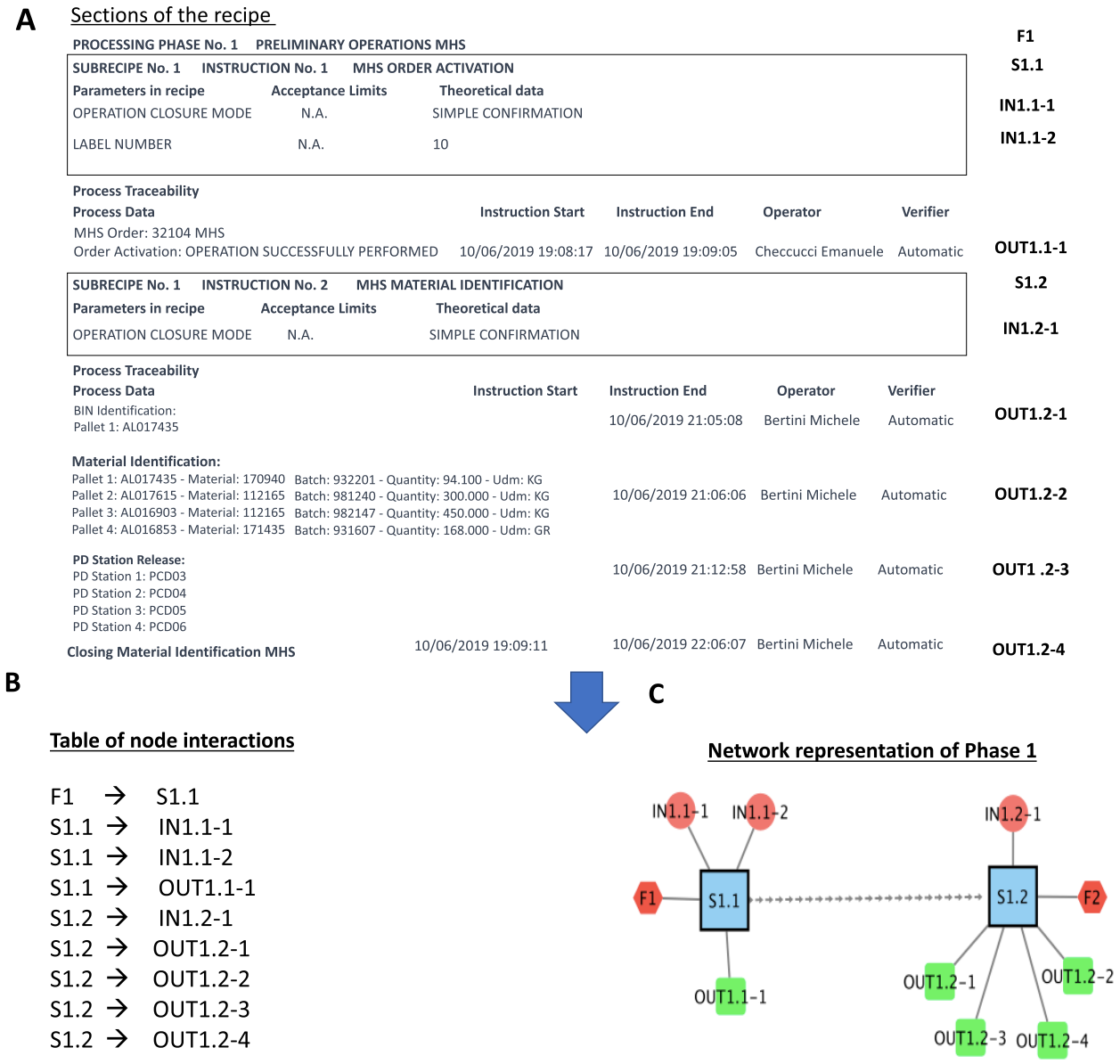


FIGURE 2. The production line data workflow for constructing graph. A) The first step is to split the recipe into sections related to the flow and dependencies between phases, sub-recipes, input data, and output messages. We use a part of the recipe that includes a phase (F1) with two sub-recipes (S1.1) and (S1.2). Each of these recipes includes input data (IN) and output messages (OUT) of the manufacturing production workflow. B) After separating the sections in the recipe, the sections are listed as nodes. Each node is connected hierarchically as the raw material flows through the processing, with phase 1 connected to its sub-recipe 1 (S1.1) and sub-recipe 1 connected to the next sub-recipe (S1.2), indicating the flow of material between them. Full list of networks of that format is shown in Supplementary materials. Then each sub-recipe connects to the input and/or output message nodes (IN, OUT). Finally, the sub-network of phase 1 is shown in C as a graph.

obtained from a pharmaceutical company. Each phase is characterized by a series of sub-recipes, each assigned a unique two prefix number, and fully defined with process-related data. Those data, as seen for Phase 1 (Supplementary Fig 2A), have been originated from batch reports, and are normally provided in pdf format. At this network construction stage, a pdf parser has been adopted that converts the recipe report, initially provided in PDF format, into a code formatted text file with well-defined node connections, so as to input network simulator and compose the graph structure. We further classify each node component within a sub-recipe as input data and be represented as input nodes in the graph. All parameter nodes are considered as inputs and hence, they are connected to the respective sub-recipe nodes. E.g for example in the batch report sample for phase 1 in Figure 2A the original component node "LABEL NUMBER" has now been considered as an input node for phase 1 (F1), assigned with label IN1.1-2. Moreover, at the end of each sub-recipe, there is an output message indicating the successful completion of the operation. These output messages are represented as

output nodes in the graph and depending on the parameters they are referred to, they are connected to the corresponding sub-recipe nodes. In Figure 2A can be seen 5 in total output nodes. Hence, Figure 2 illustrates the production line data workflow for constructing graph for phase I (F1) and its two sub-recipes S1.1 and S1.2 respectively. Fig 2A shows the batch report table for F1 phase of a given recipe, including all input and output data that required for capturing the flow and dependencies between phases and sub-recipes, in the data collection workflow. Those dependencies are mapped into network connections between stages of the manufacturing production line and presented in the edge table of Fig 2B. The final network construction (connections) of phase 1 can be seen in Fig 2C.

The production line data workflow is divided into phases, with each phase represented by a node labelled with the letter ''F'' followed by a number (e.g., F1, F2). Within each phase, there are sub-recipes identified by a two prefix numbers. Each sub-recipe is denoted by the letter ''S'' followed by its phase number (1st prefix) and its unique sequence number (e.g., S1.1, S1.2). The connections between the phases and sub-recipes are established as follows: The first sub-recipe of each phase is connected to the corresponding phase node, and subsequent sub-recipes are connected to each other. For example, F1 is connected to S1.1, and S1.1 is connected to S1.2. If there are no additional sub-recipes within a phase, the connection continues with the next phase, such as F2. Hence, S1.2 is connected to the F2 node. The nature of these connections can be either directed or undirected, depending on the flow of material in each stage. The parameters associated with the sub-recipes are referred to as input data and are represented as nodes connected to each sub-recipe in the graph. For instance, a node labelled as IN1.1-1 indicates that it belongs to sub-recipe S1.1 and is connected to its representative S1.1 node. Similarly, all parameter nodes are connected as inputs to the respective sub-recipe nodes. At the end of each sub-recipe, there is an output message indicating the successful completion of the operation. These output messages are represented as output nodes in the graph and connected to the corresponding sub-recipe nodes. For example, S1.1 has two input nodes, IN1.1-1 and IN1.1-2, and one output node, OUT1.1-1. Thus, all three nodes are connected to S1.1. We add direction to the edge when one stage is depending to another for production to indicate the material flow (e.g. S1.1 towards S1.2).

