The following publication L. Xia, P. Zheng, M. Herrera, Y. Liang, X. Li and L. Gao, "Graph Embedding-Based Bayesian Network for Fault Isolation in Complex Equipment," in IEEE Transactions on Reliability, vol. 74, no. 3, pp. 3897-3910, Sept. 2025 is available at https://doi.org/10.1109/TR.2024.3416064.

Graph Embedding-Based Bayesian Network for Fault Isolation in Complex Equipment

Liqiao Xia, *Member, IEEE*, Pai Zheng, *Senior Member, IEEE*, Manuel Herrera, *Member, IEEE*, Yongshi Liang, Xinyu Li, *Member, IEEE*, and Liang Gao, *Senior Member, IEEE*

***Abstract*—Fault isolation, or fault location, aims to identify anomalous components at the start of the maintenance process. However, fault isolation within complex equipment can be challenging due to constraints on the scarcity of labeled data and the intricate interaction among various substructures. To overcome this challenge, an embedding-based Bayesian Network (BN) probability inference is proposed to locate the fault components, where the embedding, derived from semantic meanings, can approximate the actual fault distribution within BN. Firstly, a Fault Graph (FG) is established based on the equipment's mechanical structure and its mechanisms. Then, a Multi-field hyperbolic embedding is employed to vectorize the nodes in the FG, thereby preserving the inherent logic maximally. Following this, the FG is transformed into the BN, which facilitates the prediction of the faulty component based on available evidence, using the well-trained graph embedding. An empirical study on oil drilling equipment showcases the graph embedding properties and inference performance of the proposed method by comparing it with other cutting-edge methods and traditional scenarios.**

***Index Terms*—Fault isolation, knowledge graph, reliability analysis, cognitive predictive maintenance, bayesian network.**

I. Introduction

COMPLEX equipment embodies a plethora of mechanical and electrical constituents, each indispensable to the intricate prerequisites, compositions, and processes requisite for operation. Ensuring the uninterrupted functionality of such machinery is critical for fulfilling its intended role, minimizing production expenditure, and curtailment of replacement components and retooling. Prognostics and Health Management (PHM) proffers a robust framework for guaranteeing the enduring reliability of such intricate equipment [2]. Nonetheless, an elemental facet of PHM is the precise identification of faults within the equipment, a precondition for tasks such as prognostics and fault diagnosis. Comprehending the root cause of equipment bolsters the continuous reliable operation of the machinery, thereby solidifying its longevity and efficacy [3].

Identifying the fault location in complex equipment, sometimes referred to as fault isolation, presents a formidable challenge due to the difficulty of label collection and the intricate structure of such systems [4]. First, it is really difficult to map signal data with fault components without prior acknowledgment. Besides, the operational constituents within these systems exhibit high levels of correlation and coupling, thereby complicating the task of isolating the root cause of a malfunction [5]. For instance, in a progressive cavity pump where the rotation of one screw propels another; any perturbation within a single component can trigger a domino effect, impacting the entire system. Moreover, signals obtained by these sensors can be influenced by the operational conditions of adjacent components, further exacerbating the predicament of isolating the potentially faulty component [6]. Therefore, precisely identifying the component at risk of failure persists as a significant and unresolved issue.

In the face of the limited data accessibility challenge, topological structures, such as trees or graphs, serve as viable instruments for furnishing additional knowledge for fault isolation. A classical method, Fault Tree (FT) [7], leverages Boolean logic to integrate a sequence of fault occurrences, thereby locating the component-level fault through the navigation of the tree structure. Yet, determining the logical linkages within can prove burdensome due to the intricate relationships and complex physical principles [8]. Furthermore, the semantic meaning embedded in the topology might possess limited expressiveness and may falter in reflecting the probability of each outcome [9]. Inaccurate calculations of failure probability can severely limit the identification of the root cause, hindering its widespread application.

In a bid to provide additional representational space for the semantic interpretation of fault descriptions, large-scale semantics can be organized via a Knowledge Graph (KG) [10]. This framework could comprehensively depict and leverage semantic relationships, positioning KG as a viable instrument for fault isolation. The original FT can be expanded into a Fault Graph (FG), where the scalable graph size lays the groundwork for graph embedding learning to compute the transfer probability of sequential fault events [11]. These

This research is partially funded by the Department of Science and Technology of Guangdong Province (2023A1515011557), State Key Laboratory of Ultra-precision Machining Technology, The Hong Kong Polytechnic University, HKSAR, China (1-BBR2), and Shanghai Science and Technology Program (22010500900).

Liqiao Xia, Yongshi Liang are with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China (e-mail: liqiao.xia@connect.polyu.hk, and yongshi.liang@polyu.edu.hk).

Pai Zheng is with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China, and also with State Key Laboratory of Ultra-precision Machining Technology, Department of Industrial Systems and Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China (e-mail: pai.zheng@polyu.edu.hk).

Manuel Herrera is with Institute for Manufacturing, Department of Engineering, University of Cambridge, Cambridge, CB3 0FS, UK (e-mail: amh226@cam.ac.uk).

Xinyu Li and Liang Gao are with the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China (email: lixinyu@mail.hust.edu.cn, gaoliang@mail.hust.edu.cn)

This is an extension from a conference paper [1], where the new journal supplements the detailed construction method of the fault graph, optimizes the graph embedding algorithm, and provides a more comprehensive experiment.

embedding can represent the intricate substructures inside the equipment through calculation, compensating for the drawback of FT. Nevertheless, preserving the diverse logical relationships within the FG presents a significant challenge. Hence, finding an efficacious way to integrate these relationships into graph embedding remains a key research focus.

Despite the strength of a well-embedded FG as a robust graph database, its limitation lies in its inability to represent a probabilistic inference model - a challenge that a Bayesian Network (BN) is poised to overcome [12]. Drawing from the structure of the FG, the BN explicates the conditional dependencies of multiple scenarios via a directed acyclic graph. This establishes a promising, well-initialized BN for inference, devoid of any historical training data. Nonetheless, transforming the FG into the BN may be challenged with practical impediments. Initially, while a single node within the BN often encompasses various conditions, these individual conditions are represented as distinct nodes in the FG. Moreover, inferring potential root fault components within the BN poses a challenge because the FG does not incorporate the occurrence probabilities for each condition (node), let alone the conditional probability calculation.

To address the problems aforementioned, this study proposes the integration of BN inference, grounded in hyperbolic graph embedding, with the FG to localize defective components under situations of limited data availability accurately. The fusion of FG, hyperbolic graph embedding, BN inference, facilitates an accurate computation of the likelihood of a faulty component. The subsequent scientific novelty of this work can be enumerated as follows:

1) Introduce a construction methodology of FG that encapsulates the complex interplay between various components by integrating them together into a large-scale graph. This FG considers the intricate fault logic interlacing the components, thereby providing a comprehensive perspective of complex equipment. Moreover, the large-scale and structured FG lays the solid fundament for well-trained graph embedding.

2) Propose a Multi-field hyperbolic embedding (MFH) approach to vectorize the nodes and edges, tailored specifically to capture the multiplicity of logical rules inherent in the FG, with a particular focus on low-frequency nodes. This method, is tailored for the unique characteristics of the FG, rendering it more suitable for the computation of probabilities in subsequent BN inference.

3) Establish an embedding-based BN inference algorithm to calculate the most probable faulty component. This transformation of the FG into BN format stimulates the power of graph-based probability inference, thus locating the faulty component accurately even in the absence of available training label data.

The remainder of this paper is structured as follows. Section II reviews the previous graph embedding, fault location, BN progress, and provides a clear articulation of the problem statement. Section III describes the construction of FG. Section IV offers MFH graph embedding, while Section V presents the BN inference. A case study of oil drilling equipment will be conducted for validation and it will be presented in Section VI. Finally, Section VII highlights the limitations of this study and suggests future directions.

II. Related work

A. Fault Isolation

Fault isolation methods can be categorized into two distinct categories: model-based and data-driven.

Model-based fault isolation endeavors to leverage physical principles for fault location. Representative of this approach is the Unscented Kalman Filter, which models power and thermal processes [13]. Another facet of this method involves fault detection by juxtaposing sensor and actuator readings with a mathematical model of the system [4]. Additionally, FT analysis often finds utility in this classification [7]. Nevertheless, these model-based strategies often encounter a formidable hurdle in the mathematical formulation of mechanical systems. Concurrently, ambiguous disturbances, such as measurement noise, may pose significant impediments to the successful implementation of these models.

On the other hand, data-driven strategies harness signal data from an array of sources to discern faulty components. Within a given environment, sensors are capable of gathering signals from a specific component without interference. In such a setting, the data-driven model captures data anomalies to pinpoint fault locations [14]. Time-frequency domain analysis can seize anomalous signal characteristics and subsequently correlate them with defective parts, even amidst interleaved signal acquisition [15]. Furthermore, statistical analysis has demonstrated efficacy in localizing fault components through the analysis of significant statistical metrics that underscore anomalous behavior [16].

However, fault isolation, being a form of anomaly detection, often grapples with data scarcity in data-driven models. Besides, the physical principles inside machinery often remain nebulous.

