



UNIVERSITÀ POLITECNICA DELLE MARCHE  
Repository ISTITUZIONALE

Association rules and social network analysis for supporting failure mode effects and criticality analysis:  
Framework development and insights from an onshore platform

This is the peer reviewed version of the following article:

*Original*

Association rules and social network analysis for supporting failure mode effects and criticality analysis:  
Framework development and insights from an onshore platform / Antomarioni, S.; Ciarapica, F. E.;  
Bevilacqua, M.. - In: SAFETY SCIENCE. - ISSN 0925-7535. - 150:(2022). [10.1016/j.ssci.2022.105711]

*Availability:*

This version is available at: 11566/297562 since: 2024-11-05T10:23:15Z

*Publisher:*

*Published*

DOI:10.1016/j.ssci.2022.105711

*Terms of use:*

The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. The use of copyrighted works requires the consent of the rights' holder (author or publisher). Works made available under a Creative Commons license or a Publisher's custom-made license can be used according to the terms and conditions contained therein. See editor's website for further information and terms and conditions.

This item was downloaded from IRIS Università Politecnica delle Marche (<https://iris.univpm.it>). When citing, please refer to the published version.

(Article begins on next page)

# **Association rules and social network analysis for supporting failure mode effects and criticality analysis: framework development and insights from an onshore platform**

Antomarioni, Sara<sup>1\*</sup>, Ciarapica, FilippoEmanuele<sup>1</sup>,Bevilacqua Maurizio<sup>1</sup>

<sup>1</sup>Dipartimento Ingegneria Industriale e Scienze Matematiche, Università Politecnica delle Marche, Via Brecce Bianche, 60131 Ancona (Italy)

## **Abstract**

The ongoing digital transformation enables the collection of a vast amount of data from the maintenance and production processes. Hence, the techniques traditionally applied for monitoring the operations and defining the maintenance approaches for improving the asset reliability can be accompanied by new methodologies more oriented to data analytics. In this context, this work proposes a data-driven framework for supporting the analysis of the production processes in terms of failures and related effects through the well-known and widely applied Failure Mode, Effect, and Criticality Analysis (FMECA). Indeed, after developing the Failure Mode Effects and Criticality Analysis, the results obtained for the plant under investigation are elaborated through Machine Learning techniques such as the Association Rule Mining to define the cause-effect relationships that led to a failure. The association rules extracted are then processed through the Social Network Analysis to represent such relationships, facilitate their comprehension, and identify the existence of communities of nodes to detect critical patterns and locate the most influential nodes. The knowledge of such details can provide helpful support in terms of awareness of the plant and the development of intelligent maintenance procedures.

The proposed approach is applied to an off-shore and on-shore platform to assess the impact of the theoretical analysis on the practical implementation by highlighting unknown relations among the analyzed variables and showing new cause-effect relationships.

**KEYWORDS:** Failure Mode Effects and Criticality Analysis; Risk-Based Maintenance; Association Rule Mining; Machine Learning; Social Network Analysis; Data-driven approach

## Nomenclature

<b>Acronym</b>	<b>Meaning</b>
FMECA	Failure Mode, Effect and Criticality Analysis
ARM	Association Rule Mining
SNA	Social Network Analysis
AR	Association Rule
SN	Social Network
OD	Out-degree
CFM	Criticality of the Failure Mode

<b>Set</b>	<b>Meaning</b>
$K$	Set of data
$T$	Set of transactions
$I, J$	Itemsets
$N$	Set of nodes
$E$	Set of Edges

<b>Parameter</b>	<b>Meaning</b>
$\lambda$	Failure rate
$w_i$	Weight associated with the $i$ -th edge
$n$	Number of edges outgoing from a node
$j$	Generic node

## 1. Introduction

Although failure mode, effect, and criticality analysis (FMECA) is a helpful and popular instrument for safety and reliability evaluation, it suffers from many shortcomings while employed in the company (Huang et al., 2020). In many industrial environments, characterized by several machines and equipment, taking control of the vast amount of data collected by Enterprise Information System, Maintenance & Operation Management system, and Plants Control System, can be critical (Marcucci et al., 2021). This industrial scenario well fits the application of Knowledge Discovery in Databases techniques since they aim to support the decision-making processes through the automatic extraction of valuable, unknown, and valid relationships among data.

Knowledge Discovery in Databases can be used for extending the analysis of failure modes and effects since it represents a prominent enhancement opportunity of the maintenance policy improvement and risk reduction through the application of advanced techniques, expanding the knowledge achievable through the traditional approaches (Crespo Márquez et al., 2019). In this context, the current work proposes a framework for extending the analysis of the production processes in terms of failures and related effects through the well-known and widely applied Failure Mode, Effect, and Criticality Analysis (FMECA).

Strategic data-driven FMECA has been addressed through several techniques (e.g., Bevilacqua, Ciarapica, & Mazzuto, 2018; Carpitella, Certa, Izquierdo, & La Fata, 2018; Luo & Lee, 2015). Some of these models are used to automate the identification of failure modes (e.g., Grunske et al., 2005; Tso et al., 2005), others are used to improve the risk assessment process (Ben Said et al., 2016; Viveros et al., 2018), while some integrate FMECA and remaining useful life prognosis (Aggogeri et al., 2020; Ragab et al., 2019). Less attention has been paid to developing data-driven frameworks for supporting failure analysis. Although the existing research is extremely relevant, a framework based on data analytics that analyzes and predicts the structural relationships among failure modes and effects, supporting the definition of the maintenance policy, is not present in the current literature.

The data-driven framework proposed aims to address the identified research gap by introducing an innovative procedure to support the data analysis for enlarging the current body of knowledge, facilitating the visualization, and, thus, the understanding of previously unknown paths. In particular, the development of the framework relies on the simultaneous adoption of two data-driven techniques for analyzing the output data of the FMECA: (i) the Association Rule Mining (ARM) and (ii) the Social Network Analysis (SNA). The main contribution deriving from introducing these tools can be summarized as follows: managers are supported in evaluating the hidden chain reaction of causes behind the most important failures that can harm the whole plant from a single trigger event. The ARM is applied to identify the relationships between failure modes, their effects on the system, and the adopted maintenance tasks. The rationale behind the adoption of ARM is based

on their intuitiveness, allowing their interpretation even for non-domain experts (Sara Antomarioni et al., 2021). Moreover, ARM does not require the formulation of hypotheses, allowing an unbiased analysis of the whole dataset. In this way, relationships that are not identifiable through the FMECA itself can emerge. In parallel, the SNA enables company managers to jointly explain the Association Rules (ARs) through a graphical representation of the results: the network provides a more understandable data format for the decision-makers. The SNA supports the definition of the nature of the interactions represented by the ARs, identifying communities of nodes to analyze critical patterns and locate influential nodes. Moreover, through analysis of the causal relationships that this methodology highlights, the domino effect affecting the factors that influence plant reliability and availability can be analyzed step by step.

From a practical point of view, the main novelty of this proposed approach is to support the operations and maintenance managers in standardizing the maintenance policy definition process. Indeed, it is possible to identify the failure modes having similar effects and foresee adopting a unique policy. This issue is particularly relevant in process plants with a growing focus on defining ad-hoc maintenance frameworks to execute the operations safely. Process plants are often characterized by aging assets and a lack of investments in renewals, but, in parallel, the reliability of these assets plays a vital role in avoiding catastrophic events. This criticality is particularly detectable in the case of installations characterized by high hazards, in which a failure has potentially catastrophic consequences not only on the machinery but also on people and environmental safety, as highlighted by Simonoff et al. (2010) and Wang (2002). For this reason, the validation of the proposed framework is presented through the case study of an offshore and on-shore plant for oil & gas extraction and storage to clarify the explanation of the proposed approach and to study the impact of the theoretical analysis on its practical implementation. This plants' typology is worth analyzing because the risk events may have severe consequences for people, assets, and the environment.

In the following, a summary of the rest of the paper is provided: Section 2 contains the literature review regarding data-driven approaches applied to failure analysis; Section 3 deploys the research approach

proposed in this study, while Section 4 presents the case study. Sections 5 and 6 discuss the results and conclude.

## **2. Data-driven approaches to failure analysis**

### **2.1 Data-driven and FMECA**

FMECA analysis is widely applied in the maintenance management field. Some applications involving its joint implementation, together with data-driven techniques, can be found in literature. For instance, in Savino et al. (2011), the fuzzy inference is used with the FMECA to perform the criticality evaluation considering safety aspects and production performance. In Tso et al. (2005), the automatic identification of the failure modes to perform an efficient FMECA is addressed through a framework based on hardware description languages and knowledge-based fault models. The automation of the FMECA has also been addressed in Grunske et al. (2005) through the implementation of behavior trees that support the failure mode identification by injecting faults data. In the attempt to identify all the possible relationships among components and failure modes, in Xu et al. (2020), text mining is applied to identify the potential failure modes related to a specific component. Data regarding the remaining useful life of components can be adequately analyzed through data mining techniques and used to update the FMECA data, to monitor the modification of the risk of failure, and to be used in the following projects (Aggogeri et al., 2020; Ragab et al., 2019). Bayesian Networks have been successfully applied with the FMECA, as presented by L. Liu et al. (2019) and Ben Said et al. (2016). Some approaches for improving the prioritization of components failure risk have also been proposed. For instance, in Chang & Cheng (2011), who apply the fuzzy ordered weighting average and the DEMATEL methods are used for calculating the risk assessment, and H. C. Liu et al. (2016), where the risk assessment evaluation is performed through the definition of fuzzy digraph and matrix. While, in Khorshidi et al. (2016), the overall failure index is used to guide the optimal selection of the improvement actions to implement to achieve maximum system reliability. As shown in Li et al. (2019) and Lv et al. (2018), the domain experts' judgments on failure modes can be included

in a cloud model so that they can be compared through multi-criteria decision-making approaches. In Ma et al. (2019), instead, the quality function deployment and the FMECA are integrated to analyze improvement areas of the components, taking into account the reliability aspects and customer expectations.

## 2.2 Association Rule Mining applications to the operations and maintenance field

Some authors apply rule-based techniques to measure the associations between failures and predictor events to analyze event log files (Yuan et al., 2011) or learn temporal associations from streams of events (Gunawardana et al., 2011). Other authors, instead, focus on developing fuzzy rule-based systems to related fault events and causes or effects (Renjith et al., 2018), while others focus on developing an expert system to guide the maintenance procedures (Mushiri et al., 2018). Several Association Rule Mining (ARM) applications in the operations and maintenance management field analyze breakages and malfunctioning or predict the need to maintain components (Harding et al., 2006). For instance, in Chen et al. (2005), the ARs are mined to detect the relationships among component defects and the machine causing them. Similarly, the detection of the sequences causing the breakages of components during the assembly can also be performed through the ARM to improve the quality of the product (Da Cunha et al., 2006). In Antomarioni et al. (2019), instead, the ARs are mined to find the components of an oil refinery frequently failing together, to replace them in advance. The approach proposed in Çakır et al. (2021) examine tugboat accidents using different association rule mining algorithms. In particular, Apriori, Predictive Apriori, and FP-Growth algorithms were employed to extract the association rules.