B. ALCOA+ NODE INTEGRATION

To complete network design and construction, we need to import ALCOA+ nodes within graph structure, as well. At first, we need to identify the relationships of ALCOA+ principles with batch record data parameters, and then to represent them with network connections accordingly. Hence, we need to associate the ALCOA+ principles represented by ALCOA+ nodes, with the corresponding data points into the existing network model. ALCOA+ principles, may deal with different format data types (e.g. single numerical values, binary values, time series numerical values, categorical data string values, etc) emanating from corresponding multiple different data sources. Normally, all such data types of a pharma production process, have been included within an electronic batch record (EBR). This ALCOA+ node association procedure may require inference of metadata or data annotations to the nodes or edges that represent the batch record data points. This can involve automated checks, audits, or quality control procedures at various stages of the manufacturing process. For example, the Attributable principle node, which is responsible for writing down the authorized personnel for each activity, has been directly associated to the datapoint that represents pharma system personnel. The Consistent principle, on the other hand, which is responsible for checking all operations with improper time sequence, is based on inferred knowledge of time comparison of timestamp data points. Should we have finished with relationship identification, then we could proceed with including the appropriate connections accordingly. As mentioned, batch record data could be either input (sub-recipe parameters) or output data (output messages), and hence the ALCOA+ node connections would be referred to either one of these generic node types. Fig 3 illustrates integration of ALCOA+ principle into the network model.

C. NETWORK CONNECTIVITY VALIDATION–NETWORK ANALYSIS

In the suggested approach, each ALCOA+ principle is separately considered, meaning that we have devised distinct network topologies, each corresponding to a specific ALCOA+ principle. These topologies were constructed based on the methodology steps described above. In the following, we present an overview for six (out of nine) ALCOA+ principles, describing the connectivity pattern for each one, in the ideal case that there is no rule violation (Table 1). The remaining three ALCOA+ principles require extra software module for representing their total graph structure, and hence they will be considered in a future version of our research work.

TABLE 1. Connections between ALCOA+ nodes and graph nodes.

Attributable Principle	Output nodes
Legible Principle	Input and output nodes
Contemporaneous Principle	Output nodes
Accurate Principle	Output nodes
Complete Principle	Input and output nodes
Consistent Principle	Output nodes

This connectivity matrix corresponds to an ideal, fully connected topology, named Topology 1, which will be used for comparisons with non-ideal topologies, named Topology 2, that represent ALCOA+ violation cases. In cases where a particular principle is violated, the connectivity between the ALCOA+ node and the corresponding node is severed. The