B. Graph Embedding

Graph embedding has ascended as a pivotal technique in graph analysis, transforming complex graphs into low-dimensional representations that conserve essential graph information, thereby enabling efficient computation. With the existing triples, the key issue is to define an appropriate score function to evaluate the validity of the triples, which underpins the embedding process.

Historically, graph embedding has relied on observed triples, with translational distance-based score functions emerging as a predominant method. Translating Embeddings (TransE) interprets a relation r as a translating operation between the head entity h and tail entity t, wherein $h + r$ should approximate the tail entity embedding t [17]. This has been expanded upon in various extensions, such as Translating on Hyperplanes (TransH) [18]. Alternatively, the similarity, which gauges the latent semantic meaning of entities and relations, can be computed. RESCAL (RElational SCALing), a three-way model for collective learning, employs a bilinear mapping, assigning each entity a vector representation to encapsulate its latent semantics [19]. Building upon RESCAL, DistMult (Distributed Multinomial) restricts the matrix form

specifically within a symmetric relationship [20]. Likewise, Complex Embedding (ComplEx) broadens the expression of the RESCAL model into the complex domain, targeting anti-symmetric relationships [19]. Further, semantic matching can be orchestrated through neural network architectures [21].

Contemporary methods have incorporated supplementary information to augment the quality of the embedding. A direct strategy considers diverse types of edges and entities [18]. Additionally, textual descriptions associated with the entities can be exploited [22], whereby the text context can serve to initialize the embedding or jointly align the embedding. Another source of usable information lies in logical rules. Typically encompassing $symmetry$, $anti-symmetry$, $inversion$, and $composition$, these rules harbor rich background information, thereby enhancing knowledge acquisition and inference. Consequently, various models strive to refine the score function or amalgamate the rules and knowledge facts to achieve superior embedding [23] [24].

Despite the advancements in graph embedding techniques, prior methods have yet to explore graph embedding within the context of the FG scenario, which is characterized by a plethora of semantic logical rules.

C. Bayesian Network for probability inference

KG facilitates the tracing of failure paths back to their origins. However, inferring the root cause through BN presents a formidable challenge due to the extensive prerequisite knowledge required for computing conditional probabilities. Transforming the FG into the BN emerges as a prospective solution, enabling precise location of faulty components [1].

BN has been applied in probability inference successfully. For instance, a junction tree, an extension of BN, is employed to infer its conditional probabilities. This mechanism clusters similar nodes into a singular "big node", thereby reducing time complexity [25]. Concurrently, a primary barrier in BN lies in estimating each node's independent probability. A noisy-OR gate is employed to learn the conditional probability of associated nodes in situations of data insufficiency [26]. Some systems involve human behavior, and accurately obtaining probabilities related to these behaviors proves challenging. To address this, Zheng et al. [27] proposed a multidimensional analysis model to compute the probabilities of human-related factors. Moreover, a fuzzy BN is incorporated to tackle the system's uncertainty issue, allowing for the identification of a range of conditions rather than a singular condition [28]. This fuzzy BN is better suited for practical application, as determining the specific condition of each node in the BN is a painstaking endeavor. To accommodate dynamic scenarios, dynamic BNs have been proposed to align with temporal signals, where signals correspond to nodes in the system [29].

Nonetheless, the aforementioned BN hinges on adequate data to estimate parameters within the BN. Fault isolation resides in the initial stage of the maintenance process, where data accumulation is limited or even non-existent. This conflict underscores the continuing search for more effective methodologies for fault isolation.

D. Problem Statement

To overcome the above challenges, this study aims to localize the faulty component utilizing BN based on observed patterns, as shown in Fig. 1. In order to construct the BN, denoted as $BNG = (N_1, E_1)$ with initialization, a FG $FG = (N_2, E_2)$ is established to serve as an intermediate stage in the BN architecture, where N denotes the node set and E denote the edge set. This FG is designed to apply the graph embedding method, which provides a vector V for each node. These vectors V are capable of approximating the probabilities within the BNG. Subsequently, using the preliminary BNG and the observed patterns from the sensor data, the faulty component can be inferred through the calculation of conditional probabilities, given by $P(n_i|n_j) = \frac{P(n_i,n_j)}{P(n_j)}$, where n_j refers to observed node mapping in BNG.

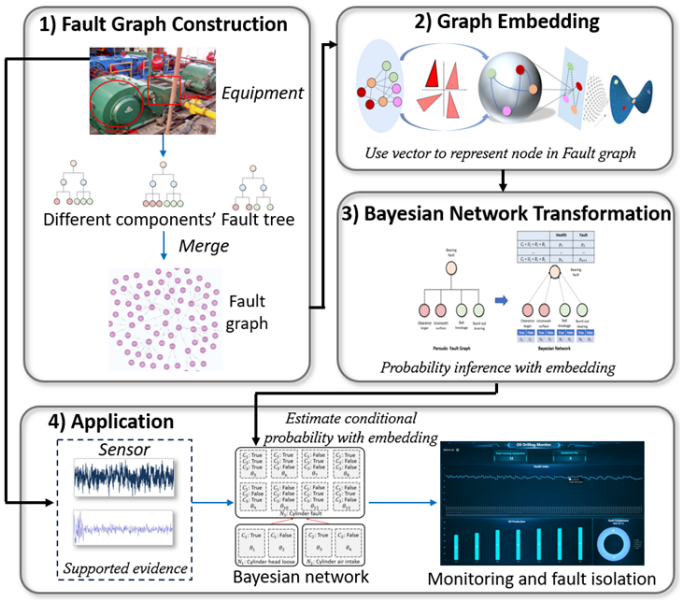

Fig. 1. Overall Framework of the Proposed Method: 1) organizing hierarchical failure relationships with a FG, 2) vectorization via Graph Embedding, and 3) estimation of conditional probabilities in an FG-based BN for 4) failure component inference from sensor evidence.

III. Fault Graph Construction

This section elucidates the process of FG generation, which serves as the foundational architecture for the BN.

A. Pseudo Logical Gate

FG is designed to delineate the connections between a critical event and its causative factors, grounded in the structural and mechanical properties of the equipment. However, identifying logical gates within complex equipment presents a strenuous task due to the synergistic interplay among components. The precise placement and intricate coupling inherent to complex equipment can obfuscate the separation of irrelevant influences in identifying causal relationships. Given the concurrent processes within the equipment, data from collection sensors inevitably amalgamate unrelated fluctuations. Consequently, while the establishment of correlative relationships is

feasible, deciphering the specific logical relation, such as AND or OR relations, poses a considerable challenge [7].

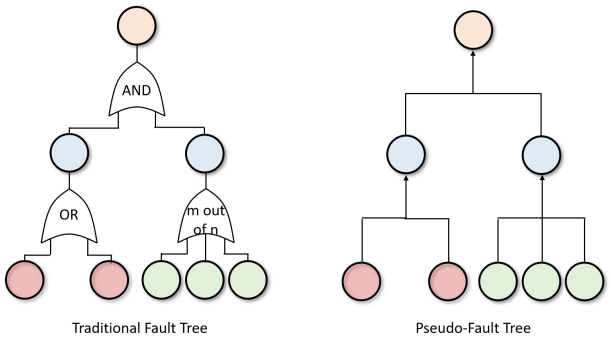

Fig. 2. Convert FT to pseudo-FT

Facilitating the development of the FG necessitates the initial establishment of an FG branch, which can be conceptualized as a pseudo-FT. Unlike a traditional FT, which describes the logical gates between various fault events, a pseudo-logical gate does not necessitate the specification of these gates due to inherent uncertainties. Nevertheless, it preserves the identical event nodes and hierarchical architecture characteristic of the conventional logic gate, as illustrated in Fig. 2. The employment of a pseudo-logical gate as an intermediary stage effectively navigates the uncertainties surrounding the logical gates, thereby laying the groundwork for the subsequent construction of the FG.

B. Fault Graph Initialization

Algorithm 1 Fault Graph Construction

Input: FT_SET: Fault Tree Set;
MS: Mechanical Structure;
Output: FG: Fault Graph

1: $FG \leftarrow\emptyset$; $N \leftarrow\emptyset$
2: **for** FT_i in FT_SET **do**
3: **for** $node_j$ in FT_i **do**
4: $\mathrm{N}_{ij} = Split(\mathrm{node}_j)$
5: //N_{ij} represents different conditions in $node_j$
6: **end for**
7: **for** $node_j$ in FT_i **do**
8: $node_j$'s neighbour $\leftarrow$node$_j$'s first order
9: **for** k in $node_j$'s neighbour **do**
10: **for** v_k in N_{ik} **do**
11: **for** v_j in N_{ij} **do**
12: $FG \leftarrow$build_edge(v_k, v_j)
13: **end for**
14: **end for**
15: **end for**
16: **end for**
17: **end for**
18: **for** $connected\_pair$ in MS **do**
19: $FT_m, FT_n \leftarrow Mapping(connected\_pair)$
20: $v_{top\_m} \leftarrow Top(FT_m)$, $v_{top\_n} \leftarrow Top(FT_n)$
21: //Top uses to retrieve the component fault's node
22: $FG \leftarrow$build_edge$(\mathrm{v}_{top_m}, v_{top_n})$
23: **end for**

Embedding can offer valuable insights into the meanings of logic gates. Specifically, edges in the FG may be characterized by varying types or weights, and their embedding expresses the gate properties inherently and semantically. For instance, in an AND relation, the simultaneous occurrence of all child events leads to a causative event. This relationship can be represented in the FG through a weighted probability relation (e.g., "may lead to") between the child events and the causative event. If all child events are active (have occurred), this results in a sufficiently high probability of the causative event. Consequently, this KG can satisfy fault event inference, providing more robust evidence and facilitating precise judgment.