Often, ARM is applied in combination with other techniques to complete the predictive framework. For example, an ARs-based optimization model is adopted in Antomarioni et al. (2019) to select the optimal component set for the predictive replacement based on the relationships among failures. In Shasha Xiao et al. (2016), ARM and Bayesian networks are integrated for maintenance decision-making in the product-service system context. In Crespo Márquez et al. (2019), ARM is applied downstream to an Artificial Neural Network to

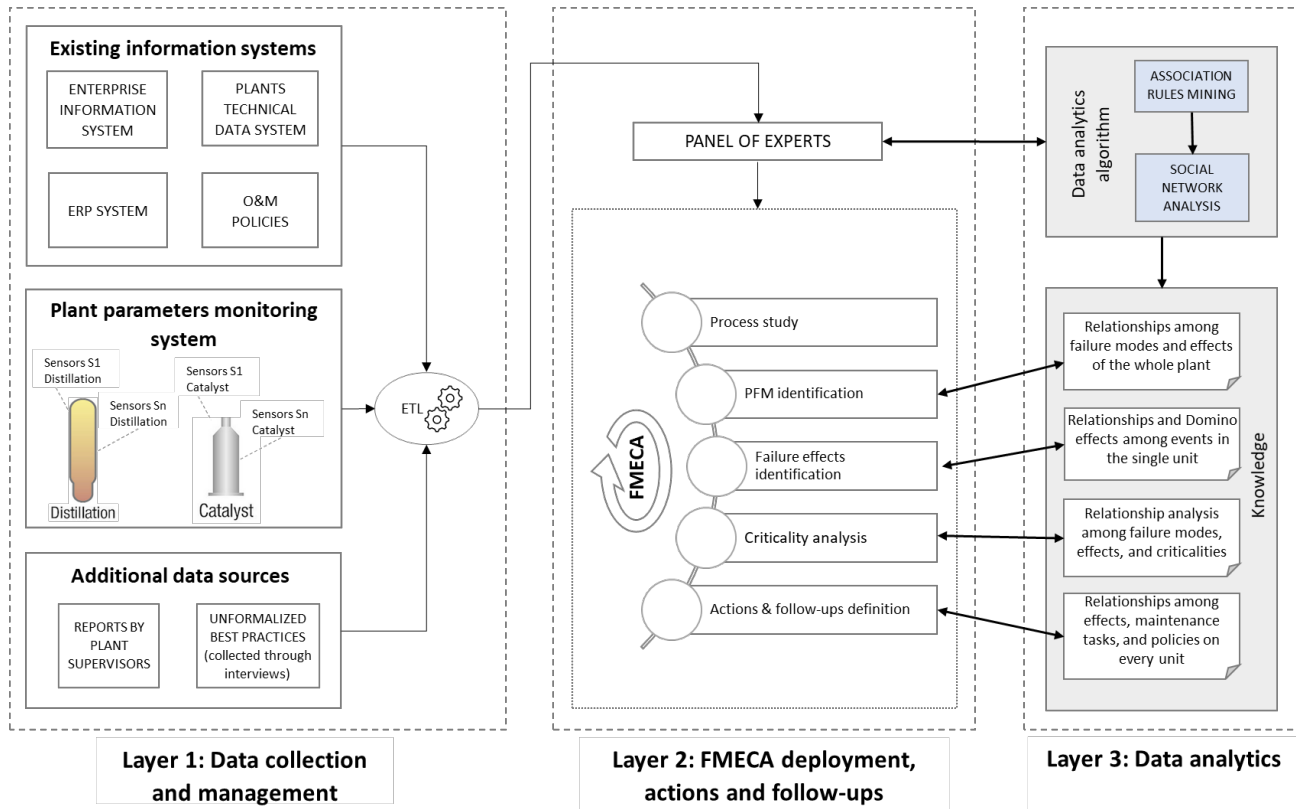
analyze the condition describing the anomaly detected. Recently, different authors used association rules to develop warning models. Jia and Zhang (2021) used association rules-based soft set theory to build an early warning model for mitigating maritime logistics service risks by analyzing the roles of information risk events in triggering other risks. Antonello et al. (2021) developed an Apriori algorithm for mining association rules for detecting functional dependencies among parts of complex technical infrastructures using databases of alarm messages. The proposed method is based on the representation of the alarm database by a binary matrix.

To the best of the authors' knowledge, the ARM application combined with the Social Network Analysis (SNA) in analyzing failure data is novel. Indeed, the latter methodology is not widely applied in the maintenance field since its aim is to investigate network's social structures through graph theory (Otte & Rousseau, 2002). Hence, SNA application mainly belongs to the sociological field (Scott, 1988). It has been applied, for instance, to study the information and knowledge flows in the construction project teams, to improve collaboration (Ruan et al., 2015), and to identify how, in the construction industry, safety communications flow within the local workers and ethnic minorities (Chan et al., 2014). In the same sector, the fatalities are analyzed through the SNA to identify common root causes (Eteifa & El-Adaway, 2018). The most effective application of SNA for the current analysis can be found in Kim et al. (2019), where the synchronous replacement of components driven by the life-cycle cost analysis is proposed. The authors use the SNA to visualize the link between the components to be replaced. Such links are defined after an interview with the maintenance experts. The combination of the ARs and SNA is firstly proposed for the study of environmental risk management as a framework for the control and improvement of the environmental performance of the company analyzed (Ciarapica et al., 2019).

As summarized before, several techniques can be applied to improve the FMECA process. However, these applications are more focused on improving the FMECA itself. Instead, the approach proposed in this work addresses the lack of a data-driven framework to support the definition of the maintenance strategies, considering the possible cascade effects related to hazardous events and their impact on plant reliability.

### **3. Research approach**

This section describes the framework proposed to analyze and manage plant failures. The framework aims to support the operations and maintenance managers in extending the knowledge obtained through traditionally applied techniques. Specifically, the framework is organized into three layers (see figure 1). Data collection and management system constitutes the first layer. The second layer consists of FMECA and action deployment. In this layer, the traditional methodologies for defining equipment inspection tasks and maintenance policies are developed. The outcomes of the second layer are taken as the input for a further layer through the description of situations frequently occurring together, namely the ARs among the failure modes, effects, and maintenance tasks or policies. This layer allows company managers to highlight unknown relations among the analyzed variables and show new cause-effect relationships. Such associations are then structured through the SNA to provide a clearer understanding of the co-occurrences and identify communities of nodes to define critical patterns and locate influential nodes. According to a supervised approach, the knowledge from the third layer will be feedback for the previous layers.



**Figure 1 Schematization of the proposed framework**

### 3.1 Layer 1: Data Collection and Management

The data collection layer involves retrieving all the input data necessary for the correct execution of the FMECA. Specifically, plant characteristics, signal records, protocols, and best practices are required to integrate the knowledge of plant features and past events. In addition, the experience of the workers could be helpful. Indeed, together with the defined protocols of the company, the unformalized practices implemented by expert technicians represent a fundamental data source to perform the analysis exhaustively. Furthermore, the quality of the study relies on this first step of the procedure. Indeed, an unreliable dataset used for the decision-making process might lead to inconsistent decisions. Multiple sources of data may have to be consulted in order to create an integrated dataset to perform the following steps of the procedure. Such data might be structured or unstructured, so pre-processing is needed to transform them into a suitable format

(e.g., structure free-text annotations). In addition, a missing data policy should be implemented to resolve all the issues related to unrecorded data (e.g., comparison with the other data sources, exclusion of the missing instance, replacement with mean values). The activities performed at this stage are conventionally named Extraction, Transformation, and Loading (ETL) process: ETL is necessary to obtain an integrated dataset that can be analyzed in the downstream phases of the approach.

### 3.2 Layer 2: FMECA deployment, actions, and follow-ups

In the second layer, the data collected in the previous one are analyzed with traditional methods to extract valuable information about failure mode, risk events, and maintenance policies. The FMECA is performed at this stage, following the recommendation provided in the US Military Standard (US Military Standard, 1980, 1983). A bottom-up approach is adopted for its execution, breaking the analyzed system down to identify its elementary components (e.g., parts or sub-systems). The objective of this step is to accurately describe the failure modes, effects, and criticality. The approach followed to carry out the FMECA is collaborative. It involves discussing the system's main features among the interdisciplinary groups of people engaged in its functioning at different employment levels (e.g., maintenance engineers, managers, technicians). In this way, several perspectives are taken into account, and a complete understanding of the system is guaranteed.

The output of this step is a dataset containing the components of the system, the potential failure modes, effects, frequency, severity, and a measure of the criticality; this information is used to define actions and follow-ups. Indeed, identifying critical components aims to guide the actions to undertake and anticipate failures or correct them. In this way, the components that should be strictly monitored are determined to define specific inspection policies and roadmaps for the corrective interventions. These analyses are made by maintenance experts that are already aware of company policies to make the action and follow-ups feasible and coherent with the company maintenance strategy. These outcomes should be included in the dataset that will be analyzed in the following steps of the procedure to enlarge the possible identifiable connections.

### 3.3 Layer 3: Data Analytics

The third step of the roadmap aims to deepen the knowledge acquired from the FMECA through the ARM and SNA. The former aims at identifying previously unknown relationships among the dataset built during the FMECA. A complete visualization of the results is provided through the latter one, which allows the representation of relations (arcs) among nodes, i.e., failure modes, effects, and criticalities. In this way, the understanding of the results is facilitated since the relationships extracted with the ARM are represented through the graph theory. The information extracted from this framework step helps identify improvement strategies to the actions and follow-ups defined in the study's second step and produce new insights for further investigation. In detail, the roadmap followed for enhancing the FMECA results can be summarized as follows:

- Relationship analysis among failure modes, effects, and criticalities in the whole plant: in this way, the network of PFM, effects, and criticalities support in identifying the riskiest events and, hence, the riskiest areas to be thoroughly monitored;
- Relationships and domino effects among potential failure modes and effects of the single unit: the most critical areas of the plant identified at the previous stage are addressed so that the domino effect among PFM and effects can be identified at a more specific level ;
- Relationship analysis among failure modes, effects, and criticalities, to take into account the most critical PFM and their effect in the attempt to define a priority scale for the interventions;
- Relationships among effects, maintenance tasks, and policies on every unit to highlight repetitive relationships among events and adjust the maintenance practices accordingly.

In order to identify whether approaches existing in literature could suit the addressed issue, the proposed approach has been compared to two recent implementations already existing in literature and based on rules generation. Specifically, the fuzzy rule-based FMECA developed in Renjith et al. (2018) and the one proposed in Mushiri et al. (2018) are used as a comparison. The former generates fuzzy rules to improve faults

prioritization and Risk Priority number definition in complex process systems, even though probability measures of such relationships are neither identified nor graphically displayed. The latter, instead, integrates the currently employed tools for preventive maintenance (i.e., FMECA and Ishikawa diagrams) with an expert system that, if appropriately filled out, provides the user with a set of rules indicating the possible actions to carry out to prevent the actual fault of the system. Even in this case, there is no representation of such relationships, nor the possibility to analyze the failure spreading.