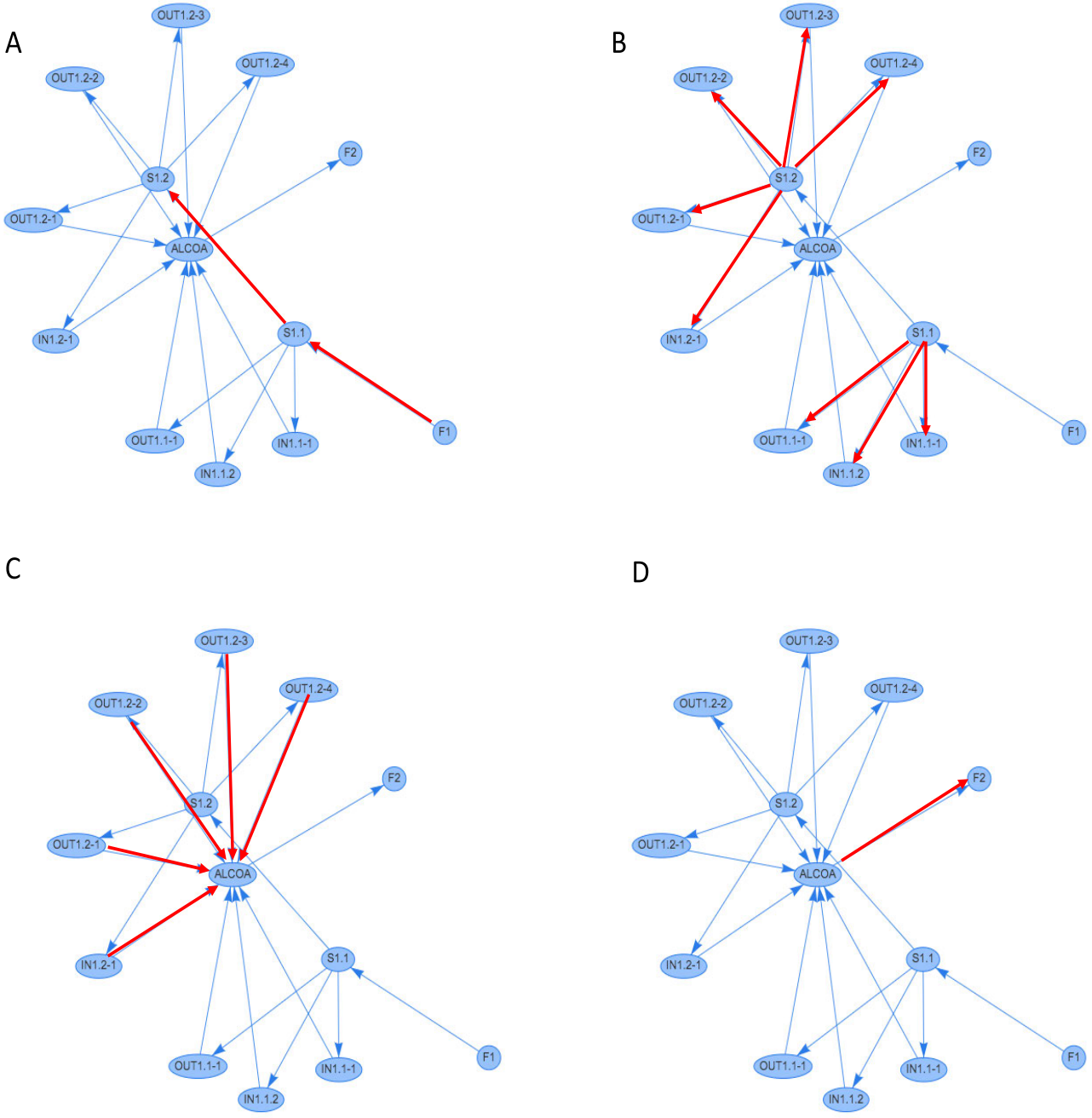


FIGURE 3. Illustrative Explanation of ALCOA+ node integration (Complete Network) for Phase 1. In this illustrative example, we elucidate the process of completing our network model to encapsulate Phase 1 activities with ALCOA+ node inclusion. A) Step 1: Phase Segmentation and Sub-Recipes As an initial step, we segment the phases within the production recipe, resulting in seven discrete phases designated as F1, F2, F3, F4, F5, F6, and F8. Focusing specifically on Phase 1 (F1), we identify two sub-recipes denoted as S1.1 and S1.2. Our approach involves establishing a sequential linkage between F1 and S1.1, followed by the subsequent connection of S1.1 with S1.2 (highlighted by the red line in the graphical representation). B) Step 2: Sub-Recipe Elaboration As indicated in subsection 2.1, each sub-recipe graph structure is consisted of two generic node types, namely inputs and outputs, characterized as 'IN' and 'OUT' nodes respectively. For instance, in S1.1, two inputs (IN1.1-1 and IN1.1-2) and one output (OUT1.1-1) are identified, and corresponding connections are established as depicted by the red line. C) Step 3: Integration of ALCOA+ Nodes The inclusion of ALCOA+ nodes is a pivotal augmentation within the network. The positioning and connections of ALCOA+ nodes are based upon the ALCOA+ principles. In a hypothetical scenario, let's consider an evaluation of completeness ALCOA+ principle. Should all nodes exhibit complete values, an ALCOA+ node is connected to both input and output nodes. However, if a value is missing, no interaction occurs between the ALCOA+ node and the affected node. D) Step 4: Successive phase connections ALCOA+ nodes are strategically connected to the succeeding phase, thus conveying the ALCOA+ assessment status across the network, as denoted by the red line. This linkage enables the propagation of ALCOA+-related assessments across successive phases, and the potential to provide segmented assessment on a per phase basis, enabling to locate those violations across production line, as well.

same topology name type applies for each principle, and hence the comparison will be held separately for each principle between the two network type Topologies, validating their connectivity mismatch. Such network modelling analyses are normally based on investigation of centrality metric. There are 4 centrality types used to analyze a manufacturing process network, namely degree, betweenness, closeness and eigenvector centralities. Among them we choose betweenness centrality, as we want to focus and emphasize on specific ALCOA+ nodes, and on their influence degree of the process flow of pharma production line, being characterized by edge connectivity across them. We can conclude by reporting batch report samples of particular sub-recipes for a given recipe, concerning ALCOA+ violation cases which lead to a decreased edge connectivity, Topology 2 graph, to be compared with ideal Topology 1 graph structure. Figure 6 depicts an overview of the proposed methodology and workflow. ALCOA+ node integration has also been included within workflow for Attributable principle. In future work, we plan to provide a complete and automated implementation for ALCOA+-based regulatory assessment, by using a single network structure, including all 9 ALCOA+ principles in a single digitalized workflow.

1) ATTRIBUTABLE PRINCIPLE

The Attributable Principle mandates that task assignments be endorsed solely by authorized individuals. We have identified in a batch recipe report sample (pdf file), that in sub-recipe 6.6 the output node 8 is not attributed with an operator, as well as in the sub-recipe 8.1, the output node 1. Consequently, we removed the two edges, finalising topology 2 for ALCOA+ attributable principle, before comparing it with topology 1 (full attributable) to record the metrics of the network.

2) LEGIBLE PRINCIPLE

The Legible Principle underscores the necessity for data to possess readability characteristics, including specific encodings, formats, or linguistic attributes conforming to established word standards. To assess adherence, a lexicon was developed and employed for scrutinizing incoming pharmaceutical parameter values. We have checked at the batch report file for such violation in text formatting and we didn't identify any word inconsistency. Therefore, topology 1 and topology 2 are same.

3) CONTEMPORANEOUS PRINCIPLE

The Contemporaneous Principle dictates that data activities must be accompanied by timestamp records signifying their occurrence times. Notably, parameters lacking timestamp values were identified, exemplified by instances where (a specific timestamp ''1970-01-01T00:00:00'' was given instead) from the database system. In the sample batch report file, we identified 12 violations in output nodes of sub recipes (OUT1.2-1, OUT1.2-2, OUT1.2-3, OUT4.15-1, OUT6.6-1, OUT6.6-2, OUT6.6-3, OUT6.6-4, OUT6.6-5, OUT6.6-6, OUT6.6-7, OUT8.1-1). Therefore, Topology 2 contains 12 less edges than Topology 1.

4) ACCURATE PRINCIPLE

The Accurate Principle emphasizes the importance of data values falling within predefined acceptable ranges. Deviations from these ranges were categorized as violations and delineated as outliers in the dataset. In the sample batch recipe report, we didn't record any violation and therefore, topology 1 and topology 2 are same.

5) COMPLETE PRINCIPLE

The Complete Principle stipulates that obligatory data fields in reports must be populated and not left empty. In practice, this principle was verified by detecting null-valued parameters. We have checked all non-null values in input and output nodes and identified 8 nodes with null values (OUT6.6-1, OUT6.6-2, OUT6.6-3, OUT6.6-4, OUT6.6-5, OUT6.6-6, OUT6.6-8, OUT8.1-1). We excluded these 8 edges with ALCOA+ node in the Topology 2.

6) CONSISTENT PRINCIPLE

The Consistent Principle, mandates that the starting dates of batch report data items must precede their respective ending dates. To ascertain consistency, timestamps were compared chronologically. In this step, we have checked batch report for each finished timestamp of each sub-recipe output, comparing it, with the initiated timestamp of the next sub-recipe. We identified, the last output (OUT4.25-1) of sub-recipe S4.25 initiated at 12/06/2019 10:12 end finished at 12/06/2019 10:12 whereas the next output of sub-recipe S5.1 (OUT5.1-1) initiated at 12/06/2019 00:33 until 12/06/2019 00:33 instead of 13/06/2019 00:33. Similarly, there is not consistency on time difference between OUT5.5-1 (end: 12/06/2019 00:36) and OUT6.1-1 (start: 12/06/2019 10:18). In addition, the last output of sub-recipe S6.9-1 ended at 13/06/2019 03:15 whereas output of sub-recipe S8.1 started without timestamp. Consequently, we removed 6 edges connected with the corresponding ALCOA+ of consistency.

VI. RESULTS

A. NETWORK VISUALIZATION

In this section, following the steps of the proposed method, we simulate in the form of a network, the production process of a pharmaceutical product based on a recipe with number 932132. In this recipe, 7 phases were recorded, each of them contains a number of sub-recipes which in total are 58. For instance, phase 1 contains two sub-recipes, phase 2 contains 7, phase 3 contains 7, phase 4 contains 25, phase 5 contains 5, phase 6 contains 11, and phase 8 contains 1 sub-recipe. Each sub-recipe contains its own set of input (IN) and output data (OUT), represented as nodes accordingly. Six networks were constructed, each representing an ALCOA+ principle that needs to be assessed. The nodes of the ALCOA+ for each network are 7, the same