In this article, the term FG refers to a specific type of KG that describes relationships between fault events. Complex equipment comprises many parts or components inextricably intertwined, indicating interactive relationships. Therefore, a specific fault event represented in FT form, which constitutes a branch in the FG, may be related to other failure events. Based on the understanding of mechanical mechanisms, different FTs can be combined into an FG through their interactions and structure, as depicted in Fig. 3.

Algorithm 1 delineates the procedure for constructing the FG. Lines 2 to 17 aim to expand each node in the FT into multiple nodes representing diverse conditions, thereby providing distinct graph embeddings for various signal patterns. The *Split* function, elaborated in line 4, is employed to divide a single fault description into a range of possible scenarios. The connections between these newly created nodes are established in lines 7 to 16, based on the relationships in the original FT. N_{ik} stands for the different conditions or states in node k. Because these nodes are connected to each other, the conditions in one node should be linked with the conditions in the next node (for loops in line 9 to 11).

Lines 18 to 23 focus on connecting different FTs, leveraging prerequisite knowledge of the mechanical structure. Each connected pair expresses two different connected components, thereby indicating a connection between their respective FTs. The proposed method incorporates the *Top* function to extract the top node in the FT (line 20), which symbolizes the most representative fault of the component. The top nodes, extracted from multiple FTs, are then interconnected to expand the FT into an FG (line 22).

Taking gearbox as an example, the top of its FT is denoted as "Gearbox Failure". The secondary layer outlines distinct failure modes, such as "Insufficient Lubrication", "Overload", and "Abnormal Vibration", which stand as separate entities without direct interlinks. The "Abnormal Vibration" category is further subdivided into conditions like "Excessive Vibration", "Insufficient Vibration", and "Normal Vibration". In parallel, the "Gearbox Failure" node is differentiated into conditions indicative of "Gearbox Malfunction" and "Gearbox Health". This framework dictates that various conditions within a node are treated as individual nodes in FG, subsequently linked to their respective upstream node (lines 7-8). For example, each vibration condition must be linked to both "Gearbox Malfunction" and "Gearbox Health" (see lines 9-11). In the analysis of complex equipment, the relationships between the gearbox and its associated components are

delineated based on the system's underlying principles and mechanical configuration (line 18-22).

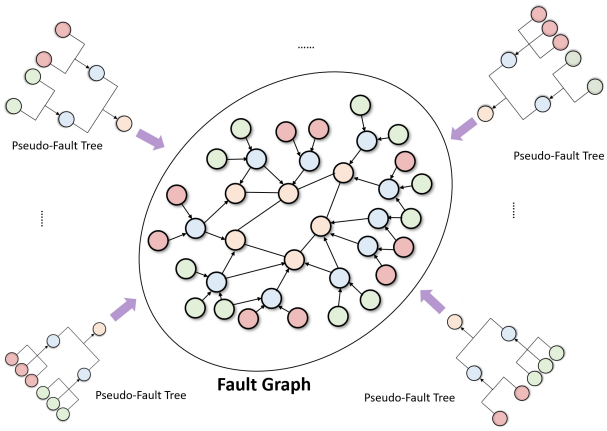

Fig. 3. Integrate pseudo-FTs to FG

IV. Multi-Field Hyperbolic embedding

A. Hyperbolic embedding

The FG is a kind of hierarchical organization, where extensive research has ascertained that any hierarchical structure can be accurately represented in hyperbolic space with minimal errors [30]. Given the logical and causal relationships inherent to the FG, this paper introduces MFH embedding, which is especially apt for inferring complex rules in the FG, as exemplified in Fig. 4.

In this study, vector multiplication is defined as the execution of an operation in tangent space through Poincaré Ball projection. Specifically, the exponential projection exp_0^c is responsible for mapping a data point from the Euclidean tangent space x^E into the hyperbolic space x^H. Conversely, the logarithmic projection log_0^c converts hyperbolic space x^H to tangent space x^E, which can be shown as follows:

$$exp_0^c(x) = tanh(\sqrt{c}\,\|x\|)\frac{x}{\sqrt{c}\,\|x\|}, \tag{1}$$

$$log_0^c(x) = arctanh(\sqrt{c}\,\|x\|)\frac{x}{\sqrt{c}\,\|x\|}, \tag{2}$$

in which *tanh* denotes the hyperbolic tangent function, and *arctanh* signifies the inverse hyperbolic tangent. The term c corresponds to the curvature parameter within the Poincaré Ball model. Both functions are employed during the rotational component of the process.

B. Hyperbolic transformation

Utilizing embedding method offers a robust approach to capturing logical relationships through parameterized spatial manipulations. Unlike conventional transformation techniques that typically focus on a select few rules, the FG approach integrates a broad spectrum of logical symbols. By amalgamating multiple transformation strategies, this approach strives to encompass a comprehensive array of relationships inherent to diverse logical systems, including *rotation*, *reflection*, and *translation*. Specifically, *rotation* addresses compositional or anti-symmetric patterns, and *reflection* encodes symmetric patterns. The following equations show the definition of *rotation*, *reflection*, and *translation*, respectively.

$$Rot_{(\Theta_r)} = Diag(G^+(\theta_1), ..., G^+(\theta_{\frac{d}{2}})), \tag{3}$$

$$Ref_{(\Phi_r)} = Diag(G^-(\phi_1), ..., G^-(\phi_{\frac{d}{2}})), \tag{4}$$

$$Tra_{(\Omega_r)} = Diag(G^0(\omega_{a,1}, \omega_{b,1}), ..., G^0(\omega_{a,\frac{d}{2}}, \omega_{b,\frac{d}{2}})), \tag{5}$$

here, $\Theta_r := (\theta_i)_{i\in 1,...,\frac{d}{2}}$, $\Phi_r := (\phi_i)_{i\in 1,...,\frac{d}{2}}$, and $\Omega_r := (\omega_{a,i}, \omega_{b,i})_{i\in 1,...,\frac{d}{2}}$, correspond to parameters that are specific to the relationship. Additionally, d is chosen to be an even integer, ensuring that the block-diagonal matrix retains hyperbolic properties. Moreover, the terms G^+, G^-, and G^0, as seen in Eq. 3-5, are delineated as:

$$G^{\pm}(\phi) = \begin{bmatrix} \cos(\phi) & \mp\sin(\phi) \\ \sin(\phi) & \pm\cos(\phi) \end{bmatrix}, \tag{6}$$

$$G^0(\theta_a, \theta_b) = \begin{bmatrix} \theta_a & \theta_b \\ \theta_a & \theta_b \end{bmatrix}. \tag{7}$$

The adoption of rotational mechanisms in complex vector spaces has demonstrated efficacy for encoding compositional structures, as highlighted in the literature [31]. These rotations are adept at capturing and inferring various patterns, including inverse relations, compositions, and both symmetrical and asymmetrical associations. Furthermore, reflections are instrumental in deducing the missing elements of symmetrical triplets by inverting them across defined axes. Translations, on the other hand, facilitate the positional shifting of embeddings within the same dimensional framework, preserving their geometric properties.

Incorporating these three principal geometrical operations, each with its distinct parameters, facilitates an exhaustive investigation of the intricate logical connections present. In this research, specific transformation configurations are chosen, namely $Rot_{(\Theta_r)}$, $Rot_{(-\Theta_r)}$, $Ref_{(\Phi_r)}$, $Ref_{(-\Phi_r)}$, and $Tra_{(\Omega_r)}$. It is noted that the linear proportionality between the columns (or rows) of $Tra_{(\Omega_i)}$ and those in another $Tra_{(\Omega_j)}$, indicating a structured relationship within the translation-based transformations, thereby rendering the proposal of $Tra_{(-\Omega_r)}$ unnecessary.