**Table 1 Comparison between the proposed approach and rule-based approaches in the literature**

<b>Criteria</b>	<b>Proposed approach (FMECA+ARM+SNA)</b>	<b>Fuzzy rule-based FMECA (Renjith et al., 2018)</b>	<b>Expert system for preventive maintenance improvement (Mushiri et al., 2018)</b>
<b>Objective</b>	Evaluate the relationships between FMECA items, their probability, and spreading	Improve faults prioritization in complex process systems	Develop an expert system to improve the preventive maintenance
<b>Definition of rules/relationships</b>	Yes	Yes	Yes
<b>Probability measure of the relationships</b>	Yes	No	Yes
<b>Visualization of the relationships/communities of nodes</b>	Yes	No	No
<b>Failure mode criticality assessment</b>	Yes	Yes	Yes
<b>Risk Priority Number calculation improvement</b>	No	Yes	No

In the following subparagraphs, the basics of ARM and SNA are provided.

### *3.3.1 Association Rule Mining*

The first application of the Association Rule Mining (ARM) involves the extraction of hidden patterns for marketing scopes from large datasets (Agrawal et al., 1993). Since then, it has spread across different fields,

like operations and production-related ones. A formal definition of the association rules and their mining is the following one:

$K=\{k_1, k_2, \dots, k_n\}$  is a set of items;  $T=\{t_1, t_2, \dots, t_m\}$  is the transaction set, each of whom is composed by an itemset, taken from  $K$ . An Association Rule is an implication  $I \rightarrow J$  such that  $I$  and  $J$  are itemsets ( $I, J \subseteq K$ ) and  $I \cap J = \emptyset$ . The itemset  $I$  is called “body” of the rule, while  $J$  is the “head”. The most common metrics for the evaluation rule quality are the support (1) and the confidence (2) (Agrawal et al., 1993). The support of the rule the itemsets  $I$  and  $J$  belonging,  $I \rightarrow J$ , is determined as:

$$Support \{I \cup J\} = \frac{\#\{I \cup J\}}{\#\{T\}} \quad (1)$$

where the function  $\#\{x\}$  indicates the cardinality of the itemset  $x$ , while  $T$  is the transaction set; hence, we can say that the support describes the number of itemsets taken from the transaction set in which both  $I$  and  $J$  appear. In other words, the support provides the joint probability of having both  $I$  and  $J$  in the same rule. The confidence, instead, is given by:

$$Confidence \{I \rightarrow J\} = \frac{Support \{I \cup J\}}{Support \{I \rightarrow true\}} \quad (2)$$

The expression  $Support \{I \rightarrow true\}$  represents all the ARs having  $I$  as body and any other possible itemset as head of the rule. Equation (2) indicates the conditional probability of having the itemset  $J$  in a transaction because it already contains  $I$ . Hence, it measures the strength of the rule.

Several algorithms for ARM exist; in the current application, the FP-growth (Han et al., 2000) is proposed since it is proven to be more efficient than other ones (e.g., Apriori, Eclat) (S Xiao et al., 2016). Scanning of  $T=\{t_1, t_2, \dots, t_m\}$  is required to determine the items that appear more frequently than the minimum support ( $min\_sup$ ) threshold, that is, a user-defined parameter. It is excluded from the analysis if an itemset does not meet the  $min\_sup$  requirement. The mining of the ARs is schematized in Table 2, also indicating the input and output considered in this work.

The definition of the min\_sup and min\_conf threshold impacts the number of rules generated: the higher the threshold, the lower the number of rules mined. In fact, setting a high threshold implies the loss of all the rules presenting a lower probability of occurrence, both in terms of support (joint probability) and confidence (conditional probability). On the other hand, a low threshold may cause the mining of many ARs, making it impossible to consider them appropriately. Thus, it is important to set a trade-off between considering the most likely events and adequately analyzing the relationships extracted.

**Table 2 Schematization of the Association Rule Mining**

<b>Input:</b>	FMECA dataset
<b>Output:</b>	Association among failure modes, effects, and criticalities frequently occurring contextually
<b>Procedure:</b>	<ol style="list-style-type: none"> <li>1: Definition of the min_sup: minimum support threshold to consider a rule</li> <li>2: Definition of min_conf: minimum confidence threshold to consider a rule</li> <li>3: Apply the FP-growth algorithm (Han et al., 2000) to select the frequent itemsets;</li> <li>4: Combination of pairs of frequent itemsets creating the ARs; delete rules not meeting the min_conf threshold.</li> </ol>

The relevant ARs are then represented through the SNA in order to facilitate the interpretation of results.

### 3.3.2 Social Network Analysis

The SNA is a methodology applied to investigate network's social structures through graph (Otte & Rousseau, 2002). The network is described by an ordered pair of nodes connected by edges,  $G=(N, E)$ , where N stands for nodes, while E stands for edges. Traditionally, the SNA is applied to study the interactions, i.e., the edges, among a set of actors (i.e., the entities object of the analysis), represented by nodes. In the current work, the SNA aims to display the events (e.g., failure modes and effects) and the related criticalities that occur concurrently with high frequency to support the ARs interpretation. The patterns that should be considered are immediately visualized.

In the proposed approach, the actors are the attributes of the FMECA, e.g., the components analyzed, the failure modes, the effects, the criticality measures, etc. Their interactions are described by the ARs and the

strength of such interactions. Indeed, the confidence values represent the relevance of interaction since it measures the probability of occurrence of the relation itself.

**Table 3 Schematization of the Social Network Analysis**

<b>Input:</b>	Frequent itemsets Association rules
<b>Output:</b>	Social networks showing Associations among failure modes, effects, and criticalities frequently occurring contextually
<b>Procedure:</b>	<ol style="list-style-type: none"> <li>1: Define the nodes of the network: Frequent Itemsets are used as nodes</li> <li>2: Define the edges of the network: the association rules among the frequent itemsets are used as edges</li> <li>3: Calculate the out-degree (<math>OD_j</math>) (Knoke et al., 2008) of each node and sort them by descending <math>OD_j</math></li> <li>4: Focus the attention on the items characterized by a high <math>OD_j</math></li> </ol>

For example, if node  $i$  is connected to node  $j$  ( $i \rightarrow j$ ), i.e., the rule  $i \rightarrow j$  is mined, it means that when the event  $i$  occurs, even the event  $j$  occurs as well. The probability of conditional events is given by the rule's confidence, used as the weight of the arc. Once the SN is built, in order to define which of the nodes might be more critical and arrange the actions and follow-ups accordingly, the Out-Degree ( $OD$ ) metric is applied: it is calculated for each node as the weighted sum of the outgoing arcs (Knoke et al., 2008); the higher the  $OD$ , the higher the strength of the node on its successors. Thus, being  $j$  the node for which the  $OD$  is determined, having  $n$  outgoing arcs, each one characterized by a weight  $w_i$ , the  $OD$  is calculated as reported in equation 3:

$$OD_j = \sum_{i=1}^n w_i \quad (3)$$

A schematization of the Social Network Analysis process used in this work is provided in Table 3.

## 4. Case study

The research approach proposed in Section 3 is applied to a case study to explain its application. The process plants analyzed, both offshore and on-shore, consist of one platform and eight wellheads, six producing only gas and two producing oil and gas; a multi-phase marine transport line to the on-shore plant, about 40 km away; an on-shore gas processing plant, located near the power plant; a pipeline for liquid hydrocarbons, which exports them to the oil export terminal. The project under investigation aims to produce gaseous and condensed hydrocarbons from the offshore reservoir and make them available on-shore. The reservoir consists of eight wellheads (Unit 100), six of which produce only gas, exiting at a pressure of 72 bar per natural flow (oil wellheads). The remaining two produce oil and associated gas, which are initially separated offshore and then pumped and compressed at the same pressure as the gas from the wellheads. Subsequently, the gas flow produced,  $2.1 \cdot 10^6 \text{ Sm}^3/\text{day}$  from the six gas wellheads and  $73080 \text{ MSm}^3/\text{day}$  (associated gas) from the two oil wellheads is sent together with the oil to a multi-phase marine line (approximately 40 km-long, Unit 130), connecting the platform with the on-shore plant, where it is immediately delivered to a Slug Catcher for liquid-gas separation (Unit 200). This gas is then mixed with the one coming from the gas-flash compression unit (Unit 360). After this mixing, a Gas Pre-Heater is collocated where the gas flow is heated by heat exchange with stable condensate flow coming out from the bottom of the Stabilizer Column. If this heat recovery is not sufficient (or if the Gas Pre-Heater is not available for any reason), a Gas Heater is also activated, which eventually allows the total by-pass of the Gas Pre-Heater. Gas heating has the objective to respect the delivery temperature at the power station inlet. After passing through the Gas Metering Station, the gas, which is the primary process fluid, becomes the power supply of an on-shore power plant located onshore (Unit 470). It will be fed in such quantity to reach its full production, while the liquid hydrocarbons will be valorized through dedicated facilities and delivered to the condensate recovery tank and then to the export terminal.

## 4.1 Data collection

The input data are the equipment list and their characteristics, in which each item represents the plant components or parts of them, whose maintenance policies are under investigation. Moreover, to evaluate the frequency of the failure modes, it is necessary to consult the company's maintenance policies and the historical data of previous failures of the plant components and reliability databases (OREDA, EIREDA, or IEEE). The failure rate  $\lambda$  (i.e., number of failures per hour) for each failure mode of each piece of equipment is retrieved from them. Moreover, on-field reports are integrated into the historical data recorded in the information systems to create a spectrum of analysis as accurately as possible. In Table 4, a summary of the units, a brief description, the number of items analyzed, and the related failure modes are reported. In Table 5, instead, the list of all the failure modes and their frequency in the FMECA document is reported. In all, 501 items and 31 different failure modes referring to the 15 units are analyzed in this application. The FMECA is performed every three months to reconsider the outcomes based on the events verified during this time interval. In this way, the framework is continuously updated with the system's actual behavior. This case study refers to the analysis conducted on March 31st, 2021.