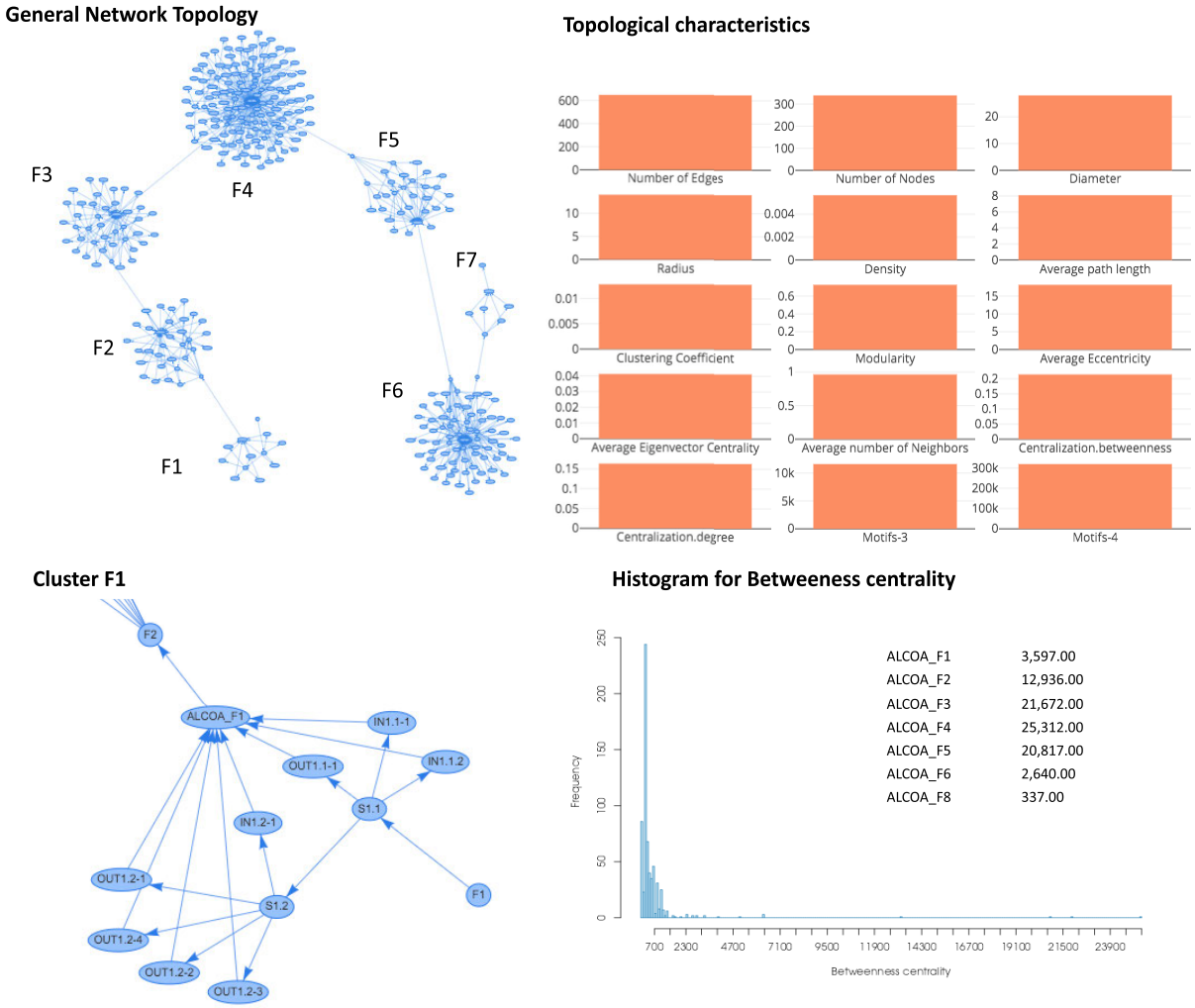
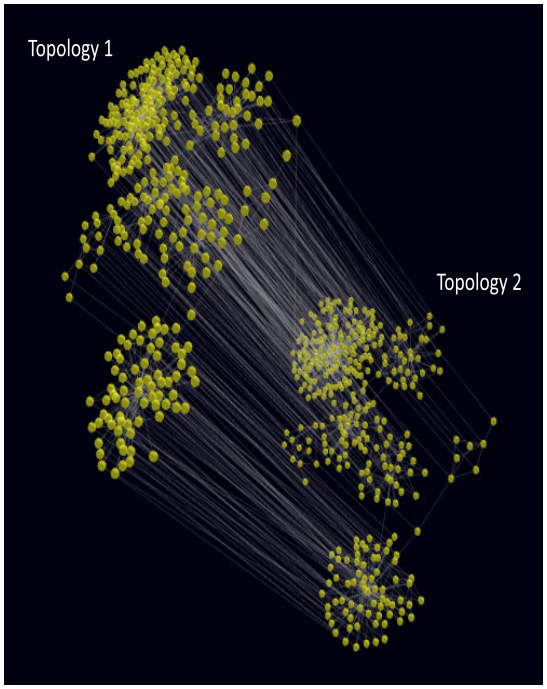


FIGURE 4. Network generation and identification of ALCOA+ centrality. Table 2 represents the definitions of each topological characteristics and what they physically mean.

as the number of phase nodes. The 7th ALCOA+ is connected to a pseudo-node representing the end of production node (FINAL). The six network topologies are given in the supplementary material as an edge list. We visually represent in 3D format the topologies for each ALCOA+ principle. Topology 2 is compared with Topology 1 of each individual ALCOA+ principle and connectivity statistics are measured using the Network Analysis Profiler (NAP) [51], a comprehensive web tool to automate network profiling and intra/inter-network topology comparison. In total, 12 networks are loaded in as a two-column binary list of connections as a tab delimited text file and in a directed topological feature. First column in the text file represents the source node and second column the target node. Then, we can automatically generate an inter-network topological analysis in order to directly compare pairwise networks for each ALCOA+ principle. Network analysis for each network is presented in Fig 4 giving values for each statistic. For the general network that presents the production process of recipe with number 932132, we have calculated the number of edges (647), the number of nodes (339), diameter (28), density (0.01), average path length (8.09), modularity (0.73), average eccentricity (18.16), average eigenvector centrality (0.04), average number of neighbours (0.96), centralization betweenness (0.21), centralization degree (0.16), motifs-3 (11,623.00) and motifs-4 (31,9732.00). In network analysis, ''motifs'' refer to specific subgraph patterns or configurations that occur with notable frequency within a larger network. Motifs are typically denoted by ''motif-N,'' where N represents the number of nodes or vertices involved in the subgraph. So, ''motif-3'' and ''motif-4'' specifically refer to subgraph patterns with three and four nodes representing a triangular relationship where nodes are all connected to each other, as a square pattern where four nodes are interconnected in a specific way, respectively. All these graph network features will be used for composing a thorough network analysis and estimating betweenness centrality metric, so as to finally compare the two Topologies, and come up with our finite

Network metrics Comparison for two topologies presenting one of the 6 ALCOA principles



	Attributable Principle		Legible Principle		Contemporaneous Principle	
	Topology 1	Topology 2	Topology 1	Topology 2	Topology 1	Topology 2
ALCOA_F1	2.616	2.600	3.597	3.597	2616	1625
ALCOA_F2	7.056	7.008	12.936	12.936	7056	6132
ALCOA_F3	10.080	10.000	21.672	21.672	10080	9250
ALCOA_F4	10.640	10.450	25.312	25.312	10640	10010
ALCOA_F5	8.667	8.453	20.817	20.817	8667	8137
ALCOA_F6	1.104	822	2.640	2.640	1104	762
ALCOA_F8	142	0	337	337	142	0
	98%		100%		89%	

