ABSTRACT

Title of Dissertation: EXPLOITING CAUSAL REASONING TO IMPROVE THE QUANTITATIVE RISK ASSESSMENT OF ENGINEERING SYSTEMS THROUGH INTERVENTIONS AND COUNTERFACTUALS

Andres Alfredo Ruiz-Tagle
Doctor of Philosophy, 2023

Dissertation Directed by: Associate Professor Katrina Groth
Department of Mechanical Engineering

The main strength of quantitative risk assessment (QRA) is to enable risk management by providing causal insights into the risk of an engineering system or process. Bayesian Networks (BNs) have become popular in QRA because they offer an improved causal structure that represents analysts' knowledge of a system and enable reasoning under uncertainty. Currently, the use of BNs for risk-informed decisions is based solely on associative reasoning, answering questions of the form "If we **observe** $X = x$, how likely is it to **observe** $Y = y$?" However, risk management in the industry relies on understanding how a system could change in response to external influences (e.g., interventions and decisions) and identifying the causes and mechanisms that could explain the outcome of past events (e.g., accident investigations and lessons learned). This dissertation shows that associative reasoning alone is insufficient to provide these insights, and it provides a framework for obtaining more complex causal insight using BNs with intervention and counterfactual reasoning.

Intervention and counterfactual reasoning must be implemented along with BNs to provide more complex insights about the risk of a system. Intervention reasoning answers queries of the form "How does **doing** $X = x$ **change** the likelihood of **observing** $Y = y$?" and can be used to inform the causal effect of interventions and decisions on the risk and reliability of a system. Counterfactual reasoning answers queries of the form "**Had** X **been** $X = x'$ in an event, instead of the observed $X = x$, **could** Y **have been** $Y = y'$, instead of the observed $Y = y$?" and can be used to learn from past events and improve safety management activities. BNs present a unique opportunity as a risk modeling approach that incorporates the complex causal dependencies present in a system's variables and allows reasoning

under uncertainty. Therefore, exploiting the causal reasoning capabilities of BNs in QRAs can be highly beneficial to improve modern risk analysis.

The goal of this work is to define how to exploit the causal reasoning capabilities of BNs to support intervention and counterfactual reasoning in the QRA of complex systems and processes. To achieve this goal, this research first establishes the mathematical background and methods required to model interventions and counterfactuals within a BN approach. Then, we demonstrate the proposed methods with two case studies concerning the risk of third-party excavation damage to natural gas pipelines in the U.S. The first case study showed that the *intervention reasoning* methods developed in this work produce unbiased causal insights into the effectiveness of implementing new excavation practices. The second case study showed how the *counterfactual reasoning* methods developed in this work can expand on the lessons learned from an accident investigation on the Sun Prairie 2018 gas explosion by providing new insights into the effectiveness of current damage prevention practices. Finally, associative, intervention, and counterfactual reasoning methods with BNs were integrated into a single model and used to assess the risk of a highly complex challenge for the future of clean energy: excavation damages to natural gas pipelines transporting hydrogen. The impact of this research is a first-of-its-kind approach and a novel set of QRA methods that provide expanded causal insights for understanding failures and accidents in complex engineering systems and processes.

EXPLOITING CAUSAL REASONING TO IMPROVE THE QUANTITATIVE RISK ASSESSMENT OF ENGINEERING SYSTEMS THROUGH INTERVENTIONS AND COUNTERFACTUALS

by

Andres Alfredo Ruiz-Tagle

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Associate Professor Katrina Groth, Chair/Advisor
Professor Enrique Lopez Droguett
Assistant Professor Michelle Bensi
Professor Jeffrey Herrmann
Professor Michel Cukier
Professor Gregory Baecher, Dean's Representative

Dedication

Para mi abuelo Julio Palazuelos y mi hermano Gonzalo Ruiz-Tagle. Estén donde estén, mi inspiran cada día.