**Table 4 Summary of the units, items, and the related failure modes**

<b>Unit</b>	<b>Description</b>	<b>Number of items analyzed</b>	<b>Potential failure modes per unit</b>
Unit 100	Oil&Gas Production Wellhead	108	232
Unit 130	Production Flowlines	6	27
Unit 190	Launching Trap	23	15
Unit 200	Oil Production Separation	10	63
Unit 220	Crude Oil Transport	22	32
Unit 230	Flare and Blow-Down	12	58
Unit 360	Gas Compression	26	107
Unit 390	Glycol Injection Pumps	44	6
Unit 420	Fuel Gas	55	137
Unit 450 O	Oil Wellhead Control Panel	24	58
Unit 450 G	Gas Wellhead Control Panel	24	58
Unit 460	Air Compressor, Air Dryer and Filters and Receivers	88	199
Unit 470	Main Power Generation System	24	46

Unit 500	Sea Water System	20	42
Unit 550	Closed Drain System	15	44

**Table 5 List of the potential failure modes and their frequency**

<b>Failure modes</b>	<b>Frequency</b>
Abnormal instrument reading	8
Abnormal output - low	20
Breakdown	14
Delayed operation	7
Erratic output	64
External leakage	29
External leakage - Utility	9
External leakage process medium	134
Fail to close on demand	96
fail to function on demand	34
Fail to open on demand	152
Fail to regulate	98
Fail to start on demand	36
Fail to stop on demand	2
Insufficient heat transfer	6
Internal leakage	8
Loss of performance	13
Low output	42
Overheating	2
Parameter deviation	16
Plugged/choked	5
Spurious operation	38
Spurious stop	47
Structural deficiency	30
Valve leakage in closed position	46
Vibration	13

#### 4.2 FMECA deployment layer

A selection of a panel of experts made up of professional figures taking part in the FMECA process has been developed. Data have to be gathered from the on-field technicians who have a direct view of the processes and failures, as well as operations and maintenance managers who consider the process from a broader perspective, taking into account both the technical aspects related to single equipment and the main features of the whole process. Other experts of the maintenance team are necessary to globally extend the study to

machine, civil structure, material, design, and process. The participants in the discussion panel, in yearly meetings, define how the equipment can fail by identifying the failure modes, their effects, and their criticality. Firstly, the functions of the equipment have to be identified to specify how failures can occur. Then, it is necessary to establish how often a failure event occurs on the specific component for each failure mode. In this work, the attribution of the frequency class is performed according to the parameters reported in Table 6.

**Table 6 Frequency class assigned to each component failure mode.**

Annual Frequency Class	Description	Increasing Annual Frequency – (number of events per year)	Meaning
0	Practical non-credible occurrence	$<10^{-6}$	Could happen
A	Rare occurrence	$[10^{-6} ; <10^{-4})$	Reported in this industry
B	Unlikely occurrence	$[10^{-4} ; 10^{-3})$	Occurred at least once in the company
C	Credible occurrence	$[10^{-3} ; 10^{-1})$	Occurred several times in the company
D	Probable occurrence	$[10^{-1} ; 1)$	Happens several times/year in the company
E	Frequent occurrence	$\geq 1$	Happens several times/year in one location

Five levels – 0 for a low level, 4 for a high one - are used to classify the severity of the effects of a failure in a qualitative way. Three different categories are evaluated when rating the severity of an effect, i.e., safety, environmental, and assets (Table 7). The severity assignment does not follow objective criteria, i.e., measurable, but is defined by the panel of experts: for example, concerning the severity of the failure on production capacity (asset), Severity 1 can be assigned to equipment stoppage, Severity 2 to quality deviations of production parameters of the output, Severity 3 or 4 to the propagation of the effects, e.g., plant stoppage. The attribution of the criticality corresponding to each of the three aspects is performed semi-quantitatively, using the matrix reported in Table 7. The Criticality indexes are calculated by multiplying the severity of each effect category (safety, environmental, asset) by the annual frequency class:

- Safety Criticality Index = Severity Safety \* Annual frequency class

- Production/Asset Criticality index = Severity Production/Asset \* Annual frequency class
- Environment Criticality index = Severity Environment \* Annual frequency class

The frequency class and severity of effects are then inserted in the risk matrix (Table 7). They are compared with the acceptability criteria: the risk matrix is applied to define the critical elements through thresholds or criteria of "acceptance" of the criticality.

**Table 7 Risk matrix used in this case study**

Severity				Annual Frequency Class					
	Safety	Environment	Asset	0	A	B	C	D	E
0	Slight health effect	Slight damage	Slight effect	<b>C3: Continuous Improvement</b>					
1	Minor health effect	Minor effect	Minor damage						
2	Major health effect	Local effect	Local damage	<b>C2: Risk Reduction Measures</b>			<b>C1: Intolerable Risk</b>		
3	One fatality	Major effect	Major damage						
4	Multiple fatalities	Extensive effect	Extensive damage						

Within the risk matrix, it is possible to read the three criticalities of the failure event and assess its acceptability threshold.

- C3: the risk is tolerable; additional impact reduction measures are not necessary, but it is sufficient to monitor performance and manage it for continuous improvement.
- C2: The risk is intolerable; the risk will become tolerable after the definition and implementation of control measures.
- C1: The risk is intolerable; further impact reduction measures are needed.

The belonging of each item to one of the three categories is determined considering the annual frequency class and discussing the severity of the effects by the multidisciplinary team. In case of uncertainty or disagreement,

the worst criticality is assigned. In addition, it is possible to calculate the Overall Critical Failure Mode (CFM), representing the overall impact of a failure mode on a specific item, considering all the types of effects it has; it is calculated as the worst of the three effects (i.e., safety, environmental and assets) of that failure mode. To identify the Criticality of the Item, regardless of the specific failure mode, it is possible to calculate the following parameter:

$$\text{Criticality of the Item} = \text{Min (CFM)} \quad (4)$$

In other words, the Criticality of the Item is defined by considering the criticality of the most critical failure mode. An excerpt of the FMECA document regarding two different items is reported in appendix A to provide an example of the basic information noted. 1196 failure modes recorded in the previous five years concerning 501 items have been analyzed in the current analysis. For each failure mode, it is specified whether a corrective, cyclical (time-based), or condition-based maintenance policy is adopted. Remarkably, the company relies on a cyclical maintenance policy to prevent the majority of failure modes (693). Instead, the condition-based approach is adopted only in a low number of cases (34). The most critical plant section is Unit 100, which shows 248 failure modes.

#### 4.3 Data Analytics layer

In the development of the data analytics layer, the recommendations proposed in the framework are applied, firstly analyzing the general situation and then exploring further the areas of interest, according to this sequence:

- STEP 1 - Relationship analysis among failure modes, effects, and criticalities of the whole plant.
- STEP 2 - Relationships among failure modes and effects of the single unit.
- STEP 3 - Relationships among effects, maintenance tasks, and policies on every unit.

In these three steps, minimum support and confidence values must be set in order to create the association rules. The thresholds' setting to define the interestingness of the rules should take into account two different factors. On the one hand, the size of the output: low support and confidence thresholds can cause the generation of many rules, possibly exponential with regards to the input size, yielding several uninteresting relationships. On the other hand, setting high thresholds can cause the loss of interesting patterns, leading to ineffective dataset exploitation. As noted in Tan, Kumar, & Srivastava (2004), there is not an absolute optimal measure to set the thresholds; it strictly depends on the application case. For this reason, the expertise of the panel of experts is required to perform this process consistently. Specifically, the support threshold  $min\_sup$  is initially set to a very low value (e.g., 0.01) and then incremented to display the most statistically significant rules for each case represented in the case study.

The analysis of the total rules simultaneously is somewhat unrealistic, even though selecting appropriate thresholds means that only the most relevant are considered. In this application, the criticality of the items considered is used as a prioritization metric. When a failure occurs, corrective interventions or control and monitoring actions are performed according to the guidelines indicated in Table 8, prioritizing the items characterized by a higher criticality. In the case of more items belonging to the same category, if the interventions cannot be performed simultaneously, the support is used as a discriminant (the higher the support, the higher the item's priority over the concurrent ones).

**Table 8 Time scale for performing corrective actions or monitoring**

<b>Items criticality rate</b>	<b>Policy adopted</b>
Criticality Index = 1	The analysis must be started immediately
Criticality Index = 2	The analysis should be started within 24-48 hours
Criticality Index = 3	The analysis should be started within a week

#### 4.3.1 STEP 1: Relationships among failure modes and effects of the whole plant

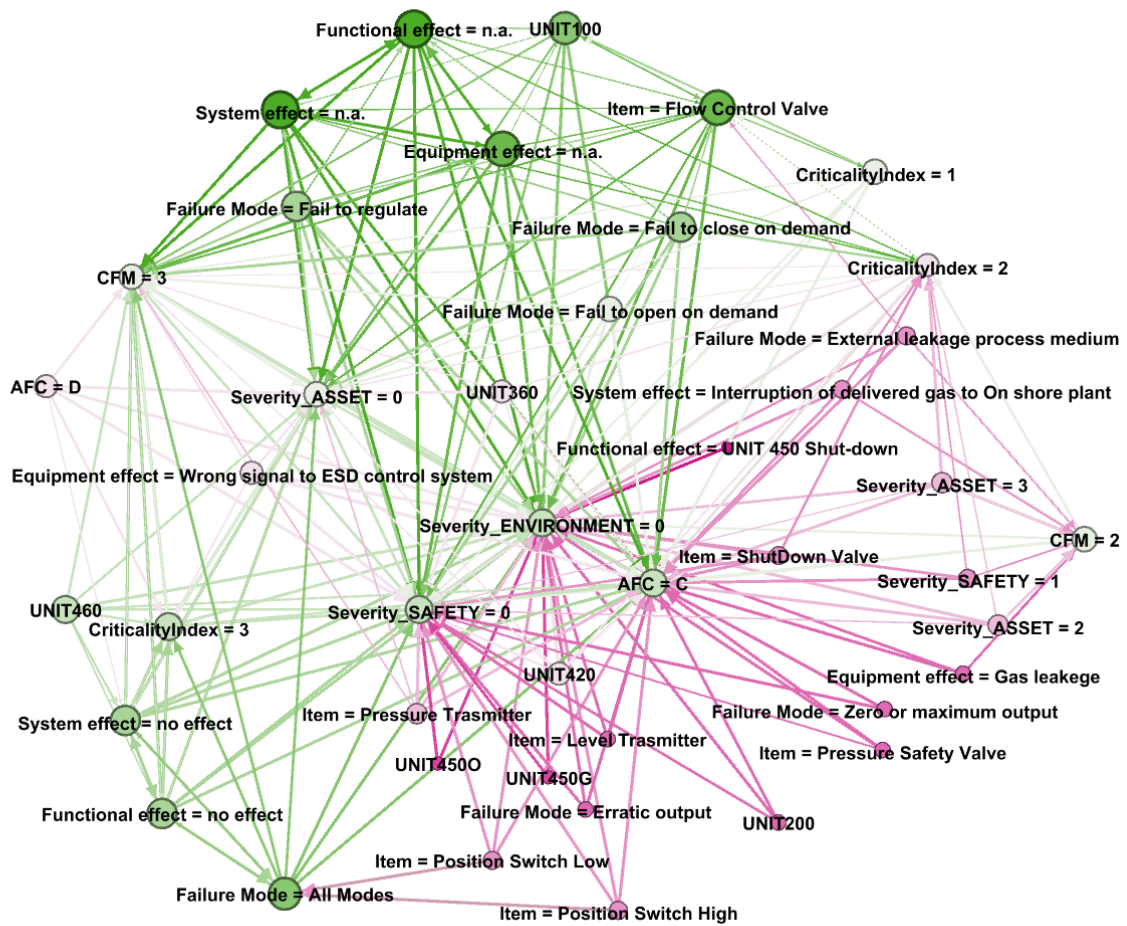
The network reported in Figure 2 is comprehensive of the principal relationships among failure modes and effects occurring across the entire production system and the annual frequency class of the failure modes, and the severity levels assessed in terms of asset, environment, and safety. In addition, to identify the most critical units and items, this information is reported in the network. The ARs' number and the SN characteristics obtained in Step 1 are shown in Table 10, varying the minimum support threshold, which is initially set to 0.01, while the minimum confidence threshold is set to 0.10, leading to the extraction of 1648 association rules. In terms of strength of the rules (i.e., confidence), there is no difference caused by the support increase from 0.01 to 0.05. However, the number of communities does not change incrementing the min\_sup since the attributes involved in the analysis at this first stage are vast; hence, none of the nodes are separated from the other ones. For sake of clarity, Figure 2 presents the network obtained by increasing the support to 0.05 and the minimum confidence to 0.25 (370 associations referring to 44 nodes). In this case, the number of ARs selected allows both an understandable visual representation and, in parallel, constitutes a considerable amount of relationships.