	Accurate Principle		Complete Principle		Consistent Principle	
	Topology 1	Topology 2	Topology 1	Topology 2	Topology 1	Topology 2
ALCOA_F1	3.597	3.597	3.597	3.597	2.616	2.600
ALCOA_F2	12.936	12.936	12.936	12.936	7.056	7.008
ALCOA_F3	21.672	21.672	21.672	21.672	10.080	10.000
ALCOA_F4	25.312	25.312	25.312	25.312	10.640	10.230
ALCOA_F5	20.817	20.817	20.817	20.817	8.667	8.058
ALCOA_F6	2.640	2.640	2.640	2.584	1.104	786
ALCOA_F8	337	337	337	329	142	0
	100%		99%		96%	

FIGURE 5. ALCOA+ assessment through network analysis. On Left two topologies are compared regarding their connectivity. On the right, rewiring of ALCOA+ edge connectivity returns betweenness centrality measures for each ALCOA+ nodes in both Topologies. The fold change difference for all principles shows that Contemporaneous ALCOA+ principle is violated by 11%.

goal, the ALCOA+ assessment. Histogram of betweenness centrality in Fig 4 shows that the majority of nodes (250) have betweenness centrality less than 700 indicating that very few are bottlenecks in the network. Three nodes, the ALCOA+ F3-F5 ranked as nodes with highest betweenness centrality with more than 20,800 pathways crossing these nodes.

B. DATA INTEGRITY ASSESMENT WITH NETWORK ANALYSIS

We can establish the general topology of the network by connecting different phases with their sub-phases/sub-recipes in an order that represents the production flow as described in the recipe report (Top left). An overview of the first phase is shown in the bottom-left group, which includes 2 sub-recipes with a direction from S1.1 to S1.2. These sub-recipes consist of a set of nodes serving as input (IN) and output (OUT) nodes, which will be linked to an ALCOA+ node corresponding to a specific principle. We have analysed the network and recorded fundamental topological measurements pertaining to network characteristics such as Number of Edges, Number of Nodes, Diameter, Radius, Density, Average Path Length, Clustering Coefficient, Modularity, Number of Self-loops, Average Eccentricity, Average Eigenvector Centrality, Average Number of Neighbours, Centralization Betweenness, Centralization Degree, Motifs-3, Motifs-4 (Top right). Apart from centrality metric estimation, many of these graph network features can be used for estimating other metrics as well, that may lead to important conclusions regarding the behavioural patterns of the simulation system they represent. For example, diameter and eccentricity are critical parameters for the magnitude of graph network model, and their numerical behaviour can be used for investigating pharma manufacturing system performance as it scales up.

Two 3D animated network Topologies are compared by performing node-level analysis to assess ALCOA+ compliance. The clusters of the Topologies showing in Figure 5, represent production line phases. In each network Topology, calculation of relevant metrics such as betweenness centrality for each node has been performed. Our focus is on the measurements of the ALCOA+ nodes. The average number of betweenness centrality in ALCOA+ nodes of network with Topology 1, which is the ideal case, is compared to the Topology 2 and the fold change returns the percentage of ALCOA+ compliance for each principle. Those percentages can be seen at the end of each ALCOA+ principle two-column list, in the table of Fig 5. In our example for the recipe with number 932132, only contemporaneous principle is less than 90% indicating that lack of timestamps in data has a result of low connectivity between ALCOA+ nodes and input nodes, therefore ALCOA+ node decreased centrality and consequently the principle itself, has been violated by 11%, or alternatively, its compliance percentage is 89%. Legible and Accurate principles are 100% complied, as the topologies

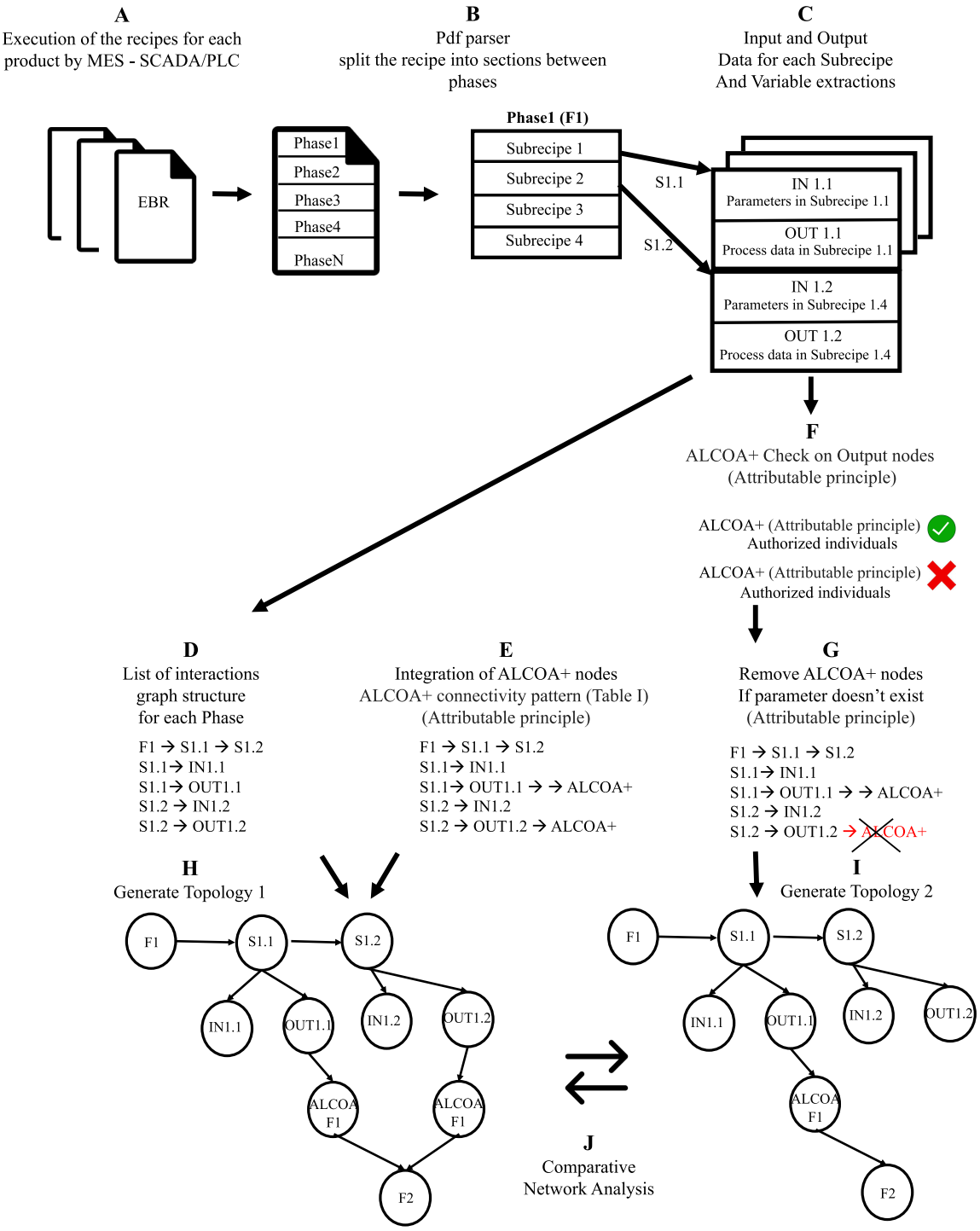


FIGURE 6. Workflow overview.

of the two network Topologies compared are the same. Finally, Attributable, Consistent and Complete principles, tend to ideal levels as well, with 98%, 96% and 99% percentages respectively. The remaining three ALCOA+ principles assessment, as mentioned will be included in a future version of our work. Moreover, as it can be seen in Fig 5, for a given principle, each ALCOA+ node has a different betweenness centrality at the end of each phase of production line and plays a crucial role in the process flow information of pharma production line. The position of ALCOA+ nodes at the end of each phase enables not only the process flow information propagation, but also the ALCOA+ violation assessments on a per phase basis. Hence via this method is feasible to locate the corresponding violations on a per phase basis across

TABLE 2. Centrality measures definitions. Understanding these centrality measures helps in characterizing different aspects of a network's structure, connectivity, and organization. The choice of centrality measure depends on the specific characteristics and goals of the network analysis. All metrics are calculated using the NAP web tool (See Proposed Methodology).