C. Hyperbolic attention

Different classes of hyperbolic isometries may have different merits. A hyperbolic attention model is proposed to synthesize integrated isometries, capable of capturing a mix of relationships in a unified manner. Before calculating the attention score, the embedding should be mapped from hyperbolic representation to tangent space representation, which is $x^E = log_0^c(x^H)$. Then different weights α can be normalized by $Softmax$ function:

$$\begin{aligned}(\alpha_{x_1}, \alpha_{x_2}, \alpha_{x_3}, \alpha_{x_4}, \alpha_{x_5}) = Softmax(a^T x_1^E, a^T x_2^E,\\ a^T x_3^E, a^T x_4^E, a^T x_5^E),\end{aligned} \tag{8}$$

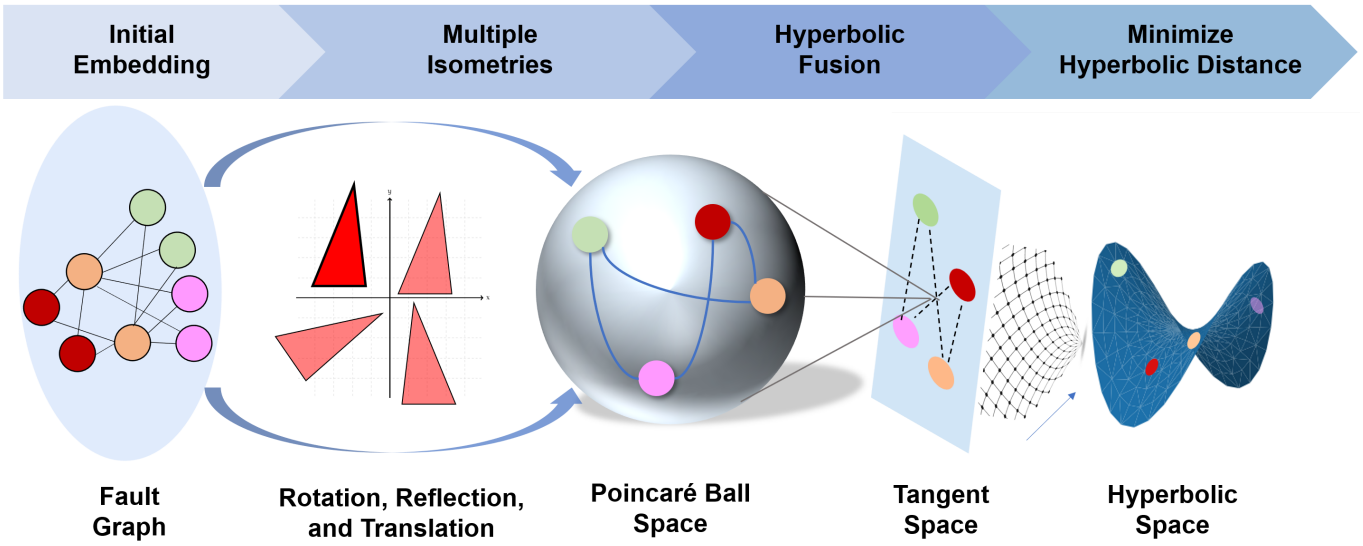

Fig. 4. Architecture of Multi-Field Hyperblic Embedding

where x_1^E to x_5^E represents the embedding of $Rot_{(\Theta_r)}$, $Rot_{(-\Theta_r)}$, $Ref_{(\Phi_r)}$, $Ref_{(-\Phi_r)}$ and $Tra_{(\Omega_r)}$, respectively. Besides, a stands for trainable vector using in the following attention mechanism. With the obtained weight coefficient, the following step is to calculate the attention score accompanied by transforming from tangent space to hyperbolic space:

$$Att(x_1^H, x_2^H, x_3^H, x_4^H, x_5^H; a) = exp_0^c(\alpha_{x_1} x_1^E, \alpha_{x_2} x_2^E, \alpha_{x_3} x_3^E, \alpha_{x_4} x_4^E, \alpha_{x_5} x_5^E), \tag{9}$$

where α is the weight obtained in Eq. 8. This attention mechanism combines different representations of the head entity initially. To consider the information involved in relation, relation embedding is then integrated into a hyperbolic transformation as shown below:

$$Q(h,r) = Att(x_1^H, x_2^H, x_3^H, x_4^H, x_5^H; a_r) \oplus^c r_r^H, \tag{10}$$

where r_r^H represents the relation embedding, $\oplus^c$ is the vector addition in hyperbolic space, the rest of the notations have occurred in previous equations.

In earlier studies, geometrical operations such as *rotation*, *reflection*, and *translation* have been established as adept in capturing the logical patterns typical of tree-like hierarchical structures. Subsequently, the computation of the hyperbolic distance is conducted using the predicted embedding along with the corresponding tail embedding, a critical step that will be incorporated into the evaluation function. The formulation of the scoring function is presented below:

$$s(h,r,t) = -d^c(Q(h,r), e_t^H)^2 + b_h + b_t, \tag{11}$$

where b_h represents the bias associated with the head entity, and b_t denotes the bias linked to the tail entity. These biases serve as the boundaries within which the scoring function operates. The term d^c is utilized to calculate the hyperbolic distance, as specified by the following equation:

$$d^c(x,y) = \frac{2}{\sqrt{c}} arctanh(\sqrt{c} \| -x \oplus^c y \|). \tag{12}$$

The scoring function is structured to assess the legitimacy of a triplet by aiming to minimize summation of h and r, subtracted by t, within the embedding space. This is reflected in Eq. 11, where a greater hyperbolic distance d^c yields a diminished score. This implies a preference for triplets that exhibit smaller distances.

Ultimately, the suggested MFH embedding methodology undergoes a training regimen that is geared toward minimizing the overall cross-entropy loss. This is achieved by employing a strategy of uniform negative sampling during the optimization:

$$\mathcal{L} = \sum_{t' \in u(v)} log(1 + exp(y_{t'} s(h,r,t'))), \tag{13}$$

where $y_{t'}$ takes the value of -1 in cases where t' is equal to t, and is set to 1 for all other instances. The minimization of the loss function is integral to aligning the node and edge embeddings as closely as possible with the logical rules set forth by our proposal.

V. Bayesian Network Inference

The principal aim of this research is to perform probabilistic inference of faults using the FG framework. To achieve this, BN is utilized to enable the probabilistic deduction process. However, the FG cannot be directly applied to the BN without undergoing a specific preliminary transformation. As depicted in Fig. 5, the FG is situated on the left, while the desired outcome, the BN, is on the right. It is crucial to recognize that each node in the BN can encompass multiple conditions, such as “Noise: High”, “Noise: Middle”, “Noise: Low”, and “Noise: None”. To transform the FG into the BN expression, it must be converted into a format where each node is associated

with its respective conditions. This fundamental transformation process is executed in Section III, Algorithm. 1.

The initial challenge involves generating supplementary nodes for the construction of the BN. Within BN, each node can embody multiple conditions, with a particular condition being expressible as a conditional probability $\theta = P(X = x_i|Parent(X) = PA\_Set)$, where x_i indicates the specific condition of node X, and PA_Set represents the trigger (reason) sets of x_i. It is crucial to highlight that the FG only takes into account a specific condition of a single node, which may result in an insufficient representation in the BN. In the original FG, it becomes necessary to unfold each node into multiple scenarios to accommodate various configurations. For instance, two separate nodes, "Cylinder head loose", and "Cylinder air intake", can cover all possible situations. Through this augmentation, the FG is thus convertible into the BN as illustrated in Fig. 6. It is noted that θ_x represents the probability of each condition.

Once the BN structure is well-constructed and the FG is aptly embedded, BN inference can be applied to deduce the conditional probability of diverse occurrences. Unlike traditional FG, BN inference necessitates information about event probability, which is absent in the original FG. To overcome this constraint, the proposed method computes the event probability based on the fine-tuned embedding, serving as an approximation of the true probability, using the following equation [32]:

$$P(X = x_{ik}) = \frac{\frac{1}{\omega}\sum_{j=1}^{\omega} V(x_{ik}, x_{ik}^{j})}{\frac{1}{\eta}\sum_{m=1}^{\eta}\frac{1}{\omega}\sum_{j=1}^{\omega} V(x_{im}, x_{im}^{j})}, \tag{14}$$

where x_{ik} denotes the k^{th} value of the variable x_i. The term ω represents the count of neighboring values for x_{ik}, whereas x_{ik}^{j} symbolizes the j^{th} neighboring node in relation to x_{ik}. The symbol η quantifies the total number of possible values for x_i. As an illustrative example, if x_i stands for *pump speed*, then x_{i1} might represent *pump speed normal*, x_{i2} could signify *pump speed excessive*, and x_{i3} might indicate *pump speed deficient*. Moreover, V quantifies the magnitude of correlation between x_{ik} and its neighbor x_{ik}^{j}, which is expressed as follows:

$$V(x_{ik}, x_{ik}^{j}) = \frac{1}{dis(x_{ik}, x_{ik}^{j})}, \tag{15}$$

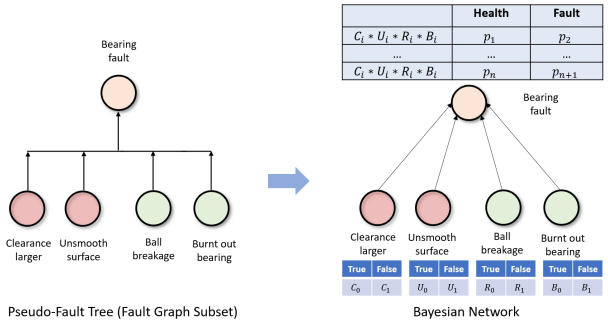

Fig. 5. Convert pseudo-FT (FG Subset) to BN

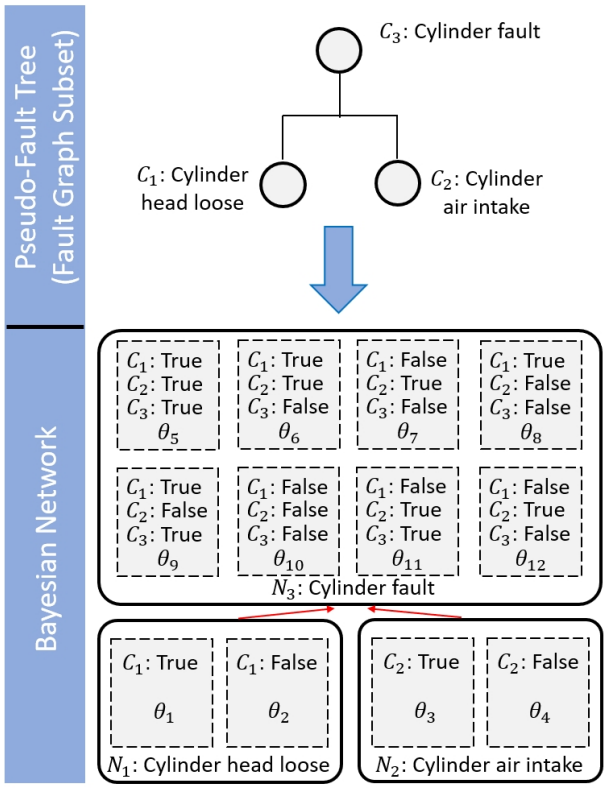

Fig. 6. BN structure based on pseudo-FT (FG Subset)

where dis denotes the distance measured in the Euclidean space.