**Table 9 - ARs' number and SN characteristics varying the minimum support threshold**

ARs and SN characteristics	min_sup				
	0.01	0.05	0.10	0.20	0.50
<b># of ARs</b>	1648	370	179	75	20
<b>Minimum confidence in the selected range of ARs</b>	0.10	0.10	0.11	0.22	0.75
<b># of communities</b>	1	1	1	1	1
<b>Out-Degree (OD) range</b>	[0.0; 11.0]	[1.0; 8.6]	[1.9; 7.0]	[0.9; 6.6]	[6.9; 7.2]

In the network of Figure 2, the thickness of the edges represents the confidence of the rule, while the dimension and the color of the nodes indicate its out-degree level – i.e., the weighted sum of its outgoing

edges: the smaller the node, the lower the OD is; similarly, the colors of the nodes indicate different OD values, following a pink-white-green scale: pink nodes are characterized by low OD, white by a medium OD and, the greener they become, the higher the OD is. As previously noticed, a high OD indicates a strong influence of the node across the network; further developing the possible chain effects in these areas is worthy of investigation. Hence, the first unit considered for further analysis is Unit 100 (OD=6.64). Among the failure modes, the ones characterized by the highest out-degree are "Fail to close on demand" (OD=7.67) and "Fail to regulate" (OD=7.34). Both the failure modes are linked to the item "Flow control valve": the association rules "Failure Mode = Fail to regulate" → "Item = Flow Control Valve" and "Failure Mode = Fail to close on demand" → "Item = Flow Control Valve" are characterized by support of 0.05 and confidence of 0.65. The latter value shows that when the two failure modes occur, there is a 65% probability that the faulty component is the flow control valve. Indeed, one of the effects related to the failure of this item at the equipment level is the gas leakage in 13.18% of cases ("Item description = Flow Control Valve" → "Equipment effect = Gas leakage", conf = 0.1318).



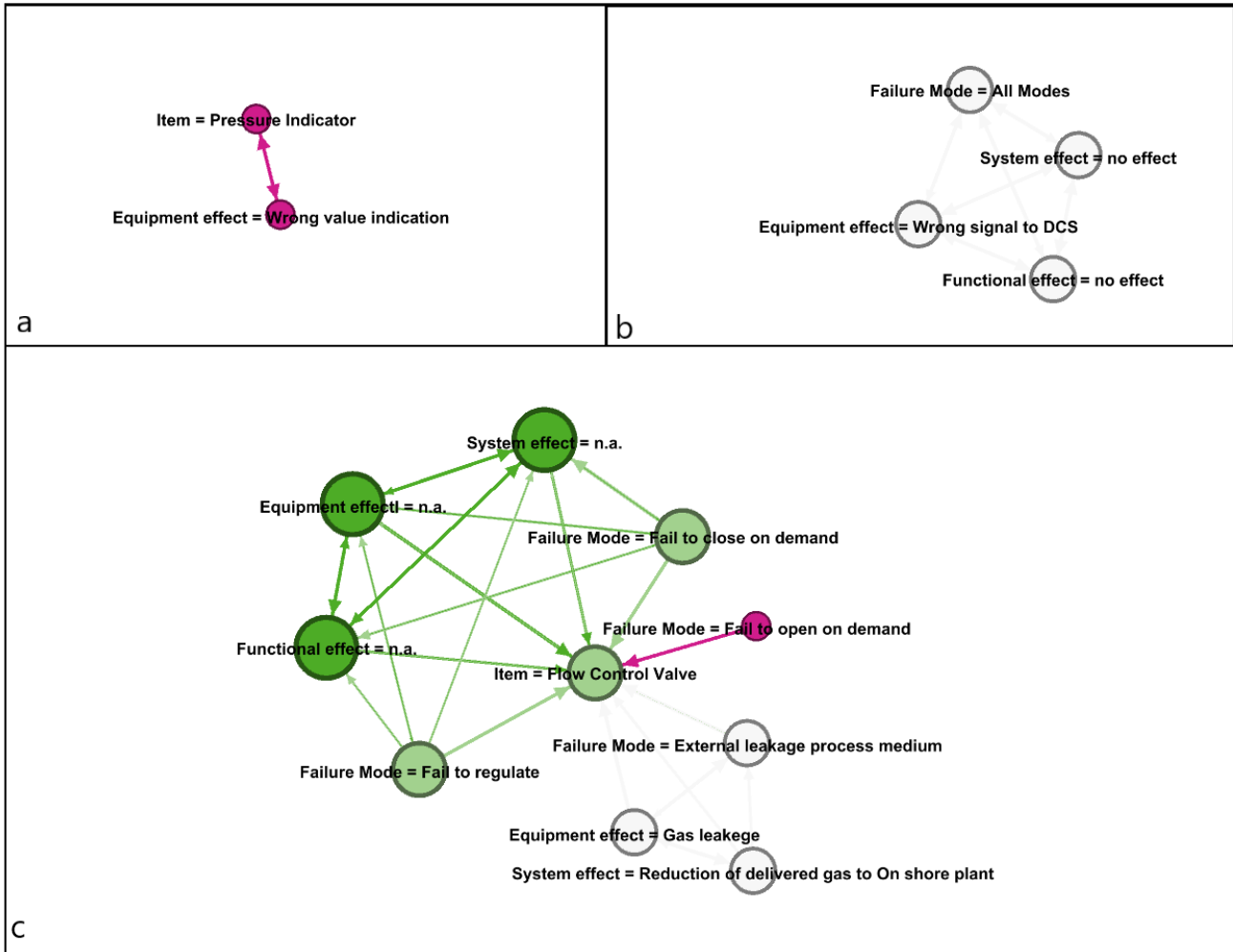
**Figure 2 Relationships among items, criticalities, failure modes, and effects of the whole plant; node dimension is proportional to the OD (the bigger the node, the higher the OD). Color scale, pick-white-green, represents growing OD levels.**

#### 4.3.2 STEP 2: Relationships among items, failure modes, and effects of every unit

The analysis carried out for the whole plant has been repeated for every unit to further the knowledge of every part of the plant. This analysis aims at highlighting the domino effect among failure modes within the unit. For example, the following SN (Figure 3) reports the relationships regarding failure modes and effects of the Oil&Gas production wellhead unit (Unit 100). To provide a clear representation of the SNs, following the same procedure deployed in Table 10, the minimum support threshold is set to 0.10.

Three communities of nodes are highlighted in the Unit 100 case (Figures 3a, 3b, and 3c). The first community of nodes reported in Figure 3a, “Item = Pressure indicator”  $\leftrightarrow$  “Equipment effect = wrong value indication” indicates two events only related among them. In other words, both the rule “Item = Pressure indicator”  $\rightarrow$  “Equipment effect = wrong value indication” and “Equipment effect = wrong value indication”  $\rightarrow$  “Item = Pressure indicator” are defined. The support and the confidence of both the rules are, respectively, 0.13 and 1.00: specifically, the two attribute-value relationships appear together in 13% of the dataset instances and, since the confidence is 100% for both sides of the rule, in every instance where “Item = Pressure indicator”, there is also “Equipment effect = wrong value indication”, and vice versa.

Similarly, in Figure 3b, the community comprises four nodes: one is representative of all the failure modes (“Failure mode = all modes”), while the other three nodes regard the related effects. Specifically, at the equipment level, the effect that verifies in the case of every failure mode is a wrong signal recorded by the distributed computer system (“Equipment effect = Wrong signal to DCS”). In contrast, no effect is related to such a failure mode at the system and functional level. The rules describing this relationship among the failure mode and the three effects are characterized by the highest confidence (0.81): “Failure mode = all modes”  $\rightarrow$  “Equipment effect = wrong signal to DCS”; “Failure mode = all modes”  $\rightarrow$  “System effect = no effect”; “Failure mode = all modes”  $\rightarrow$  “Functional effect = no effect”.



**Figure 3 Relationships among items, failure modes, and effects of Unit 100; node dimension is proportional to the OD (the bigger the node, the higher the OD). Color scale, pick-white-green, represents growing OD levels.**

The community of nodes reported in Figure 3c is more complex: four failure modes are represented, two effects at each level, and an item. The node "Item = Flow control valve" has an important function. Indeed, it represents the joint between two opposite sides of the community. If a significant control is dedicated to this item, indeed, it is possible to take control of all the failure modes and effects related to it. For example, suppose the failure mode is "External leakage process medium" then, in 65% of cases. In that case, there is a gas leakage at the equipment level ("Failure mode = leakage process medium" → "Equipment effect = gas leakage") and a reduction of delivered gas to the on-shore plant ("Failure mode = leakage process medium" →

“System effect = Reduction of delivered gas to On-shore plant”) occurs with the same probability (confidence = 0.65).

Remarkably, in the SN reported in Figure 3c, there are two different categories of the effects: on the upper side of the SN, we can see the equipment, functional, and system effects labeled with “n.a.”. This means that these effects are not assessable. Thus it is not possible to monitor or prevent them. On the other hand, effects like “Equipment effect = gas leakage” and “System effect = reduction of delivered gas to the on-shore plant” can be critical in terms of safety or asset integrity. Hence, these aspects will be analyzed further in the third step of the data-analytics phase.