Num. of Edges	Indicates the overall level of connectivity in the network
Num. of Nodes	Represents the size of the network in terms of individual elements.
Diameter	Provides a measure of the "longest path" or the maximum communication distance in the network.
Radius	Represents the "shortest maximum distance" from any node to all other nodes in the network.
Density	Describes how connected the network is relative to its maximum potential connectivity.
Avg. Path Length	Represents the average communication distance in the network.
Clustering Coeff.	Indicates the presence of tightly connected groups or communities within the network.
Modularity	Quantifies the network's structure in terms of distinct, internally connected groups.
Num. of Self-loops	Indicates the presence of self-connections within the network.
Eccentricity	Represents the average distance from a node to all other nodes.
Avg. Eigenvector Centrality	Measures the influence of nodes based on the influence of their neighbors.
Avg. Num. of Neighbours	Describes the average level of local connectivity for nodes.
Centralization Betweenness	Indicates the concentration of control or communication pathways in the network.
Centralization Degree	Indicates the concentration of connections or links in the network.
Motifs-3, Motifs-4	Captures recurring substructures or motifs in the network.

production line, thus providing a more thorough monitoring view of production line failures.

VII. CONCLUSION

In this paper we presented a novel approach for assessing pharma regulatory compliance by adopting a graph network modelling of a pharmaceutical line process associated with a network analysis method which are analysed using a graph theory that would be applied to electronic batch as well as to traditional paper-based records. This approach simplifies the compliance procedure and thus its required processing burden, as it benefits graph structure, via checking only for network connectivity changes along ALCOA+ nodes, instead of managing vast amounts of word semantics and complex relationships that other conventional modelling approaches require, especially when dealing with scalable and complex pharma manufacturing production lines, with multiple batches and recipe combinations. In future work, we plan to provide a complete and automated implementation for ALCOA+-based regulatory assessment, by using a single network structure. We are also intent to validate the proposed approach on various pharma industry conditions, and in particular in cases where the complexity of a production line is scaling up.

SUPPLEMENTARY INFORMATION

Below is the link to the electronic supplementary material.

Supplementary Figure

Supplementary File

DECLARATION

The authors have no competing interests to declare that are relevant to the content of this article.

Informed consent was obtained from all human participants before the start of the study.

DATA AVAILABILITY

The authors confirm that the data sets used in this study have been provided by SPuMoNI project industrial partners and are available from the corresponding author on reasonable request.

REFERENCES

[1] World Health Organisation. (2016). *WHO Guidance on Good Practices and Record Management Practices*. [Online]. Available: https://www.gmp-compliance.org/files/guidemgr/WHO_TRS_996_annex05.pdf

[2] WHO Drug Information. (2019). *Guideline on Data Integrity*. [Online]. Available: https://cdn.who.int/media/docs/default-source/medicines/norms-and-standards/guidelines/inspections/trs1033-annex4-guideline-on-data-integrity.pdf?sfvrsn=6218a4e6_4&download=true

[3] B. T. Zhong, L. Y. Ding, H. B. Luo, Y. Zhou, Y. Z. Hu, and H. M. Hu, "Ontology-based semantic modeling of regulation constraint for automated construction quality compliance checking," *Autom. Construct.*, vol. 28, pp. 58–70, Dec. 2012, doi: 10.1016/j.autcon.2012.06.006.

[4] K. Sapkota, A. Aldea, M. Younas, D. A. Duce, and R. Banares-Alcantara, "Automating the semantic mapping between regulatory guidelines and organizational processes," *Service Oriented Comput. Appl.*, vol. 10, no. 4, pp. 365–389, Dec. 2016, doi: 10.1007/s11761-016-0197-2.

[5] K. Sapkota, A. Aldea, D. A. Duce, M. Younas, and R. Bañares-Alcántara, "Towards semantic methodologies for automatic regulatory compliance support," in *Proc. 4th Workshop Ph.D. Students Inf. Knowl. Manage.* Scotland, U.K.: ACM, Oct. 2011, pp. 83–86, doi: 10.1145/2065003.2065021.

[6] L. Hailemariam and V. Venkatasubramanian, "Purdue ontology for pharmaceutical engineering—Part I. Conceptual framework," *J. Pharmaceutical Innov.*, vol. 5, no. 3, pp. 88–99, Oct. 2010, doi: 10.1007/s12247-010-9081-3.

[7] L. Hailemariam and V. Venkatasubramanian, "Purdue ontology for pharmaceutical engineering—Part II. Applications," *J. Pharmaceutical Innov.*, vol. 5, no. 4, pp. 139–146, Dec. 2010, doi: 10.1007/s12247-010-9091-1.

[8] J. Morbach, A. Wiesner, and W. Marquardt, "Onto CAPE 2.0—A (Re) usable ontology for computer-aided process engineering," in *Computer Aided Chemical Engineering*, vol. 25. Amsterdam, The Netherlands: Elsevier, 2008, pp. 991–996. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S157079460880171X

[9] M. B. Sesen, P. Suresh, R. Banares-Alcantara, and V. Venkatasubramanian, "An ontological framework for automated regulatory compliance in pharmaceutical manufacturing," *Comput. Chem. Eng.*, vol. 34, no. 7, pp. 1155–1169, Jul. 2010, doi: 10.1016/j.compchemeng.2009.09.004.

[10] A. Yunianta, A. Hoirul, A. Satria, A. Bramantoro, I. Syamsuddin, N. Yusof, A. Omran, and K. Alsubhi, "OntoDI: The methodology for ontology development on data integration," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 1, 2019, doi: 10.14569/ijacsa.2019.0100121.

[11] V. Venkatasubramanian, C. Zhao, G. Joglekar, A. Jain, L. Hailemariam, P. Suresh, P. Akkisetty, K. Morris, and G. V. Reklaitis, "Ontological informatics infrastructure for pharmaceutical product development and manufacturing," *Comput. Chem. Eng.*, vol. 30, nos. 10–12, pp. 1482–1496, Sep. 2006, doi: 10.1016/j.compchemeng.2006.05.036.

[12] H. Alosert, J. Savery, J. Rheaume, M. Cheeks, R. Turner, C. Spencer, S. S. Farid, and S. Goldrick, "Data integrity within the biopharmaceutical sector in the era of Industry 4.0," *Biotechnol. J.*, vol. 17, no. 6, Jun. 2022, Art. no. 2100609, doi: 10.1002/biot.202100609.