Calculating the joint probability of a pair of events forms the bedrock of BN inference. Three distinct kinds of dependency relations that can characterize the joint probability $P(X = x_i, Y = y_i)$: $independent$, $causal$, and $share\ ancestor$. Initially, the $independent$ relationship denotes that the occurrence of event x_i is not contingent upon event y_i. The mathematical representation for this relationship is as follows:

$$P(X = x_{ik}, Y = y_{ik}) = P(X = x_{ik}) * P(Y = y_{ik}). \tag{16}$$

Moreover, the term $causal$ relationship is used to signify that one event is a predecessor of another. For instance, should y_i be a precursor to x_i, the joint probability in this scenario would be articulated as:

$$P(X = x_{ik}, Y = y_{ik}) = P(Y = y_{ik}) * \sum_{l=1}^{n^{y_i}} \frac{V(x_{ik}, y_{il})}{\sum_{p=1}^{N} V(x_{ip}, y_{il})}, \tag{17}$$

in which n^{y_i} denotes the count of nodes associated with y_i, while N reflects the total number of potential outcomes for the variable X. Furthermore, the concept of $share\ ancestor$ applies to two events that have a common predecessor, identified here as Z. In such a case, the expression for the joint

probability is formalized as:

$$P(X=x_{ik},Y=y_{ik})=\sum_{l=1}^{N_z}P(Z=z_l)*\sum_{p=1}^{n^{z_i}}\frac{V(x_{ik},z_{ip})}{\sum_{q=1}^{N}V(x_{iq},z_{ip})}$$
$$*\sum_{p=1}^{n^{z_i}}\frac{V(y_{ik},z_{ip})}{\sum_{q=1}^{N}V(y_{iq},z_{ip})}, \tag{18}$$

where the term $\sum_{l=1}^{N_z}P(Z=z_l)$ is utilized to aggregate the probabilities across all pertinent nodes of the variable Z. Drawing from Bayes' theorem, the computation of conditional probability is given by the following equation:

$$P(A|B)=\frac{P(A,B)}{P(B)}. \tag{19}$$

When incorporating multiple pieces of evidence. The estimation of the conditional probability is expressed by the subsequent formula:

$$P(A|B,C...Z)=\frac{P(A,B,...,Z)}{P(B,...,Z)}. \tag{20}$$

As a result, any malfunction in the complex equipment will be reflected in the FG. This evidence will be used to search the fault components with the most fault probability. To better illustrate the proposed method, a unified calculation process is provided as Algorithm 2:

Algorithm 2 Unified Calculation Process for Graph Embedding-Based Bayesian Network

Input: FT_SET: Fault Tree Set;
MS: Mechanical Structure;
Output: Fault component with probability
1: $FG \leftarrow Algorithm\ 1(FT\_SET, MS)$;
2: $V_{FG} \leftarrow Multi-Field\ Hyperbolic\ Embedding(FG)$;
3: Transfer FG into BNG;
4: Approximate single event probability of each node from V_{FG} by Eq.(15);
5: Approximate conditional probability of each component from V_{FG} by Eq.(16-20);

VI. Case Study

Oil-drilling equipment, a piece of complex equipment composed of multi-level components with hierarchical failure relationships, is examined as a case study, depicted in Fig. 7.

The experimental setup is bifurcated into two segments: the embedding of FG (Section IV) and the inferential process of BN (Section V). FG serves as the semantic foundation for graph-based reasoning, constructed from unstructured and scattered failure relationships. When the FG is converted into a BN and combined with empirical evidence from monitoring, it is capable of pinpointing the component most likely to fail. To assess the efficacy of FG embedding, a link prediction task is utilized, owing to its practical advantage of not requiring extra labeling for evaluation. Considering the challenge of obtaining a sufficient number of labels for affirming the outcomes of inference, the BN inference phase is evaluated via an unsupervised approach. It is critical to emphasize that the foundation of the FG is a key element in both sets of experiments, which is detailed in Section III.

Fig. 7. Oil drilling equipment

A. Oil drilling equipment-based FG construction

The oil drilling equipment is composed of multiple mechanical elements, for example, the liquid end and power end. These units link to additional subordinate components forming a tiered architecture. In line with the current FT and mechanical configurations, distinct FTs can be interlinked to form an FG by connecting nodes at the highest hierarchy, as shown in Fig. 3. In this FG, each node symbolizes a distinct malfunction, with a range of potential conditions. Take "cross head aging" node as an example, it might encapsulate several states like "cross head: wear", "cross head: scratch", and "cross head: health". To clarify this node's inherent vagueness, it is decomposed into several nodes, each delineating a precise state, and new connecting edges are introduced. The refined graph is then archived in Neo4j, a standard graph-oriented database [33], as depicted in Fig. 8. This database arrangement enables the swift retrieval of pertinent nodes for conducting BN inference.

B. Embedding Experiment Setting

1) Evaluated Metrics: The evaluated metrics for the link prediction experiment are shown below:

$$Hit@n=\frac{1}{N}\sum_{i=1}N(rank_i\leq n), \tag{21}$$

where the term *rank* denotes the position of a particular item within a ranking sequence, while *(.)* signifies the inclusion of the desired item within the anticipated list of rankings. For the purposes of evaluation in this study, the metrics Hit@1, Hit@3, and Hit@10 are employed, with their respective definitions referenced in Eq. 21. Additionally, the Mean Reciprocal Rank (MRR), a holistic measure, is utilized as it accentuates the overall sequence order:

$$MRR=\frac{1}{N}\sum_{i=1}N(\frac{1}{rank_i}). \tag{22}$$

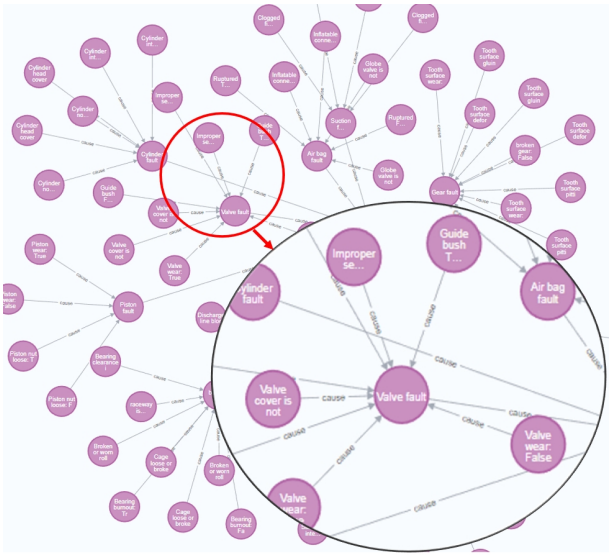

Fig. 8. Fault Graph in Neo4j

2) Model Settings: Ensuring the efficiency and stability of the introduced model, key parameter configurations are detailed in Table I. It is important to note that existing edges are treated as positive instances, whereas edges that are absent are considered negative instances. Meanwhile, the model undergoes an embedding initialization using the word2vec [34].

TABLE I
KEY HYPERPARAMETER IN MFH

Hyperparameter	Value	Hyperparameter	Value
Dropout rate	0.1	Batch size	64
Learning rate	0.001	Epoch	1000
Embedding size	256	Optimize function	Adagrad [35]

Cutting-edge models for comparison: to prove the superiority of the proposed MFH model, other cutting-edge models are adopted to conduct the link prediction task, they are: Complex embedding (ComplEx) [11], TransE [17], Multi-relational poincaré graph embeddings (MurE) [36], Canonical tensor decomposition (CP) [37], Relational rotation embedding (RotE) [31], Hyperbolic embedding (ATTH) [38].

C. Embedding Performance

The task of predicting links is employed as a practical means to assess the quality of embeddings generated by the proposed MFH, with the understanding that a well-representative embedding is more likely to accurately predict links. In this research, the performance of the newly proposed MFH model is compared with that of existing models in the realm of link prediction, using a variety of datasets that are both publicly available and proprietary, including the FG introduced in this study.