4.3.3 STEP 3: Relationships among effects, maintenance tasks, and policies on Unit 100

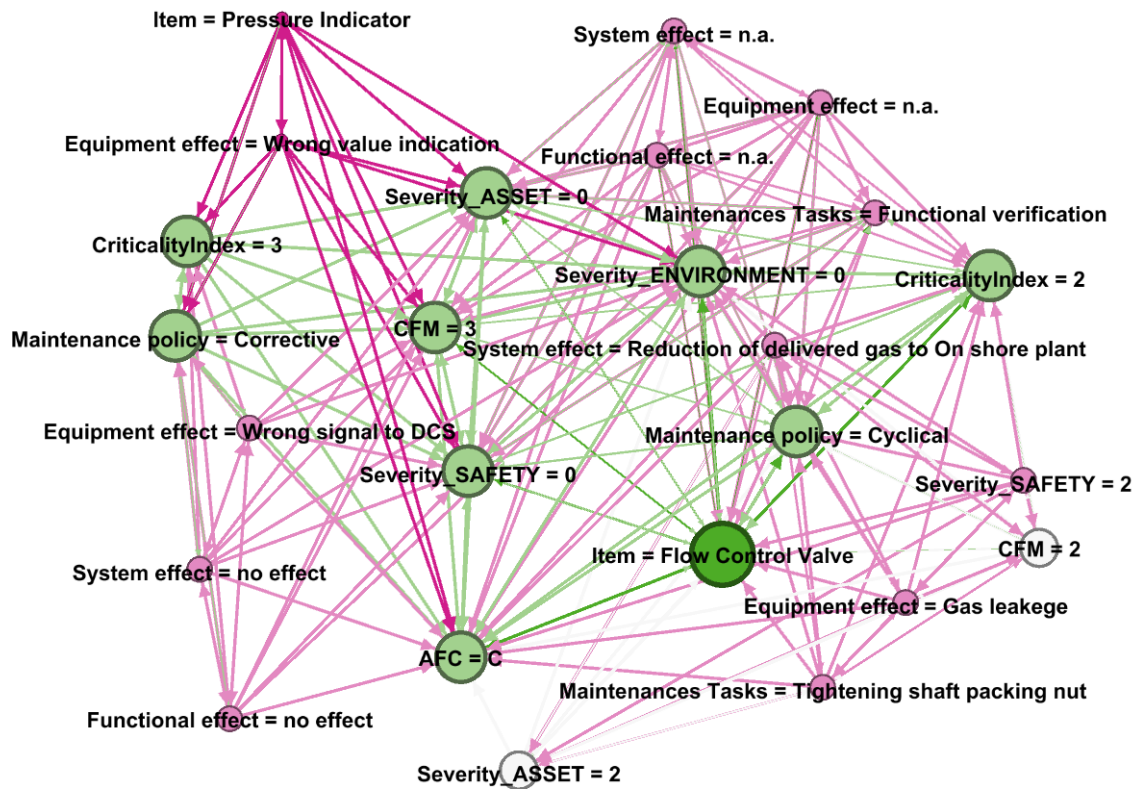


Figure 4 Relationships among effects, maintenance tasks, and policies of Unit 100; node dimension is proportional to the OD (the bigger the node, the higher the OD). Color scale, pick-white-green, represents growing OD levels.

The relationships among effects, maintenance tasks, and policies are analyzed to assess the maintenance policies' effectiveness and identify best practices. In Figure 4, the interrelations regarding maintenance tasks and policies, items, effects at all levels, criticality indexes, and severity are presented for Unit 100. Indeed, it is considered the most critical one, according to the considerations reported in section 4.3.1. In this case, the minimum support threshold is set to 0.1, while the minimum confidence at 0.20: 360 rules are presented in the SN of Figure 4; according to the procedure deployed in Table 10, these values represent the most appropriate trade-offs. In section 4.3.2, it is pointed out how the node "Item = Flow control valve" should be considered for its influence across the SN; the severity impact of a failure on this item is always null for the environmental aspects ("Item = Flow control valve" → Severity\_ENVIRONMENT = 0, confidence = 1.00) hence no concern rises

on this topic. A different situation is pointed out for what concerns asset integrity. According to the following rules, indeed, the severity can be 0 (thus causing only slight damages to the asset) or 2, i.e., causing local damages:

- “Item = Flow control valve” → “Severity\_ASSET = 0”, confidence = 0.72,
- “Item = Flow control valve” → “Severity\_ASSET = 2”, confidence = 0.28.

As presented in the bullet list, the former event is far more likely than the latter, with a probability of 72% versus 28%: slight damage to the assets is more likely than local damage. The flow control valve is associated with a Criticality index = 2 (confidence = 1.00) from the global analysis. Thus, it is assumed that there is the need for measures aiming at reducing the impact of the effects related to this item and that the procedures must start between 24 and 48 hours from the failure mode occurrence.

As testified by the following rules, no effect is shown at a system and functional level. Still, only at an equipment level, the maintenance policy adopted is corrective:

- System effect = no effect → Maintenance policy = Corrective, confidence = 1.00;
- Functional effect = no effect → Maintenance policy = Corrective, confidence = 1.00;
- Equipment effect = Wrong signal to DCS → Maintenance policy = Corrective, confidence = 1.00;
- Equipment effect = Wrong value indication → Maintenance policy = Corrective, confidence = 1.00;

However, when the maintenance policy defined for a specific failure mode is corrective, the effect related to it could also be non-assessable (n.a.) both at a system and functional level (Maintenance policy = Corrective → System effect = n.a., confidence = 0.21; Maintenance policy = Corrective → Functional effect = n.a., confidence = 0.21). In this case, the functional verification is foreseen with different percentages depending on the level of the effect (system, functional, or equipment level):

- In case of a non-assessable effect at the equipment level, the functional verification is always required (Equipment effect = n.a. → Maintenance task = Functional verification, confidence = 1.00);
- In case of a non-assessable effect at the functional level, the functional verification is required in 81% of the situations (Functional effect = n.a. → Maintenance task = Functional verification, confidence = 0.81);
- In case of a non-assessable effect at the system level, the functional verification is required in 83% of cases (System effect = n.a. → Maintenance task = Functional verification, confidence = 0.83).

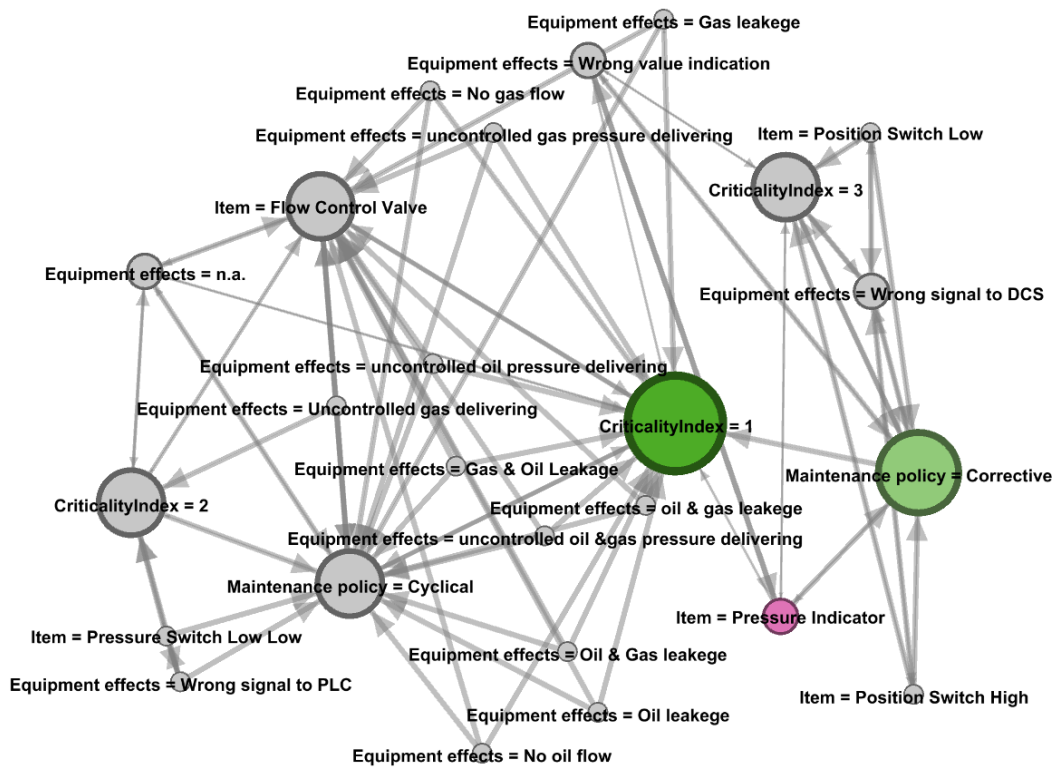
Several effects are related to a time-based preventive maintenance policy. At the equipment level, both gas leakage and no gas flow are prevented through cyclical maintenance interventions (“Equipment effect = Gas leakage” → “Maintenance policy = Cyclical”, confidence = 1.00; “Equipment effect = No gas flow” → “Maintenance policy = Cyclical”, confidence = 1.00), as well as the reduction or absence of gas flow at a functional (“Functional effect = Reduction of delivered gas to Unit 190 Launching Trap” → “Maintenance policy = Cyclical”, confidence = 1.00; “Functional effect = no gas flow to Unit 190 Launching Trap” → “Maintenance policy = Cyclical”, confidence = 1.00) and system level (“System effect = Reduction of delivered gas to On shore plant” → “Maintenance policy = Cyclical”, confidence = 1.00; “System effect = no gas flow to onshore plant” → “Maintenance policy = Cyclical”, confidence = 1.00).

#### *4.3.4 An example of a decision-making step based on SNA*

The analysis of the Social Network helps identify the chain of nodes that leads to high-risk events. For example, a potential misleading maintenance policy application leads to high-risk events in the case shown below. Knowing this sequence of nodes helps to break this chain to reduce the risk level of the processes under investigation. In Figure 5, for example, the SN showing the associations among items, equipment effects, maintenance policies, and criticality indexes is reported. The attention should be focused on the nodes “Criticality Index = 1”, “Maintenance policy = Corrective”, and “Item = Pressure Indicator” that are, for the

readers' convenience, colored. The two biggest nodes of the network (i.e., having the highest OD values, so the strongest influence of the node across the network) are "Criticality Index = 1" and "Maintenance policy = Corrective".

Moreover, there is an association rule between these nodes. Since the "Criticality Index = 1" indicates the highest risk for the plant, the association with a corrective maintenance policy is not justifiable and should be further inquired. Specifically, the item associated with these values is the pressure indicator. Indeed, in some cases, a wrong indication does not affect the functional and system levels. On the contrary, sometimes, it may imply a wrong setting on the plant's downstream machinery and uncontrolled pressure values. A modification in the maintenance policy is defined to avoid cascade effects. Periodical inspection routes are established to control the physical integrity of the pressure indicators, and specific alerts are set on the management system to advise the operators in case of unexpected pressure values. If this issue, as is the case, is noticed during the annual re-execution of the FMECA, it is immediately corrected, and the modification can be immediately implemented. Else, the modification can be introduced as a new best practice, and, at the next FMECA meeting, the revision is formalized.



**Figure 5 Relationships among effects, maintenance policies, and criticality indexes; node dimension is proportional to the OD (the bigger the node, the higher the OD).**

#### 4.3.5 Procedure validation

The mining of the ARs on a well-known data analytics platform ensures that the calculations are performed correctly. However, there is a three-step validation of the results before and during their implementation.

1. The first check is performed after the FMECA is carried out. Indeed, the objective is to avoid missing data at the first instance and prepare a starting dataset as complete as possible. For this reason, the FMECA is carried out by multidisciplinary teams, and the results have to be validated by every one of them and the plant chief. The document is updated periodically to be extended with the latest events and keep up with modifications of the plant.

2. The procedure implemented on the selected software (i.e., RapidMiner, in the case study) is double-checked by the engineers in charge of this step to fix any possible unnoticed error before the actual implementation.
3. During the first stages of the implementation, the failure events occurring are compared with the rules extracted to verify if the probability distribution of their occurrence and effects reflects the one described by the support and confidence of the ARs mined. The beginning of the procedure implementation dates back to April 2020, and data concerning the following eight months are considered. In 87.5%, there is a correspondence of the two compared cases, showing an acceptable accuracy of the proposed approach.