Acknowledgments

After four years of amazing experiences and hard work, this chapter has come to an end and it is time to show my gratitude to all of you who have been beside me throughout this incredible journey.

First and foremost, I would like to thank my advisor, Dr. Katrina Groth. This has been an incredible experience. I was lucky enough to be mentored by someone that shares my values and my passion for doing meaningful work. You are not only an amazing engineer but also an amazing person. I will always be grateful for your mentorship, insights, support, and openness. It has been a pleasure to work with you for these four years and make contributions toward a safer world.

I would also like to thank my committee members, Dr. Enrique Lopez, Dr. Michelle Bensi, Dr. Jeffrey Herrmann, Dr. Michel Cukier, and Dr. Gregory Baecher. I am very grateful to have been mentored by all of you, whether it was in a resilience seminar, class, or the qualifying exam. Your insights in the proposal were fundamental to strengthening the final version of this work. Additionally, I want to thank Dr. Enrique Lopez, the person that made all of this possible. Your mentorship and support date back to the beginning of my academic journey in Chile, where you pushed me to pursue graduate studies in the U.S. I will always be grateful. Also, I would like to acknowledge Ernest Lever at GTI Energy, whose mentorship has been

invaluable. My time at Chicago with GTI Energy encouraged me to pursue a career in safety and it was an experience I will always remember.

To my colleagues in the SyRRA lab, thank you. Your support and friendship is something I will always value. I will always be grateful for coming out of this program with not only great co-workers but also great friends. In particular, I would like to thank Vinnie, Austin, and Ramin. Sharing this experience together has been amazing and I am so happy to see how all of you are accomplishing your dreams. You deserve it.

Manu, you have been my rock throughout these years. You inspire me to pursue my passions, and your love and support bring a smile to my face every day. After almost 1,200 pictures of flowers, I look forward to giving them to you in person when we move to Albany. I could not have done this without you.

Being more than 8,000 km (5,000 miles) apart from my family is not easy. However, I never felt you apart. Your love and support are something I have always cherished and for that I am extremely grateful. You are my inspiration and the reason for where I am right now. Words do not do justice to expressing how happy I am to have you all in my life. Thank you Mom, Dad, Gonza, Teté, Tita, Tata Julio, Tata Alfredo, and Martita, for everything.

To my amazing friends, the old and the new, thank you very much. To Valdés, Schaad, and Brian, "los de siempre." Your more than 20 years of friendship are invaluable. To my university friends, Vale, Pipe, Seba, Fernando, Mati, Chelo, and Sibona, thank you for your support and friendship throughout these years. To the amazing friends I made in the U.S., "los latinos," and my amazing roommates, thank

you so much. To everyone else I have failed to call out specifically, thank you for being part of this journey.

Lastly, I would like to acknowledge the support of the National Science Foundation Grant No. 2045519, the U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration's Pipeline Safety Research and Development Program Award No. 693JK31910002POTA, and the A. James Clark School of Engineering at the University of Maryland. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Pipeline and Hazardous Materials Safety Administration, the Department of Transportation, or the U.S. Government. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Additionally, I would like to acknowledge the support of the Marvin Roush Fellowship and the Chilean National Agency for Research and Development Graduate Studies Scholarship. Lastly, I would like to acknowledge the support of the National Science Foundation Grant No. 2045519, the U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration's Pipeline Safety Research and Development Program Award No. 693JK31910002POTA, and the A. James Clark School of Engineering at the University of Maryland. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Pipeline and Hazardous Materials Safety Administration, the Department of Transportation, or the U.S. Government. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Additionally, I would like to acknowledge the support of the Marvin Roush Fellowship in Risk and Reliability at the University of Maryland and the Chilean National Agency for Research and Development Graduate Studies Scholarship.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BaNTERA	Bayesian Network for Third-party Excavation Risk Assessment
BN	Bayesian Network
CCE	Comparative Causal Effect
CDE	Controlled Direct Effect
CDF	Cumulative Distribution Function
CGA	Common Ground Alliance
CPF	Conditional Probability Function
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
ET	Event Tree
FT	Fault Tree
HyRAM	Hydrogen Risk Assessment Model
ID	Influence Diagram
IE	Intervention Effect
iid	Independent and Identically Distributed
IM	Importance Measure
OSHA	Occupational Safety and Health Administration
PHMSA	Pipeline and Hazardous Materials Safety Administration
QRA	Quantitative Risk Assessment
RT	Research Task
SHyTERP	Safe Hydrogen Transportation through Existing and Repurposed Pipelines
SRA	Society of Risk Analysis
TPD	Third-Party Excavation Damage