[13] R. S. Patil, S. B. Kulkarni, and V. L. Gaikwad, "Artificial intelligence in pharmaceutical regulatory affairs," *Drug Discovery Today*, vol. 28, no. 9, Sep. 2023, Art. no. 103700, doi: 10.1016/j.drudis.2023.103700.

[14] I. Kavasidis, E. Lallas, H. C. Leligkou, G. Oikonomidis, D. Karydas, V. C. Gerogiannis, and A. Karageorgos, "Deep transformers for computing and predicting ALCOA+data integrity compliance in the pharmaceutical industry," *Appl. Sci.*, vol. 13, no. 13, p. 7616, Jun. 2023, doi: 10.3390/app13137616.

[15] Y. Chen, O. Yang, C. Sampat, P. Bhalode, R. Ramachandran, and M. Ierapetritou, "Digital twins in pharmaceutical and biopharmaceutical manufacturing: A literature review," *Processes*, vol. 8, no. 9, p. 1088, Sep. 2020, doi: 10.3390/pr8091088.

[16] C. Castaldello, P. Facco, F. Bezzo, C. Georgakis, and M. Barolo, "Data-driven tools for the optimization of a pharmaceutical process through its knowledge-driven model," *AIChE J.*, vol. 69, no. 4, Apr. 2023, Art. no. e17925, doi: 10.1002/aic.17925.

[17] E. N. Lallas, I. Santouridis, G. Mountzouris, V. C. Gerogiannis, and A. Karageorgos, "An SQWRL-based method for assessing regulatory compliance in the pharmaceutical industry," *Appl. Sci.*, vol. 12, no. 21, p. 10923, Oct. 2022, doi: 10.3390/app122110923.

[18] T. Funke and T. Becker, "Complex networks of material flow in manufacturing and logistics: Modeling, analysis, and prediction using stochastic block models," *J. Manuf. Syst.*, vol. 56, pp. 296–311, Jul. 2020, doi: 10.1016/j.jmsy.2020.06.015.

[19] T. Becker, M. Meyer, and K. Windt, "A manufacturing systems network model for the evaluation of complex manufacturing systems," *Int. J. Productiv. Perform. Manage.*, vol. 63, no. 3, pp. 324–340, Apr. 2014, doi: 10.1108/ijppm-03-2013-0047.

[20] H. Blunck, V. Vican, T. Becker, and K. Windt, "Improvement heuristics for manufacturing system design using complex network figures," *Proc. CIRP*, vol. 17, pp. 50–55, Jan. 2014, doi: 10.1016/j.procir.2014.01.063.

[21] S. Orchard, "Molecular interaction databases," *Proteomics*, vol. 12, no. 10, pp. 1656–1662, May 2012, doi: 10.1002/pmic.201100484.

[22] H.-J. Park and K. Friston, "Structural and functional brain networks: From connections to cognition," *Science*, vol. 342, no. 6158, Nov. 2013, Art. no. 1238411, doi: 10.1126/science.1238411.

[23] L. Peel, T. P. Peixoto, and M. De Domenico, "Statistical inference links data and theory in network science," *Nature Commun.*, vol. 13, no. 1, p. 6794, Nov. 2022, doi: 10.1038/s41467-022-34267-9.

[24] J. L. Gardy, D. J. Lynn, F. S. L. Brinkman, and R. E. W. Hancock, "Enabling a systems biology approach to immunology: Focus on innate immunity," *Trends Immunology*, vol. 30, no. 6, pp. 249–262, Jun. 2009, doi: 10.1016/j.it.2009.03.009.

[25] G. A. Pavlopoulos, M. Secrier, C. N. Moschopoulos, T. G. Soldatos, S. Kossida, J. Aerts, R. Schneider, and P. G. Bagos, "Using graph theory to analyze biological networks," *BioData Mining*, vol. 4, no. 1, p. 10, Dec. 2011, doi: 10.1186/1756-0381-4-10.

[26] A.-L. Barabási, N. Gulbahce, and J. Loscalzo, "Network medicine: A network-based approach to human disease," *Nature Rev. Genet.*, vol. 12, no. 1, pp. 56–68, Jan. 2011, doi: 10.1038/nrg2918.

[27] T. Charitou, P. I. Kontou, I. A. Tamposis, G. A. Pavlopoulos, G. G. Braliou, and P. G. Bagos, "Drug genetic associations with COVID-19 manifestations: A data mining and network biology approach," *Pharmacogenomics J.*, vol. 22, nos. 5–6, pp. 294–302, Dec. 2022, doi: 10.1038/s41397-022-00289-1.

[28] S. A. Kennedy et al., "Extensive rewiring of the EGFR network in colorectal cancer cells expressing transforming levels of $KRAS^{G13D}$," *Nature Commun.*, vol. 11, no. 1, p. 499, Jan. 2020, doi: 10.1038/s41467-019-14224-9.

[29] T. Charitou, K. Bryan, and D. J. Lynn, "Using biological networks to integrate, visualize and analyze genomics data," *Genet. Selection Evol.*, vol. 48, no. 1, p. 27, Dec. 2016, doi: 10.1186/s12711-016-0205-1.

[30] T. Becker and D. Wagner-Kampik, "Complex networks in manufacturing and logistics: A retrospect," in *Dynamics in Logistics*, M. Freitag, H. Kotzab, and N. Megow, Eds. Cham, Switzerland: Springer, 2021, pp. 57–70, doi: 10.1007/978-3-030-88662-2_3.

[31] A. M. Chmiel, J. Sienkiewicz, K. Suchecki, and J. A. Hołyst, "Weighted networks at the Polish market," in *Econophysics of Markets and Business Networks*, A. Chatterjee and B. K. Chakrabarti, Eds. Milan, Italy: Springer, 2007, pp. 127–138, doi: 10.1007/978-88-470-0665-2_9.

[32] V. Batagelj, N. Kejžar, and S. Korenjak-Černe, "Analysis of the customers' choice networks: An application on Amazon books and CDs data," *Adv. Methodol. Statist.*, vol. 4, no. 2, Jul. 2007, doi: 10.51936/ugam9616.

[33] D.-F. Zhang, Q.-S. Gao, and Z.-F. Li, "Critical quality chain analysis and evaluation based on quality loss in service industry," in *Proc. Chin. Control Decis. Conf.*, Shandong, China, Jul. 2008, pp. 1588–1592, doi: 10.1109/CCDC.2008.4597585.

[34] S. Hu and B. Dong, "A multidimensional data flow driven-based quality fluctuation evaluation for manufacturing process," *Math. Problems Eng.*, vol. 2021, pp. 1–9, Apr. 2021, doi: 10.1155/2021/5514056.

[35] H. K. Kim, J. K. Kim, and Q. Y. Chen, "A product network analysis for extending the market basket analysis," *Expert Syst. Appl.*, vol. 39, no. 8, pp. 7403–7410, Jun. 2012, doi: 10.1016/j.eswa.2012.01.066.

[36] T. Palo and J. Tähtinen, "A network perspective on business models for emerging technology-based services," *J. Bus. Ind. Marketing*, vol. 26, no. 5, pp. 377–388, Jun. 2011, doi: 10.1108/08858621111144433.