The entire process is then reiterated every year once the FMECA is updated.

## **5. Discussion**

The proposed research approach aims at supporting the development of the FMECA through further data-driven analysis, extracting the relationships among the failure modes, effects, and maintenance tasks assigned.

The contributions that can be highlighted from this framework are reported in the following subsections.

### **5.1 Theoretical implications**

In order to analyze the outcome of the FMECA, especially in the case of large and critical plants, a tool enabling the analysis of a vast amount of data is more suitable than the traditional statistical techniques. In this context, the ARM allows company managers the extraction of patterns characterized by potentially unknown relationships. Moreover, since the dataset can be analyzed all at once, it is unnecessary to formulate further research hypotheses, leaving all data-driven search possibilities unrestricted. Defining all the possible itemsets represents an NP-hard problem since the possible combinations of the items have a size  $2^q - 1$  ( $q$  being the

number of items in the dataset under investigation), excluding the non-valid and empty sets. This issue makes the dimensionality of the input space a critical aspect. However, due to the anti-monotonicity (or downward-closure) property of the support, the definition of the frequent itemset is more efficient since none of the infrequent itemsets is a subset of a frequent itemset. Additionally, the data must be cleaned before starting the ARM process. Indeed, the presence of redundancy in data impacts both results quality and algorithms efficiency.

## 5.2 Practical implications

The proposed method enables company managers to connect multidimensional and multidisciplinary concepts (e.g., failure mode, equipment criticality, failure effects at different levels, and maintenance policies adopted). From a practical point of view, this approach enables a significant control of the process analysis, specifically in the maintenance field. Firstly, having a procedure defining how the traditional FMECA can further analyze the failure modes, effects, and maintenance procedure, can provide strategic support and a growth opportunity for the company. Indeed, the more reliable the failure analysis is, the more consistent the achievable benefits.

As stated before, SNA helps know how a failure mode impacts the possible effect identified and whether these effects are corrected or monitored cyclically, making the understanding of the interactions very intuitive both for domain experts and non-experts. On the other hand, a missing connection among nodes can help experts understand whether some critical details have been missed during the first stages of the analysis. For example, inconsistencies during the data collection and management phases can be noticed and amended before making decisions on incoherent information. It is also easier to implement corrections and improvements promptly by integrating the actions defined after the FMECA or, even further upstream, by adding a failure mode or an effect to equipment.

From an engineering point of view, this visualization is useful for defining the event chains that are more critical since they can act as a trigger for other effects, not only at an equipment level but also at a functional or system level. In this way, specific resources can be dedicated to analyzing the most critical areas and the critical components, like defining ad-hoc strategies or analyzing the structural improvement of such areas. Furthermore, identifying the possible causes of concatenating failure modes and effects on items allows identifying the critical items from a traditional risk-assessment perspective and considering the patterns extracted from the data-driven analytics. Indeed, the nodes acting as bridges in the SN, thus connecting different communities of nodes, has to be taken into particular consideration. For such nodes (e.g., the flow control valve in Figure 3c), specific monitoring activities can be planned to prevent the spreading of failures or effects across the network. Indeed, in process industries, it is vital to be aware of the possible propagation of the effects due to the hazardous nature of the production process and the deriving danger.

The adoption of the proposed framework is also helpful in the case of re-layout of the plant or designing similar ones. Indeed, the reliability and failure modes should be considered even in the early stages of realizing a production system: anticipating these issues helps define proper strategies to deal with them and organize the operations and maintenance activities accordingly.

The adoption of new techniques, like ARM and SNA, does not entirely change the company's failure analysis procedure, and they are adopted as an addition to the present one. In this way, the change for the workers is not radical and allows gradual habituation to the new methodologies. This aspect is important in guaranteeing that the personnel willingly accept new methodologies without completely abandoning the previous habits, avoiding possible resistance to change.

## 6. Conclusions

This work proposes a framework aiming to deepen the failure analysis by applying two data-driven techniques, the Association Rule Mining and the Social Network Analysis. The framework involves the data collection to carry out the well-known Failure Modes Effects and Criticality Analysis process, and an analysis of its outcome follows it through the Association Rule Mining. Indeed, the Association Rule Mining aims to define the co-occurrence of events, like failure modes and the related effects, to have a clearer idea of the dynamics describing the possible failure in a company. Thanks to the Social Network Analysis, the Association Rules are represented as a network of nodes (the attributes of the Failure Modes Effects and Criticality Analysis) and direct edges (the Association Rules among the nodes). The possibility of having a graphical representation of the association rules facilitates the global understanding of the context, highlighting which failure modes and effects are related among them and detecting a possible lack of information to have a clear view of the process and implement improvement actions.

The proposed research approach is applied to a real-life case study of an onshore/offshore plant used for oil and gas extraction. The results of this implementation highlighted the existence of potentially missing information and, thus, the need for improvement of the data collection to avoid this lack and the necessity of significant responsiveness in case of specific failure modes.

Applying the two data-driven techniques, Association Rule Mining and the Social Network Analysis, is rather innovative in the maintenance management field both singularly and jointly. The innovative contribution is not on the single technique but on how the two techniques are implemented. Indeed, they are applied in sequence providing information about communities and possible chains of failures along the plant. The idea of combining them after the deployment of the traditional Failure Modes Effects and Criticality Analysis process implies the integration of these new methodologies into this kind of analysis, not its elimination. In this way, the current body of knowledge in the failure analysis area may be extended, capitalizing on data analytics techniques'

benefits. Future research directions regards other case studies, so that standard procedures belonging to the same field of research might be provided.

An important limitation of this work regards the time dimension. This aspect is not considered in extracting the ARs. It would indeed be interesting to analyze the temporal distance between the occurrence of two (or more) faults in the process. In this way, it would be possible to identify the modifications of the relationships along time and precise scheduling of the maintenance activity so that the newly introduced policies and practices could be compared with the previous one. During the implementation of the proposed approach, it is fundamental to have the support of the whole team involved in the study. In case of member reluctance, the collaborative approach at the basis of the FMECA procedure and the further analysis of the outcomes might be ineffective since opinions from different expertise areas ensure a complete understanding of the production environment.

## **Appendix A**

In Table A1, an excerpt of the FMECA document regarding two different items is shown to provide an example of the basic information reported. The table reports the ID number of the item analyzed, followed by its synthetic description that makes it understandable, and the list of failure modes. In addition, the effects deriving from the failure modes are reported. A specification is made in considering the effects, individuating three different levels:

- Equipment effect, describes the failure at the item level;
- Functional effect, describes the failure at a facility functional level (i.e., inside the same unit);
- System effect, describes the effect on the whole plant (i.e., more than one unit is involved);

Moreover, the annual frequency class (AFC), severity levels and their combination (ICA, ICS, ICE) are reported, as well as the criticality in terms of assets, environment and safety. The overall criticality failure mode CFM (for each failure mode) and the item criticality CI (for each item) are also expressed.

**Table A1 Excerpt of the data coming from the FMECA process**

TAG Number	Item description	Failure Mode Description	Equipment effect	Functional effect	System effect	AFC	Severity Safety	Severity Asset	Severity Environment	ICS	ICA	ICE	CSS	CSA	CSE	CFM	CI
0130361000S SSV070	Flow Control Valve	Fail to close on demand	n.a.	n.a.	n.a.	C	0	0	0	C0	C0	C0	3	3	3	3	2
	Flow Control Valve	Fail to open on demand	No oil flow	no oil flow to unit 200 Oil Production Separation	no oil flow to on shore plant	C	0	0	0	C0	C0	C0	3	3	3	3	
	Flow Control Valve	Fail to regulate	n.a.	n.a.	n.a.	C	0	0	0	C0	C0	C0	3	3	3	3	
	Flow Control Valve	External leakage process medium	Gas & Oil Leakage	Reduction of delivered oil & GAS to Unit 200 Oil Production Separation	Reduction of delivered oil & gas to On shore plant	C	1	0	0	C1	C0	C0	3	3	3	3	
	Flow Control Valve	External leakage process medium	Gas & Oil Leakage	Reduction of delivered oil & GAS to Unit 200 Oil Production Separation	Reduction of delivered oil & gas to On shore plant	C	1	2	0	C1	C2	C0	3	2	3	2	
0130361000S SV071	Flow Control Valve	Fail to close on demand	n.a.	n.a.	n.a.	C	0	0	0	C0	C0	C0	3	3	3	3	2
	Flow Control Valve	Fail to open on demand	No oil flow	no oil flow to unit 200 Oil Production Separation	no oil flow to on shore plant	C	0	0	0	C0	C0	C0	3	3	3	3	
	Flow Control Valve	Fail to regulate	n.a.	n.a.	n.a.	C	0	0	0	C0	C0	C0	3	3	3	3	
	Flow Control Valve	External leakage process medium	Gas & Oil Leakage	Reduction of delivered oil & GAS to Unit 200 Oil Production Separation	Reduction of delivered oil & gas to On shore plant	C	0	0	0	C0	C0	C0	3	3	3	3	

Flow Control Valve	External leakage process medium	Gas & Oil Leakage	Reduction of delivered oil & GAS to Unit 200 Oil Production Separation	Reduction of delivered oil & gas to On shore plant	C	1	2	0	C1	C2	C0	3	2	3	2
...	...	...	...	...											

---

## References

- Aggogeri, F., Adamini, R., Aivaliotis, P., Borboni, A., Eytan, A., Merlo, A., Németh, I., Taesi, C., & Pellegrini, N. (2020). Robotic System Reliability Analysis and RUL Estimation Using an Iterative Approach. *Advances in Intelligent Systems and Computing*, 980, 134–143. [https://doi.org/10.1007/978-3-030-19648-6\\_16](https://doi.org/10.1007/978-3-030-19648-6_16)
- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *ACM SIGMOD Record*. <https://doi.org/10.1145/170036.170072>
- Antomarioni, S., Pisacane, O., Potena, D., Bevilacqua, M., Ciarapica, F. E., & Diamantini, C. (2019). A predictive association rule-based maintenance policy to minimize the probability of breakages: application to an oil refinery. *International Journal of Advanced Manufacturing Technology*. <https://doi.org/10.1007/s00170-019-03822-y>
- Antomarioni, Sara, Bevilacqua, M., Potena, D., & Diamantini, C. (2019). Defining a data-driven maintenance policy: an application to an oil refinery plant. *International Journal of Quality & Reliability Management*, 36(1), 77–97. <https://doi.org/10.1108/IJQRM-01-2018-0012>
- Antomarioni, Sara, Lucantoni, L., Ciarapica, F. E., & Bevilacqua, M. (2021). Data-driven decision support system for managing item allocation in an ASRS: A framework development and a case study. *Expert Systems with Applications*, 185, 115622. <https://doi.org/10.1016/J.ESWA.2021.115622>
- Ben Said, A., Shahzad, M. K., Zamai, E., Hubac, S., & Tollenaere, M. (2016). Experts' knowledge renewal and maintenance actions effectiveness in high-mix low-volume industries, using Bayesian approach. *Cognition, Technology and Work*, 18(1), 193–213. <https://doi.org/10.1007/s10111-015-0354-y>
- Bevilacqua, M., Ciarapica, F. E., & Mazzuto, G. (2018). Fuzzy cognitive maps for adverse drug event risk management. *Safety Science*, 102, 194–210. <https://doi.org/10.1016/j.ssci.2017.10.022>
- Carpitella, S., Certa, A., Izquierdo, J., & La Fata, C. M. (2018). A combined multi-criteria approach to support FMECA analyses: A real-world case. *Reliability Engineering and System Safety*, 169, 394–402. <https://doi.org/10.1016/j.res.2017.09.017>
- Chan, A., Javed, A. A., Wong, F., Hon, C., Zahoor, H., & Lyu, S. (2014). The application of social network analysis in the construction industry of Hong Kong. *International Conference on Emerging Trends in Engineering, Management and Sciences*.
- Chang, K. H., & Cheng, C. H. (2011). Evaluating the risk of failure using the fuzzy OWA and DEMATEL method. *Journal of Intelligent Manufacturing*, 22(2), 113–129. <https://doi.org/10.1007/s10845-009-0266-x>
- Chen, W. C., Tseng, S. S., & Wang, C. Y. (2005). A novel manufacturing defect detection method using

association rule mining techniques. *Expert Systems with Applications*.  
<https://doi.org/10.1016/j.eswa.2005.06.004>