Chapter 1: Introduction

1.1 Research Motivation

Complex engineering systems (CES), such as power plants and oil & gas infrastructure, are critical to society's economic stability and security. Their failure to operate as intended can cause major consequences. As such, quantitative risk assessment (QRA) has become key in informing the risk management of these systems and processes because QRA provides causal insights into how failures and accidents occur and how to mitigate them [2]. The impact of QRA on engineering has been broad and significant for many industries, and its implementation can be seen in the generation of causal insights toward solving complex challenges, such as nuclear safety and regulation [3] or providing lessons learned in catastrophes such as the Macondo blowout (Deepwater Horizon) in the Gulf of Mexico [4]. Furthermore, risk assessment and QRA are considered crucial in the development of new industrial applications such as hydrogen technologies to tackle climate change (e.g., international safety codes and standards for re-fuelling stations [5], and the U.K.'s plan to enable hydrogen-based appliances in residential homes [6])

In the past decade, Bayesian networks (BNs) have become one of the most popular methods for modern QRAs because they can quantify uncertainty, provide an improved causal structure that represents an analyst's knowledge of a system or process, and enable more extensive causal insights about risk [7, 8]. These features can be seen in a wide set of applications ranging from analyzing accidents in complex socio-technical systems to dynamically assessing the likelihood of incident scenarios in nuclear power plants [9, 10].

Obtaining causal insights to inform decision-making is a core goal of QRA [11–15]. The causal modeling capabilities of BNs have improved the accuracy and depth of the insights provided by traditional QRA approaches such as fault trees and event trees [16]. Still, the methods for obtaining causal insights in modern BN-based QRAs have been limited to reasoning about associations between the variables describing a system's risk. Associations can only provide insights on **observations** obtained (or hypothesized) from a system or process (e.g., quantifying correlations [17, 18]). However, risk managers are commonly more interested in deeper causal insights involving the effect of decisions and interventions on the risk of a system or from evaluating and learning about the mechanisms that could explain the outcome of past events [11,19]. Interventions and counterfactuals deal with the more complex challenge of providing insights on how a system or process can be **changed**, which is crucial for risk management. Unfortunately, no methods have been established to enable intervention and counterfactual reasoning in QRAs based on BNs, thus limiting the depth of causal insights that modern QRAs can provide on a system's risk.

The most common type of decision support in risk assessments is based on associative reasoning, which provides causal insights into a system or process by answering queries of the form *"If we **observe** $X = x$, how likely is it to **observe** $Y = y$?"* An example of associative reasoning can be seen in the yearly recommendations report published by the Common Ground Alliance (CGA), an organization that promotes damage prevention across underground utilities in the U.S. In the past years, CGA has shown in their report that there is a positive association between excavation damages and an excavator's lack of awareness of damage prevention practices, such as calling 811 before digging [20]. This insight was provided as a recommendation that stresses the need to increase the awareness of damage prevention practices among excavators. However, the CGA has also found that, even though damage prevention awareness is at an all-time high, excavation damage trends have not changed in the past years [21, 22].