[37] P. Tsankov, "Overview of network-based methods for analyzing financial markets," *Proc. Tech. Univ. Sofia*, vol. 71, no. 1, Mar. 2021, doi: 10.47978/tus.2021.71.01.01.

[38] M. E. Beber and T. Becker, "Towards an understanding of the relation between topological characteristics and dynamic behavior in manufacturing networks," *Proc. CIRP*, vol. 19, pp. 21–26, Jan. 2014, doi: 10.1016/j.procir.2014.05.005.

[39] R. Vrabič, G. Škulj, and P. Butala, "Anomaly detection in shop floor material flow: A network theory approach," *CIRP Ann.*, vol. 62, no. 1, pp. 487–490, 2013, doi: 10.1016/j.cirp.2013.03.131.

[40] A. S. Mata and S. C. Ferreira, "Multiple transitions of the susceptible-infected-susceptible epidemic model on complex networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 91, no. 1, Jan. 2015, Art. no. 012816, doi: 10.1103/physreve.91.012816.

[41] M. E. J. Newman, "Spread of epidemic disease on networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 66, no. 1, Jul. 2002, Art. no. 016128, doi: 10.1103/physreve.66.016128.

[42] C. Dubé, C. Ribble, D. Kelton, and B. McNab, "A review of network analysis terminology and its application to foot-and-mouth disease modelling and policy development," *Transboundary Emerg. Diseases*, vol. 56, no. 3, pp. 73–85, Apr. 2009, doi: 10.1111/j.1865-1682.2008.01064.x.

[43] G. Rossi, R. L. Smith, S. Pongolini, and L. Bolzoni, "Modelling farm-to-farm disease transmission through personnel movements: From visits to contacts, and back," *Sci. Rep.*, vol. 7, no. 1, p. 2375, May 2017, doi: 10.1038/s41598-017-02567-6.

[44] Q. Liu, T. Li, and M. Sun, "The analysis of an *SEIR* rumor propagation model on heterogeneous network," *Phys. A, Stat. Mech. Appl.*, vol. 469, pp. 372–380, Mar. 2017, doi: 10.1016/j.physa.2016.11.067.

[45] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili, "Theory of rumour spreading in complex social networks," *Phys. A, Stat. Mech. Appl.*, vol. 374, no. 1, pp. 457–470, Jan. 2007, doi: 10.1016/j.physa.2006.07.017.

[46] N. Fujiwara, J. Kurths, and A. Díaz-Guilera, "Synchronization in networks of mobile oscillators," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 83, no. 2, Feb. 2011, Art. no. 025101, doi: 10.1103/physreve.83.025101.

[47] T. Peron, B. Messias F. de Resende, A. S. Mata, F. A. Rodrigues, and Y. Moreno, "Onset of synchronization of Kuramoto oscillators in scale-free networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 100, no. 4, Oct. 2019, Art. no. 042302, doi: 10.1103/physreve.100.042302.

[48] H. F. de Arruda, L. D. F. Costa, and D. R. Amancio, "Using complex networks for text classification: Discriminating informative and imaginative documents," *EPL Europhy. Letters*, vol. 113, no. 2, p. 28007, Jan. 2016, doi: 10.1209/0295-5075/113/28007.

[49] G. A. Wachs-Lopes and P. S. Rodrigues, "Analyzing natural human language from the point of view of dynamic of a complex network," *Expert Syst. Appl.*, vol. 45, pp. 8–22, Mar. 2016, doi: 10.1016/j.eswa.2015.09.020.

[50] E. H. Klijn, "Analyzing and managing policy processes in complex networks: A theoretical examination of the concept policy network and its problems," *Admin. Soc.*, vol. 28, no. 1, pp. 90–119, May 1996, doi: 10.1177/009539979602800104.

[51] T. Theodosiou, G. Efstathiou, N. Papanikolaou, N. C. Kyrpides, P. G. Bagos, I. Iliopoulos, and G. A. Pavlopoulos, "NAP: The network analysis profiler, a web tool for easier topological analysis and comparison of medium-scale biological networks," *BMC Res. Notes*, vol. 10, no. 1, p. 278, Dec. 2017, doi: 10.1186/s13104-017-2607-8.

THEODOSIA CHARITOU is currently a Postdoctoral Research Fellow with the Department of Computer Science and Biomedical Informatics, University of Thessaly, where she is also a member of the Applied Informatics and Digital Technologies (AIDigiLab) Laboratory. She is also an Adjunct Lecturer in computer science and bioinformatics with the University of Thessaly. Her previous work experience includes research position in several International Health Institutes, USA, Australia, and Europe, where her main role was focusing on implementing computational approaches to study complex diseases. Her scientific research interests include bio-medical data processing using machine and deep learning methods and studying whole genome using next generation sequencing technologies and big data analysis with bioinformatics and network science.

EFTHIMIOS LALLAS is currently an Associate Professor with the School of Technology, University of Thessaly, Greece, where he is also a member of the Applied Informatics and Digital Technologies (AIDigiLab) Laboratory. His previous work experience includes key job positions as the Core Network Deputy Director of HUAWEI and participations on several major European Research Programs. He is the author of many scientific papers and monographs in international conferences and journals and a peer reviewer as well. His scientific research interests include network system technologies (broadband and optical wireless-sensor IoT systems), network on chip architectures (NoC), network computing system design and techno economical evaluation, network system security and blockchain, cognitive devices and systems, and intelligent home networks.

VASSILIS C. GEROGIANNIS received the Diploma degree in computer engineering and the Ph.D. degree in software engineering from the University of Patras, Greece. He is currently a Professor and the Head of the Department of Digital Systems, University of Thessaly (academic subject: analysis, design of systems, and projects with emphasis on decision making). He is also an Adjunct Professor with Hellenic Open University. He is also a Visiting Professor with the IPAG Business School, France, and Siauliai State University of Applied Sciences, Lithuania. Since 1992, he has been participating in several projects. He is the author/coauthor of more than 130 papers published in international journals/conference proceedings and cited in a plethora of citations. He is also the coauthor/editor of three scientific books. He is a member of editorial board, a guest editor, and a reviewer of international journals. He is also the conference chair/program chair and an invited speaker of international conferences. He has received the "best paper award" in three international conferences. He is also a member of the Management Board of the Hellenic National Academic Recognition Information Centre (NARIC), the Council for Research, Innovation in Thessaly; the Management Committee of the Entrepreneurship Innovation Research Institute of the Research Center IASON, University of Thessaly; the Management Committee of the Technical Chamber Central and Western Greece; the Central Assembly of the Technical Chamber of Greece; and the Scientific Committee of Electronics Engineers in the Technical Chamber of Greece.

ANTHONY KARAGEORGOS received the degree in applied mathematics and in computer science and the Ph.D. degree in agent-based software engineering. He is currently a Professor in intelligent information systems with the School of Technology, University of Thessaly, Greece, and the Head of the Applied Informatics and Digital Technologies (AIDigiLab) Laboratory. Among his main research involvements in AIDig-iLab are the development of intelligent systems for personalized e-commerce and mass manufacturing. He has received funding for research projects in Greece and EU, which resulted in patents and several international journals and conference publications. He has extensive experience in self-organizing systems and decentralized intelligence. His current research interests include combining immersive reality and decentralized intelligence, self-organizing and intelligent systems, and the IoT and data analytics.