Table. II illustrates the performance of various models. For the FB15K dataset [39], sourced from the Freebase Knowledge Base, the model presented in this study achieved the top score in terms of MRR and Hit@1, while it ranked second in both Hit@3 and Hit@10 metrics. Additionally, this model scored highest in Hit@10 when evaluated on the WN18RR dataset, which is derived from WordNet [40]. It is worth noting that while the model showcased strong performance, it did not outperform all others on the WN18RR dataset.

In the context of the FG dataset, the strengths of the suggested model are more evident. It achieves the highest scores in MRR, Hit@1, and Hit@10, while taking the second spot in Hit@3. The model demonstrates a clear advantage compared to its counterparts. For example, the model's Hit@10 score is 0.835, outpacing the runner-up model, which scores 0.824. However, it's important to note that the model does not significantly outdo the second-best model in other datasets, with a marginal difference in scores (0.453 against 0.452 in FB15K, and 0.558 against 0.554 in WN18RR).

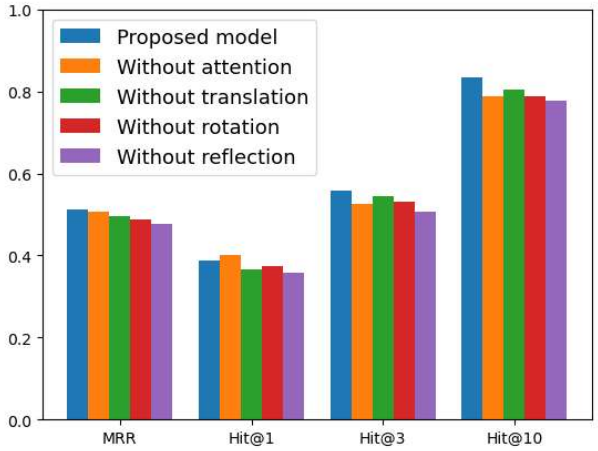

Fig. 9. Ablation study of MFH Embedding

To ascertain the contributions of distinct modules, an ablation study is conducted, as represented in Fig. 9. In this study, the four primary components—attention, translation, rotation, and reflection—are sequentially removed, each denoted with the prefix "Without". Overall, Fig. 9 reveals that these four scenarios underperform compared to the proposed model, with the exception of Hit@1, where the proposed model ranks second. The results of this ablation study underscore that the proposed model, inclusive of all tailored modules, yields the highest performance, thereby buttressing the individual contributions of various modules and the overall efficacy of the MFH's architecture.

The loss curves of the training and testing datasets are provided to validate that the model has been appropriately trained. Both these loss curves converge swiftly, as depicted in Fig. 10. During the initial 10 epochs, the model's learning phase is marked by pronounced fluctuations in both training and testing loss, a result of substantial weight adjustments introducing loss variability. Following this period, the training loss not only declines but also stabilizes beneath the testing loss, a trend in line with expectations, as the model is directly adapting to the training data without optimization against the testing set. Despite this progress, persistent undulations in the testing loss hint at the model's imperfect generalization to unseen data.

TABLE II
LINK PREDICTION RESULT IN OPEN DATASET AND PROPOSED FAULT GRAPH

Model	FB15K				WN18RR				FG (our dataset)			
	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
ComplEx	0.255	0.178	0.277	0.411	0.378	0.365	0.382	0.402	0.450	0.303	0.500	0.777
TransE	0.238	0.151	0.263	0.414	0.173	0.29	0.259	0.447	0.496	0.351	0.569	0.824
MurE	**0.286**	0.201	**0.309**	**0.453**	0.421	0.35	0.459	0.554	0.483	0.342	0.55	0.798
CP	0.259	0.182	0.281	0.414	0.383	0.37	0.388	0.407	0.332	0.235	0.324	0.611
RotatE	0.284	0.201	0.308	0.452	**0.448**	**0.392**	**0.476**	0.551	0.501	0.378	0.543	0.830
ATTH	0.284	0.201	**0.309**	0.452	0.42	0.353	0.452	0.544	0.511	0.378	**0.571**	0.796
Ours	**0.286**	**0.204**	0.308	0.452	0.416	0.35	0.453	**0.558**	**0.513**	**0.388**	0.559	**0.835**

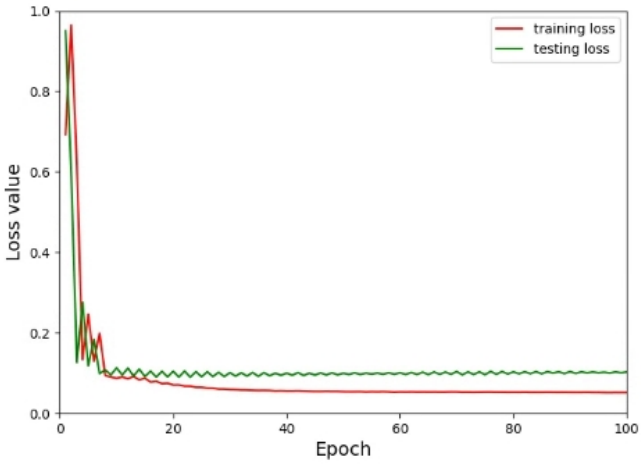

Fig. 10. Loss curve of MFH embedding

D. Bayesian Inference Performance

The effectively trained embedding can enhance the process of fault localization by applying Bayesian inference. As described in Section V, the likelihood of each initiating event can be determined in light of the observed events.

The embedding initialized BN accurately pinpoints fault locations given the available evidence. The FG provides a solid base (graph embedding) for calculating posterior probabilities. The FT also proves capable of embedding nodes effectively, though it typically deals with graphs of a smaller scale. Therefore, it's crucial to compare the effectiveness of FG and FT. Fig. 11 and 12 illustrate a comparison of the posterior probability of fault events in the BN, using the embeddings derived from both FG and FT. This evaluation is conducted on two critical components of the proposed oil drilling equipment, namely the Power end and the Liquid end, respectively.

Fig. 11 presents an analysis of the Power end, demonstrating that as the quantity of evidence increases, there is a corresponding rise in likelihood. Additionally, the posterior probability derived from the FG surpasses that originating from the FT. The larger the amount of evidence, the wider the gap between both probabilities. Similarly, Fig. 12 echoes the same finding, albeit the gap is not as pronounced as in Fig. 11. Despite the probabilities being virtually identical when the number of provided evidence is two in Fig. 12, all other scenarios have displayed a significant lift from FT to FG. A plausible explanation for this could be that the FG, given its larger graph size, provides a sufficient training context for graph embedding.

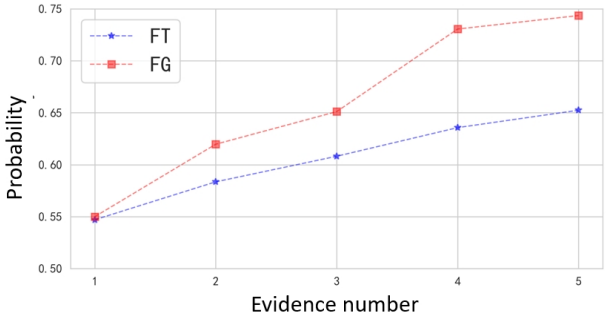

Fig. 11. Power end fault probability with evidence numbers

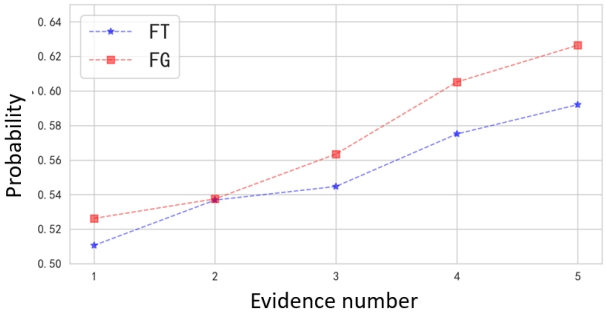

Fig. 12. Liquid end fault probability with evidence numbers

In the realm of fault isolation tasks, obtaining a substantial volume of practical labels poses a significant challenge. Consequently, this study makes use of a limited number of available labels to evaluate the performance of the proposed methods. It is noteworthy that the proposed method is designed to operate

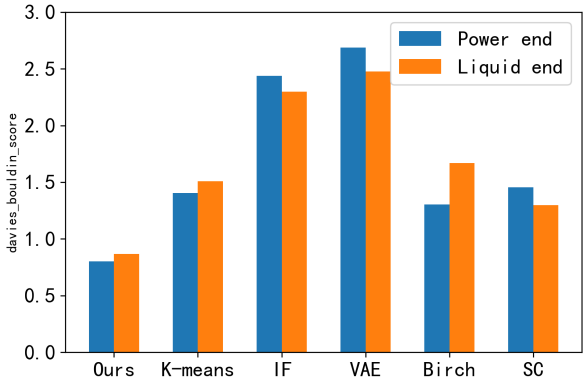

Fig. 13. Unsupervised learning comparison in davies bouldin score metrics

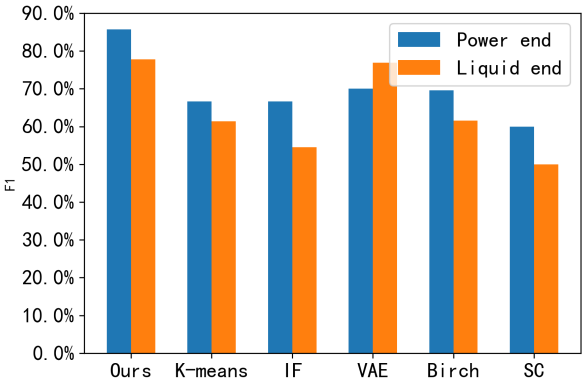

Fig. 14. Unsupervised learning comparison in F1 metrics

without the need for labeled data. As such, state-of-the-art unsupervised learning methods, including K-means [41], Isolate Forest (IF) [42], Variational Autoencoder(VAE) [43], Birth [44], and Spectral clustering (SC) [45], are employed as comparative models. The actual labels are used in this experiment to evaluate the outputs of these methods.