The use of BNs in QRA has expanded the depth of insights obtained from associative reasoning and can provide, for example, diagnostic inference capabilities in complex engineering processes. Our work on BNs for the risk assessment of excavation damage [23] shows the benefits of improved associative reasoning capabilities through BNs. Going back to the example on the CGA's yearly recommendation report, our work with BNs expanded the insights obtained by the CGA on the positive relationship between excavation damage and a lack of awareness of damage prevention practices by diagnosing that most excavators involved in damages are professional excavators that decide not to execute preventive measures even though they are aware of them.

However, causal insights based on associations can only provide information on the statistical relationship between the variables describing a system or process and have been known to be prone to bias and to provide oversimplified recommendations [8, 11]. Therefore, traditional QRA has used several methods to expand insights beyond associations and provide decision-makers with more valuable information based on complex causal reasoning. These improved insights are based on more complex causal reasoning based on interventions and counterfactuals and have been employed on popular QRA methods such as importance measures [2] and Accimap analysis [24].

First, intervention reasoning answers queries of the form *"Does **doing** $X = x$ **change** the likelihood of **observing** $Y = y$?"* and can be used to inform the causal effect of interventions and decisions on the risk and reliability of a system or process. An example of intervention reasoning in traditional QRA can be seen in importance measures analysis with fault trees [2]. Importance measures are a crucial quantitative method in QRA to prioritize actions/policies by quantifying how much a system's reliability changes if a component's reliability is intervened to be perfect or imperfect [12, 25]. However, insights from importance measures can be misleading as fault trees are known to oversimplify causality and provide an inaccurate representation of a complex system or process [26, 27].

Second, counterfactual reasoning answer queries of the form *"**Had** X **been** $X = x'$ in an event, instead of the observed $X = x$, **could** Y **have been** $Y = y'$, instead of the observed $Y = y$?"* and can be used to learn from past events and improve risk management activities. An example of counterfactual reasoning can be found in Accimap analysis based on "but-for" accident investigations [24, 28]. Accimaps have been widely used in the industry to determine the contribution of a variable on an accident's outcome through counterfactual statements such as "but-for this variable, the accident would not have occurred" [28, 29]. Accimaps can provide an improved causal structure of a system. However, they can only generate qualitative insights that do not account for uncertainties in an event [30].

These more complex causal insights beyond associations have been crucial in the development and adoption of QRA in the industry [12]. However, as shown above, these insights could only be produced due to the simplified representation of causality adopted in traditional QRA models. In fact, Zio and Aven [31] stated that "how to understand causality in a risk context" is a foundational issue prevalent in modern QRA, showing that reasoning about causality in complex engineering systems and processes is not straightforward. BNs present a unique opportunity as a risk modeling approach that incorporates the complex causal dependencies present on a system's variables as well as their uncertainty. Therefore, exploiting the causal reasoning capabilities of BNs in QRAs can be highly beneficial to improve modern risk analysis.

1.2 Problem Statement and Research Objectives

The main strength of QRA is its ability to provide causal insights into the risk of an engineering system or process. Modern QRAs have used a BN approach due to their improved causal modeling capabilities. However, an appropriate understanding of causality, in addition to the sole use of associative reasoning capabilities of BNs in QRA, is limiting how analysts reason about risk and provide valuable causal insights based on interventions and counterfactuals that have been demonstrated to be highly relevant to different industries and applications.

Therefore, the goal of this dissertation is to exploit the causal reasoning capabilities of BNs to support intervention and counterfactual reasoning in the QRA of complex systems and processes. This research has three main objectives designed to achieve the dissertation's goal:

- **Objective 1 (O1): Establish** the mathematical background and methods needed to perform intervention reasoning with BNs in risk assessment.
- **Objective 2 (O2): Establish** the mathematical background and methods needed to perform counterfactual reasoning with BNs in risk assessment.
- **Objective 3 (O3): Demonstrate** the proposed intervention and counterfactual reasoning methods on a comprehensive QRA of a complex engineering system or process.