- Ciarapica, F., Bevilacqua, M., & Antomarioni, S. (2019). An approach based on association rules and social network analysis for managing environmental risk: A case study from a process industry. *Process Safety and Environmental Protection*. <https://doi.org/10.1016/j.psep.2019.05.037>
- Crespo Márquez, A., de la Fuente Carmona, A., & Antomarioni, S. (2019). A Process to Implement an Artificial Neural Network and Association Rules Techniques to Improve Asset Performance and Energy Efficiency. *Energies*, 12(18), 3454. <https://doi.org/10.3390/en12183454>
- Da Cunha, C., Agard, B., & Kusiak, A. (2006). Data mining for improvement of product quality. *International Journal of Production Research*, 44(18–19), 4027–4041. <https://doi.org/10.1080/00207540600678904>
- Eteifa, S. O., & El-Adaway, I. H. (2018). Using Social Network Analysis to Model the Interaction between Root Causes of Fatalities in the Construction Industry. *Journal of Management in Engineering*, 34(1), 04017045. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000567](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000567)
- Grunske, L., Lindsay, P., Yatapanage, N., & Winter, K. (2005). An automated failure mode and effect analysis based on high-level design specification with behavior trees. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3771 LNCS, 129–149. [https://doi.org/10.1007/11589976\\_9](https://doi.org/10.1007/11589976_9)
- Gunawardana, A., Meek, C., & Xu, P. (2011). A Model for Temporal Dependencies in Event Streams. In *Advances in Neural Information Processing Systems* (Vol. 24).
- Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 29(2), 1–12. <https://doi.org/10.1145/335191.335372>
- Harding, J. A., Shahbaz, M., Srinivas, & Kusiak, A. (2006). Data mining in manufacturing: A review. In *Journal of Manufacturing Science and Engineering, Transactions of the ASME* (Vol. 128, Issue 4, pp. 969–976). American Society of Mechanical Engineers Digital Collection. <https://doi.org/10.1115/1.2194554>
- Huang, J., You, J. X., Liu, H. C., & Song, M. S. (2020). Failure mode and effect analysis improvement: A systematic literature review and future research agenda. *Reliability Engineering and System Safety*, 199(January 2019), 106885. <https://doi.org/10.1016/j.res.2020.106885>
- Khorshidi, H. A., Gunawan, I., & Ibrahim, M. Y. (2016). Data-Driven System Reliability and Failure Behavior Modeling Using FMECA. *IEEE Transactions on Industrial Informatics*, 12(3), 1253–1260. <https://doi.org/10.1109/TII.2015.2431224>
- Kim, J., Han, S., & Hyun, C. T. (2019). Identification and Reduction of Synchronous Replacements in Life-Cycle Cost Analysis of Equipment. *Journal of Management in Engineering*, 35(1), 04018058. [https://doi.org/10.1061/\(ASCE\)ME.1943-5479.0000673](https://doi.org/10.1061/(ASCE)ME.1943-5479.0000673)
- Knoke, D., & Yang, S. (n.d.). Network fundamentals. *Methods.Sagepub.Com*. Retrieved April 9, 2020, from <https://methods.sagepub.com/base/download/BookChapter/social-network-analysis/n2.xml>
- Li, X., Ran, Y., Zhang, G., & He, Y. (2019). A failure mode and risk assessment method based on cloud model. *Journal of Intelligent Manufacturing*. <https://doi.org/10.1007/s10845-019-01513-9>
- Liu, H. C., Chen, Y. Z., You, J. X., & Li, H. (2016). Risk evaluation in failure mode and effects analysis using fuzzy digraph and matrix approach. *Journal of Intelligent Manufacturing*, 27(4), 805–816.

<https://doi.org/10.1007/s10845-014-0915-6>

- Liu, L., Fan, D., Wang, Z., Yang, D., Cui, J., Ma, X., & Ren, Y. (2019). Enhanced GO methodology to support failure mode, effects and criticality analysis. *Journal of Intelligent Manufacturing*, 30(3), 1451–1468. <https://doi.org/10.1007/s10845-017-1336-0>
- Luo, S. H., & Lee, G. G. (2015). Applying failure mode and effects analysis for successful knowledge management. *Total Quality Management and Business Excellence*, 26, 62–75. <https://doi.org/10.1080/14783363.2012.733263>
- Lv, J., Xu, S., Zhang, R., Xiao, H. , & Chen, Z. (2018). Safety Analysis of Metro Turnouts Based on Fuzzy FMECA. *2018 IEEE International Conference on Internet of Things (IThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 0, 599–606. [https://doi.org/10.1109/Cybermatics\\_2018.2018.00123](https://doi.org/10.1109/Cybermatics_2018.2018.00123)
- Ma, H., Chu, X., Xue, D., & Chen, D. (2019). Identification of to-be-improved components for redesign of complex products and systems based on fuzzy QFD and FMEA. *Journal of Intelligent Manufacturing*, 30(2), 623–639. <https://doi.org/10.1007/s10845-016-1269-z>
- Marcucci, G., Antomarioni, S., Ciarapica, F. E., & Bevilacqua, M. (2021). The impact of Operations and IT-related Industry 4.0 key technologies on organizational resilience. *Production Planning and Control*. <https://doi.org/10.1080/09537287.2021.1874702>
- Mushiri, T., Hungwe, R., & Mbohwa, C. (2018). An artificial intelligence based model for implementation in the petroleum storage industry to optimize maintenance. *IEEE International Conference on Industrial Engineering and Engineering Management, 2017-December*, 1485–1489. <https://doi.org/10.1109/IEEM.2017.8290140>
- Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441–453. <https://doi.org/10.1177/016555150202800601>
- Ragab, A., Yacout, S., Ouali, M. S., & Osman, H. (2019). Prognostics of multiple failure modes in rotating machinery using a pattern-based classifier and cumulative incidence functions. *Journal of Intelligent Manufacturing*, 30(1), 255–274. <https://doi.org/10.1007/s10845-016-1244-8>
- Renjith, V. R., Jose kalathil, M., Kumar, P. H., & Madhavan, D. (2018). Fuzzy FMECA (failure mode effect and criticality analysis) of LNG storage facility. *Journal of Loss Prevention in the Process Industries*, 56, 537–547. <https://doi.org/10.1016/j.jlp.2018.01.002>
- Ruan, X., Ochieng, D. E. G., Price, A. D., & Egbu, C. O. (2015). Time for a Real Shift to Relations: Appraisal of Social Network Analysis Applications in the UK Construction Industry. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2623032>
- Savino, M. M., Brun, A., & Riccio, C. (2011). Integrated system for maintenance and safety management through FMECA principles and fuzzy inference engine. *European Journal of Industrial Engineering*, 5(2), 132–169. <https://doi.org/10.1504/EJIE.2011.039870>
- Scott, J. (1988). Social Network Analysis. *Sociology*, 22(1), 109–127. <https://doi.org/10.1177/0038038588022001007>
- Simonoff, J. S., Restrepo, C. E., & Zimmerman, R. (2010). Risk management of cost consequences in natural gas transmission and distribution infrastructures. *Journal of Loss Prevention in the Process Industries*, 23(2),

269–279. <https://doi.org/10.1016/j.jlp.2009.10.005>

Tan, P. N., Kumar, V., & Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems*, 29(4), 293–313. [https://doi.org/10.1016/S0306-4379\(03\)00072-3](https://doi.org/10.1016/S0306-4379(03)00072-3)

Tso, K. S., Tai, A. T., Chau, S. N., & Alkalai, L. (2005). On automating failure mode analysis and enhancing its integrity. *Proceedings - 11th Pacific Rim International Symposium on Dependable Computing, PRDC 2005, 2005*, 287–292. <https://doi.org/10.1109/PRDC.2005.42>

US Military Standard. (1980). *MIL-STD-1629a. Procedures for performing a failure mode, effect and criticality analysis.*

US Military Standard. (1983). *MIL-STD-1629A. Procedures for performing a failure mode, effect and criticality analysis.*

Viveros, P., Nikulin, C., López-Campos, M., Villalón, R., & Crespo, A. (2018). Resolution of reliability problems based on failure mode analysis: an integrated proposal applied to a mining case study. *Production Planning and Control*, 29(15), 1225–1237. <https://doi.org/10.1080/09537287.2018.1520293>

Wang, J. (2002). Offshore safety case approach and formal safety assessment of ships. *Journal of Safety Research*, 33(1), 81–115. [https://doi.org/10.1016/S0022-4375\(02\)00005-1](https://doi.org/10.1016/S0022-4375(02)00005-1)

Xiao, S, Hu, Y., Han, J., Zhou, R., Cirp, J. W.-P., & 2016, U. (2016). Bayesian networks-based association rules and knowledge reuse in maintenance decision-making of industrial product-service systems. *Elsevier*. <https://www.sciencedirect.com/science/article/pii/S2212827116300166>

Xiao, Shasha, Hu, Y., Han, J., Zhou, R., & Wen, J. (2016). Bayesian Networks-based Association Rules and Knowledge Reuse in Maintenance Decision-Making of Industrial Product-Service Systems. *Procedia CIRP*, 47, 198–203. <https://doi.org/10.1016/j.procir.2016.03.046>

Xu, Z., Dang, Y., Munro, P., & Wang, Y. (2020). A data-driven approach for constructing the component-failure mode matrix for FMEA. *Journal of Intelligent Manufacturing*, 31(1), 249–265. <https://doi.org/10.1007/s10845-019-01466-z>

Yuan, Y., Zhou, S., Sievenpiper, C., Mannar, K., & Zheng, Y. (2011). Event log modeling and analysis for system failure prediction. *IIE Transactions (Institute of Industrial Engineers)*, 43(9), 647–660. <https://doi.org/10.1080/0740817X.2010.546385>