From the perspective of machine learning, the proposed method falls under the category of unsupervised learning. Therefore, the *davies bouldin score*, defined as the average similarity measure between each cluster and its most similar one, is employed to assess the performance of the proposed method. Fig. 13 reveals that the proposed model achieves the lowest score, signifying that the distances within clusters are small, while the distances between clusters are relatively large.

To further substantiate our method, a limited number of labels are gathered to assess its performance in a classification task. Thus, the F1 score is applied as a metric, where $F1 = \frac{2*Recall*Precision}{Recall+Precision}$. Fig. 14 illustrates the performance of the proposed model and other state-of-the-art unsupervised learning models in the Power end and Liquid end, respectively. The results show that the BN embedding-based inference outperforms other models in both the Power end and the Liquid end. A plausible explanation for this could be the integration of fault event-causal relations as prior knowledge in the proposed BN embedding-based inference, which narrows the search space and results in a more optimal outcome.

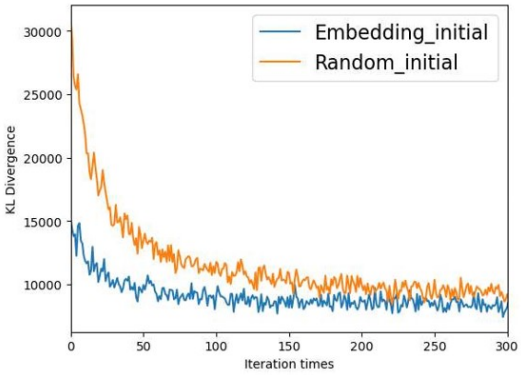

Fig. 15. Power end fault distribution comparison

One of the principal advantages of this method is that graph embedding can approximate the ground truth conditional probability within the BN, even in the absence of genuine data

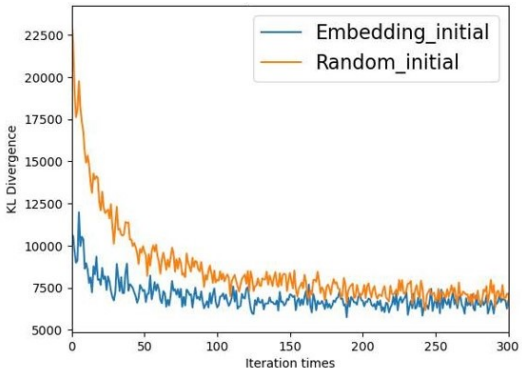

Fig. 16. Liquid end fault distribution comparison

support, as graph embedding arises from the semantic meaning of the FG. To evaluate the efficacy of graph embedding-based conditional probability within the BN, the training processes of BN initialized from both graph embedding-based and random initialization are compared. The training processes will be contrasted relative to the variance in fault distribution, where a minimal amount of labeled data will be inputted into the BN to update the conditional probability, approximating the ground truth distribution. Throughout the training process, the fault distribution will adjust accordingly. This adjusted fault distribution can then be compared with the ground truth distribution using Kullback-Leibler Divergence (KL-Divergence), a measure of disparity between two distributions. As the value of KL-Divergence increases, so does the difference between the two distributions.

Fig. 15 and 16 display the fault distribution for the power end and liquid end, respectively. All the curves in both graphs converge as the number of iterations increases, bringing both distributions closer to the ground truth distribution. Moreover, the blue curves exhibit relatively low KL-Divergence values at the initial stage, signifying that the distribution from the graph embedding-based initial embedding is closer to the ground truth distribution. Compared to the curve from random initialization, the graph embedding-based approach converges more quickly to the ground truth distribution. This accelerated convergence assists engineers and operators in locating faults, especially when the available data is insufficient.

Fig. 17. Oil drilling monitor interface, where the pie chart indicates the fault probability of different components

To apply this proposed method in the monitoring of oil drilling equipment, a maintenance-oriented monitoring plat-

form is developed for the practical operation of the equipment. Fig. 17 presents a visual depiction of the platform. The upper part features a temporal evaluation of the health index, providing a dynamic overview of equipment health. The lower part, represented by a pie chart, showcases the distribution of fault probabilities, with a particular emphasis on the component most likely to exhibit a fault. This platform thus provides an interactive and intuitive interface for ongoing equipment monitoring and maintenance.

VII. Conclusion

Fault isolation often necessitates the integration of various forms of expert knowledge for accurate determination, particularly in the face of data scarcity. However, such prior knowledge is occasionally incomplete or obfuscated by complex structures, making its efficient utilization especially challenging for complex equipment. Addressing this gap, the paper presents a methodology that arranges varied knowledge specific to the domain into the FG structure. This is utilized to pinpoint faulty components by leveraging an embedding-based BN inference. This approach addresses the issue of conducting quality assessments on accumulated fault semantics when labeled data is scarce. As far as we are aware, this is the first application of graph embedding in fault isolation.

Although this study introduces a new approach, there are some limitations to consider. First, the shortage of labeled data means unsupervised learning is used for comparison with inference results instead of supervised learning. Secondly, the FG necessitates that the equipment possesses a hierarchical structure of failure modes, which may not be applied to equipment with a simple mechanical structure. Looking ahead, future studies could look into creating a detailed FG that can suggest appropriate maintenance actions for identified faults. The proposed method could also be refined to tackle temporal challenges by modeling inter-temporal relationships, drawing on Dynamic Bayesian Network principles to effectively capture time-dependent dynamics.

References

[1] L. Xia, P. Zheng, K. Keung, C. Xiao, T. Jing, and L. Liu, "From fault tree to fault graph: Bayesian network embedding-based fault isolation for complex equipment," *Manufacturing Letters*, vol. 35, pp. 983–990, 2023.

[2] E. Zio, "Prognostics and health management methods for reliability prediction and predictive maintenance," *IEEE Transactions on Reliability*, vol. 73, no. 1, pp. 41–41, 2024.

[3] J. Guo, Z. Li, and M. Li, "A review on prognostics methods for engineering systems," *IEEE Transactions on Reliability*, vol. 69, no. 3, pp. 1110–1129, 2019.

[4] Y. Xi, Y. Cui, X. Tang, Z. Li, and X. Zeng, "Fault location of lightning strikes using residual analysis based on an adaptive kalman filter," *IEEE Access*, vol. 7, pp. 88126–88137, 2019.

[5] L. Xia, Y. Liang, P. Zheng, and X. Huang, "Residual-hypergraph convolution network: A model-based and data-driven integrated approach for fault diagnosis in complex equipment," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1–11, 2022.

[6] M. Chen, H. Shao, H. Dou, W. Li, and B. Liu, "Data augmentation and intelligent fault diagnosis of planetary gearbox using ilofgan under extremely limited samples," *IEEE Transactions on Reliability*, vol. 72, no. 3, pp. 1029–1037, 2023.

[7] C. Zhou, Q. Chang, H. Zhao, M. Ji, and Z. Shi, "Fault tree analysis with interval uncertainty: a case study of the aircraft flap mechanism," *IEEE Transactions on Reliability*, vol. 70, no. 3, pp. 944–956, 2020.

[8] C. Zhu, Y. Jiang, G. Liu, and T. Zhang, "Integration frameworks and intelligent research in dynamic fault tree: A comprehensive review and future perspectives," *Quality and Reliability Engineering International*, vol. 39, no. 7, pp. 3157–3178, 2023.

[9] A. Rauzy and C. Blériot-Fabre, "Towards a sound semantics for dynamic fault trees," *Reliability Engineering & System Safety*, vol. 142, pp. 184–191, 2015.

[10] L. Xia, Y. Liang, J. Leng, and P. Zheng, "Maintenance planning recommendation of complex industrial equipment based on knowledge graph and graph neural network," *Reliability Engineering & System Safety*, vol. 232, p. 109068, 2023.

[11] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *International conference on machine learning*, pp. 2071–2080, PMLR, 2016.

[12] L. Xia, P. Zheng, X. Li, R. X. Gao, and L. Wang, "Toward cognitive predictive maintenance: A survey of graph-based approaches," *Journal of Manufacturing Systems*, vol. 64, pp. 107–120, 2022.

[13] Z.-X. Zhang, X.-S. Si, C.-H. Hu, X.-X. Hu, and G.-X. Sun, "An adaptive prognostic approach incorporating inspection influence for deteriorating systems," *IEEE Transactions on Reliability*, vol. 68, no. 1, pp. 302–316, 2018.