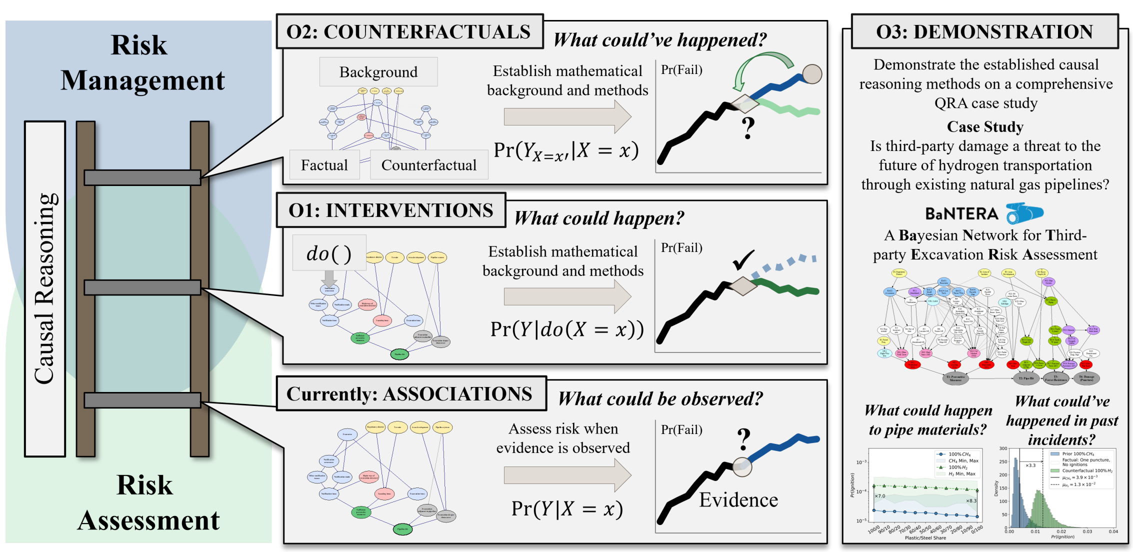


Figure 1.1: Research objectives for exploiting the causal reasoning capabilities of modern BN-base QRAs through intervention and counterfactual reasoning.

Causal reasoning can be represented by a *Ladder of Causality* as proposed by Judea Pearl [32]. Moving upward in the three rungs of this ladder, we find a more complex level of causal reasoning which requires information on the previous rung to be executed. The three levels (rungs) of causal reasoning present in the ladder of causality are (1) associations, (2) interventions, and (3) counterfactuals. In this work, the ladder of causality presented in Figure 1.1 is used as a visual representation of the three main objectives addressed in this work. As seen in the figure, modern QRAs based on BNs are based solely on associative reasoning (rung 1 of the ladder). The results of this dissertation, however, will move QRAs upward in the ladder to enable insights based on interventions (rung 2) and counterfactuals (rung 3).

1.3 Research Approach

The objectives described above represent a guideline for moving modern BN-based QRA methods from the first rung of the ladder of causality (i.e., association reasoning) toward the upper rungs of intervention and counterfactual reasoning. Three tasks were carried out to meet these objectives:

- **Research task 1 (RT1)**: *Establish the mathematical background and methods for intervention reasoning in risk assessment.* Currently, causal insights from modern QRAs based on BNs are based on associative reasoning. The objective of this task is to move up to the second rung of the ladder of causality. Therefore, RT1 starts by identifying the conditions that a BN model should comply with to be able to provide causal insights based on interventions and to illustrate its use in QRA. Although there are several potential applications to intervention reasoning in QRA (e.g., causal discovery and importance measure analysis), we will narrow this research activity toward contributing to the problem of estimating the causal effect of decisions on the variables describing an engineering system or process. Then, a causal BN-based methodology is proposed to support risk-informed decision-making, and its application is demonstrated in an illustrative case study on the problem of third-party excavation damages (TPD) to natural gas pipelines in the U.S.

- **Research task 2 (RT2)**: *Establish the mathematical background and methods for counterfactual reasoning in risk assessment.* As will be explained in Chapter 5, the accuracy of the results of counterfactual methods for BNs in QRA is highly sensitive to the exhaustiveness of the causal model of the system or process to which the methods are applied to. Therefore, in order to move up in the ladder of causality from interventions to counterfactuals, RT2 is divided into the following sub-tasks to enable counterfactual reasoning in BNs for QRA:

  - **Research task 2.1 (RT2.1)**: *Develop a comprehensive BN model on the risk of TPD to natural gas pipelines.* The aim of this research task is to develop a comprehensive BN model for the problem of TPD to natural gas pipelines in the U.S. This task is needed to enable and demonstrate the counterfactual reasoning methods that will be established to achieve O2 as counterfactual insights are highly sensitive to the comprehensiveness of a causal model. To develop a causal BN model for TPD, we first define a taxonomy and high-level architecture that includes the processes, actors, and objects involved in a third-party excavation process. Then, we define a causal structure for the BN model and processed the data from multiple sources to parameterize it. After parameterizing the BN, we will verify and validate the model using nationwide statistics on third-party damages to natural gas pipelines in the U.S.

  - **Research task 2.2 (RT2.2)**: *Establish and demonstrate counterfactual methods.* The comprehensive BN-based risk model defined in RT2.1 will enable us to move up in the ladder of causality toward counterfactual reasoning. Therefore, RT2.2 starts by identifying the conditions that a BN model, in general, should comply with to be able to provide causal insights based on counterfactuals and to illustrate its use in risk analysis. We narrowed this research task toward contributing to the problem of learning lessons from past accidents. A causal BN-based methodology is proposed to provide causal insights from accident investigations based on counterfactual reasoning, which will be demonstrated in a case study. In particular, the 2018 natural gas explosion caused by third-party excavators in the city of Sun Prairie, WI [33] is used as a case study to demonstrate the proposed counterfactual reasoning methods in this task.

- **Research task 3 (RT3)**: *Demonstrate the established causal reasoning meth-*

ods on a comprehensive QRA on the risk of TPD to natural gas pipelines transporting hydrogen: The final research task of this work is to demonstrate how BNs for QRA can provide causal insights on all three rungs of the ladder of causality. To do this, we perform a comprehensive BN-based QRA using the causal reasoning methods established in RTs 1 and 2. The specific engineering application is to assess the risk of TPD to natural gas pipelines transporting hydrogen in the U.S. Hydrogen is an essential fuel for a decarbonized future. However, the risk and safety implications of transporting this gas through existing pipelines are a major challenge to its widespread use. In order to fulfill O3 and contribute to the safe transportation of hydrogen through natural gas pipelines, we first perform an exhaustive literature review and data collection on the topic. We select the HyRAM QRA methodology, proposed by Groth et al., [34] to perform this research task and to guide the development of a causal BN model to assess the risk of TPD to natural gas and hydrogen pipelines. The proposed risk model is based on the BN that results from RT2.1 and incorporates data and assumptions from multiple sources to simulate a third-party excavation process in the U.S. and its potential consequences in a pipeline damage scenario for both natural gas and hydrogen settings. Four independent studies based on the three rungs of the ladder of causality are designed to inform utility companies about the risk of TPD to their future hydrogen pipelines. The results of these studies are based on the methods established in RTs 1 and 2 and will provide causal insights to inform recommendations for the safe transportation of hydrogen through natural gas pipelines.