[14] A. He and X. Jin, "Deep variational autoencoder classifier for intelligent fault diagnosis adaptive to unseen fault categories," *IEEE Transactions on Reliability*, vol. 70, no. 4, pp. 1581–1595, 2021.

[15] O. Naidu and A. K. Pradhan, "A traveling wave-based fault location method using unsynchronized current measurements," *IEEE Transactions on Power Delivery*, vol. 34, no. 2, pp. 505–513, 2018.

[16] S. A. Rocha, T. G. Mattos, R. T. Cardoso, and E. G. Silveira, "Applying artificial neural networks and nonlinear optimization techniques to fault location in transmission lines—statistical analysis," *Energies*, vol. 15, no. 11, p. 4095, 2022.

[17] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," *Advances in neural information processing systems*, vol. 26, 2013.

[18] Q. Wang, Z. Mao, B. Wang, and L. Guo, "Knowledge graph embedding: A survey of approaches and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 12, pp. 2724–2743, 2017.

[19] M. Nickel, V. Tresp, H.-P. Kriegel, *et al.*, "A three-way model for collective learning on multi-relational data.," in *Icml*, vol. 11, pp. 3104482–3104584, 2011.

[20] B. Yang, W.-t. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," *arXiv preprint arXiv:1412.6575*, 2014.

[21] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "A semantic matching energy function for learning with multi-relational data," *Machine Learning*, vol. 94, no. 2, pp. 233–259, 2014.

[22] Z. Wang, J. Li, Z. Liu, and J. Tang, "Text-enhanced representation learning for knowledge graph," in *Proceedings of International joint conference on artificial intelligent (IJCAI)*, pp. 4–17, 2016.

[23] Q. Wang, B. Wang, and L. Guo, "Knowledge base completion using embeddings and rules," in *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[24] S. Guo, Q. Wang, L. Wang, B. Wang, and L. Guo, "Jointly embedding knowledge graphs and logical rules," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 192–202, 2016.

[25] Y. Zheng, F. Zhao, and Z. Wang, "Fault diagnosis system of bridge crane equipment based on fault tree and bayesian network," *The International Journal of Advanced Manufacturing Technology*, vol. 105, no. 9, pp. 3605–3618, 2019.

[26] X. Feng, J.-c. Jiang, and W.-f. Wang, "Gas pipeline failure evaluation method based on a noisy-or gate bayesian network," *Journal of Loss Prevention in the Process Industries*, vol. 66, p. 104175, 2020.

[27] W. Qiao, Y. Liu, X. Ma, and Y. Liu, "Human factors analysis for maritime accidents based on a dynamic fuzzy bayesian network," *Risk analysis*, vol. 40, no. 5, pp. 957–980, 2020.

[28] M. Kordestani, A. Zanj, M. E. Orchard, and M. Saif, "A modular fault diagnosis and prognosis method for hydro-control valve system based on redundancy in multisensor data information," *IEEE Transactions on Reliability*, vol. 68, no. 1, pp. 330–341, 2018.

[29] H. Wang, Q. Huang, and Z. S. Li, "A dynamic bayesian network control strategy for modeling grid-connected inverter stability," *IEEE Transactions on Reliability*, vol. 71, no. 1, pp. 75–86, 2021.

[30] F. Sala, C. De Sa, A. Gu, and C. Ré, "Representation tradeoffs for hyperbolic embeddings," in *International conference on machine learning*, pp. 4460–4469, PMLR, 2018.

[31] Z. Sun, Z.-H. Deng, J.-Y. Nie, and J. Tang, "Rotate: Knowledge graph embedding by relational rotation in complex space," *arXiv preprint arXiv:1902.10197*, 2019.
[32] J. Wang, K. Yue, L. Duan, Z. Qi, and S. Qiao, "An efficient approach for multiple probabilistic inferences with deepwalk based bayesian network embedding," *Knowledge-Based Systems*, vol. 239, p. 107996, 2022.
[33] "Neo4j." https://neo4j.com/.
[34] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, 2017.
[35] A. Lydia and S. Francis, "Adagrad—an optimizer for stochastic gradient descent," *Int. J. Inf. Comput. Sci*, vol. 6, no. 5, pp. 566–568, 2019.
[36] I. Balazevic, C. Allen, and T. Hospedales, "Multi-relational poincaré graph embeddings," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
[37] T. Lacroix, N. Usunier, and G. Obozinski, "Canonical tensor decomposition for knowledge base completion," in *International Conference on Machine Learning*, pp. 2863–2872, PMLR, 2018.
[38] I. Chami, A. Wolf, D.-C. Juan, F. Sala, S. Ravi, and C. Ré, "Low-dimensional hyperbolic knowledge graph embeddings," *arXiv preprint arXiv:2005.00545*, 2020.
[39] K. Toutanova and D. Chen, "Observed versus latent features for knowledge base and text inference," in *Proceedings of the 3rd workshop on continuous vector space models and their compositionality*, pp. 57–66, 2015.
[40] C. Shang, Y. Tang, J. Huang, J. Bi, X. He, and B. Zhou, "End-to-end structure-aware convolutional networks for knowledge base completion," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 3060–3067, 2019.
[41] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020.
[42] S. Hariri, M. C. Kind, and R. J. Brunner, "Extended isolation forest," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1479–1489, 2019.
[43] D. P. Kingma, M. Welling, *et al.*, "An introduction to variational autoencoders," *Foundations and Trends® in Machine Learning*, vol. 12, no. 4, pp. 307–392, 2019.
[44] T. Zhang, R. Ramakrishnan, and M. Livny, "Birch: an efficient data clustering method for very large databases," *ACM sigmod record*, vol. 25, no. 2, pp. 103–114, 1996.
[45] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, pp. 395–416, 2007.

VIII. Biography Section

Liqiao Xia (Member, IEEE) received the Ph.D. degree in the Department of Industrial and Systems Engineering at the Hong Kong Polytechnic University, M.S. degree from Hong Kong University of Science and Technology in 2020, and the B.S. degree from South China Agricultural University in 2017. His research interests include Knowledge graph; Predictive maintenance; Reliability analysis.

Pai Zheng (Senior Member, IEEE) is currently an Assistant Professor in the Department of Industrial and Systems Engineering, at the Hong Kong Polytechnic University. Dr. Zheng received the Dual Bachelor's Degrees in Material (major) and Computer (minor) Science and Engineering from Huazhong University of Science and Technology, China, in 2010, Master's Degree in Mechanical Engineering from Beihang University, China in 2013, and the Ph.D. degree in Mechanical Engineering from the University of Auckland, New Zealand in 2017. His research interests include smart product-service systems, human–robot collaboration, and intelligent manufacturing systems.

Dr. Zheng is a member of ASME/IEEE/CMES. He serves as the Associate Editor for *Journal of Intelligent Manufacturing* and *Journal of Cleaner Production*, an Editorial Board Member of the *Journal of Manufacturing Systems* and *Advanced Engineering Informatics*, and the guest editor/reviewer for several high impact international journals in the manufacturing.

Manuel Herrera (Member, IEEE) is a Senior Research Associate (SRA) in distributed intelligent systems at the University of Cambridge (UK). His research focuses on predictive analytics and statistical learning in complex systems for the optimal operation and management of smart and resilient critical infrastructure, with a particular emphasis on urban water systems. He also has considerable experience in telecommunication, transport, and maritime sectors through his involvement in projects funded by Industry, the UK government and research councils. Dr Herrera is a fellow of the Royal Statistical Society (RSS) and committee member of the RSS Special Interest Group in statistical engineering. He is also the recipient of the Frank Hansford-Miller fellowship 2021 in applied statistics, awarded by the WA Branch of the Statistical Society of Australia; and the recipient of the Excellence in Research award 2022, granted by the Institute for Manufacturing, Dept. of Engineering, at the University of Cambridge.

Yongshi Liang received her B.Sc. degree and M.Eng. degree from South China Agricultural University in 2017 and 2019. She is currently working as a research associate in the Department of Industrial and System Engineering at the Hong Kong Polytechnic University. Her research interests include Knowledge graph; Natural Language Processing; Augmented Reality Head-Up Display.

Xinyu Li (Member, IEEE) received the Ph.D. degree in industrial engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2009.

He is currently a Professor with the Department of Industrial and Manufacturing Systems Engineering, State Key Laboratory of Intelligent Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, HUST. He has authored or coauthored more than 90 refereed papers. His research interests include intelligent algorithm, Big Data, machine learning, etc.

Liang Gao (Senior Member, IEEE) received the B.Sc. degree from Xidian University, Xi'an, China, in 1996, and the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2002, both in mechatronic engineering.

He is a Professor with the Department of Industrial and Manufacturing System Engineering, Deputy Director of the State Key Laboratory of Intelligent Manufacturing Equipment and Technology, and the Vice Dean of Research and Development Office, HUST. He has authored or coauthored more than 200 refereed papers indexed by SCIE, and has authored 7 monographs. His research interests include operations research and optimization, Big Data, and machine learning.

Dr. Gao is currently the Co-Editor-in-Chief for IET Collaborative Intelligent Manufacturing, and an Associate Editor of Swarm and Evolutionary Computation and the Journal of Industrial and Production Engineering.