1.4 Dissertation Overview

Chapter 2 provides a literature review on the necessary background information required to address the three main objectives of this dissertation. This review starts with a summary of QRA, its objectives, and its benefits toward informing the risk management of complex engineering systems and processes. Then, we show the role of causality in the modeling process and insights provided by QRAs, emphasizing how causal reasoning in QRA is based on association, interventions, and counterfactuals. BNs are then presented as an improved causal model for the QRA of complex systems and processes, including its basic mathematical background and probabilistic inference methods. The research conducted in this dissertation is illustrated and demonstrated through case studies on the problem of TPD in natural gas and hydrogen pipeline systems. Therefore, this literature review ends with a presentation of the problem of TPD in the U.S. and its position as a future potential threat to the safe transportation of hydrogen through existing natural gas pipelines.

Chapter 3 shows the results of RT1. This chapter provides a framework that enables the modeling of decisions in a BN-based QRA model that accounts for potential confounding variables between a decision's outcome of interest and the variable to be intervened. Furthermore, we develop and propose a set of metrics describing the magnitude of the causal effect of a decision on a system or process. An illustrative case study on the risk of TPD is used to show how using associative reasoning methods, instead of the proposed intervention reasoning methods, can potentially misguide decision-making with biased causal insights on the effectiveness of implementing new excavation best practices. In addition, the same case study is used to show how an incorrect understanding of the difference between associative and intervention reasoning with BNs can lead to erroneous implementations of widely used methods in the industry, such as risk importance measures.

Chapter 4 shows the results of RT2.1, providing the development process and results of building BaNTERA, a probabilistic Bayesian network model for third-party excavation risk assessment in the U.S. BaNTERA's capabilities for risk-informed decision support are presented in three ways: verification of the model's performance, validation of its damage rate predictions with historical industry data, and application in multiple case study scenarios. Preliminary results indicate that BaNTERA offers valuable risk insight, including and beyond a probability estimation of third-party damage. Using the best available industry data and previous models derived from multiple sources, different probabilistic inference methods are used with BaNTERA to assist in pipeline damage prevention and risk mitigation.

Chapter 5 shows the results of RT2.2. This chapter provides a novel probabilistic approach to counterfactual reasoning called "possible worlds" counterfactuals. This methodology enables the integration of an analyst's causal knowledge about a system (in the form of a BN-based risk assessment model) with the best available evidence about an event of interest (e.g., an accident). As a result, counterfactual hypotheses, commonly used in risk analysis, can now be rigorously assessed through causally-sound probabilistic methods. We demonstrate the capabilities of "possible worlds" counterfactuals with a real-world case study on the 2018 Sun Prairie gas explosion and show how the proposed approach can provide additional lessons and insights beyond those provided by authorities after the event.

Chapter 6 shows the results of RT3, demonstrating the proposed intervention and counterfactual reasoning methods on a comprehensive QRA study on the threat of TPD to future hydrogen transportation through existing natural gas pipelines. A preliminary risk model is developed and used in 4 independent studies based on associative, intervention, and counterfactual reasoning. The first study compares the probability of puncture, ignited release, and thermal harm caused by third-party damage in hydrogen and natural gas contexts. Based on intervention reasoning, the results of the second study show the effect of hydrogen on the risk of TPD to different pipeline materials used in gas distribution systems. Based on counterfactual reasoning, the third study evaluates the probability of alternative consequences to past incidents experienced by a utility partner's transmission systems had they been transporting hydrogen instead of natural gas. Based on associative reasoning, the results of study 4 identify the most relevant risk-influencing factors affecting TPD to distribution and transmission pipelines transporting hydrogen. The result and insights from these four studies are used to inform recommendations for future TPD risk mitigation activities in a hydrogen context.

Chapter 7 concludes this dissertation by summarizing the key results and contributions of the previous chapters to the field of risk analysis. We list the work products associated with this dissertation research, including publications, presentations, codes, and software. Lastly, future directions are provided to move forward into analyzing novel methods toward exploiting the causal insights that improved causal models based on BNs can provide in QRAs.

Chapter 2: Background and Literature Review

The following chapter presents the fundamental background information required to conduct the research outlined in Chapter 1, including the definition of QRA, the role of causality in QRA, causal BNs, and details about third-party excavation as a major threat to natural gas pipelines and the future of hydrogen delivery in the U.S.

2.1 Quantitative Risk Assessment: Definition, Objectives, and Benefits

According to the Society of Risk Analysis (SRA) [35], risk can be defined in relation to the consequences of an activity or process with respect to something that humans value, normally focusing on negative, undesired consequences. In order to describe and measure risk, Terje Aven [19] defines the following quadruplet to characterize it:

$$\text{Risk} = \{A, C, U, K\} \tag{2.1}$$

Where,

- A: is an event that might occur.
- C: are the event consequences (e.g., safety, societal, environmental, security, financial, among others.).
- U: is a measure of uncertainty (usually based on probability) associated with C.
- K: is the background knowledge (and judgment of its strength) that supports C and U.

All activities of an organization involve risks that should be managed as they may affect the achievement of their objectives (e.g., strategic initiatives, operations, processes, and projects, among others). Risk management involves all activities of an organization to handle risks, such as prevention, mitigation, adaptation, and communication. Risk assessment provides a structured process to explore priorities, build consensus, encourage discourse among interested parties, and build a common basis for safety discussions to enable risk management. To do this, risk assessment attempts to answer the following three questions in order to support decision-making [36]:

1. What can happen? (i.e., What can go wrong?)
2. How likely is it that that will happen?
3. If it does happen, what are the consequences?

Assessing the risk of engineering systems and processes is a complex challenge, as data is usually scarce, not all risk scenarios (hazardous events) are known to analysts, and the accuracy of risk assessments is difficult to prove [13]. QRA has become a key process in the risk management of engineering systems by providing a scientifically-sound, defensible, and transparent basis for risk assessment [12, 15]. Unfortunately, QRAs have been predominately used as a mechanical procedure for determining what is an acceptable (or unacceptable) risk based on predefined criteria. This form of QRA takes the form:

$$\text{Risk} = U \times C \leq R' \tag{2.2}$$

Where Risk is the risk of an event, U is measured as the probability of the event (e.g., frequency of an event per year), C is a measure of the consequence of the event (e.g., number of fatalities in an area), and R' is an unacceptable risk threshold (e.g., an unacceptable number of fatalities per year in an area).

QRA presents, in principle, more objective information and accurate data than qualitative risk analyses. However, a quantitative assessment such as the one in Equation 2.2 still suffers from the inherent challenges of assessing the risk of a system: it is unlikely that a QRA will provide a completely faithful representation of reality [37–40]. As such, several prominent authors [2, 12, 15, 19] have stressed that the strength of QRA is not to provide a single number to measure risk, but to provide insights into finding ways to manage risks and support decision-making. As said by Apostolakis: *"QRA results are never the sole basis for decision-making by* *responsible groups. In other words, safety-related decision-making is risk-informed, not risk-based"* [15]. The benefits of QRA as an approach to generate risk-informed insights are clearly stated in the ISO 31010 [41], including:

- Understanding risk and its potential impact upon objectives;
- Providing information to decision makers;
- Contributing to the understanding of risks to assist in the selection of treatment alternatives;
- Identifying the important contributors to risks in systems;
- Comparing risks in alternative systems, among others.

As can be seen, detailed QRAs are not important *per se*, but rather the importance lies in the risk-informed decisions that make an activity safe [14].

2.2 The Role of Causality in QRA

Risk assessments are, in their majority, based on models that represent a system or process. A set of conceptual constructs are used to represent the risk of the system, and these constructs are typically mathematical and heavily based on assumptions about the system [19]. According to Aven, a model-based risk assessment framework should take into consideration all the elements of the quadruplet $\{A, C, U, K\}$ in Equation 2.1 and proposed the framework of Figure 2.1.

Figure 2.1 defines a risk model $G(X)$ with input variables X. The model $G(\bullet)$ relates X to quantities of interest Y. Uncertainty plays a fundamental role in this

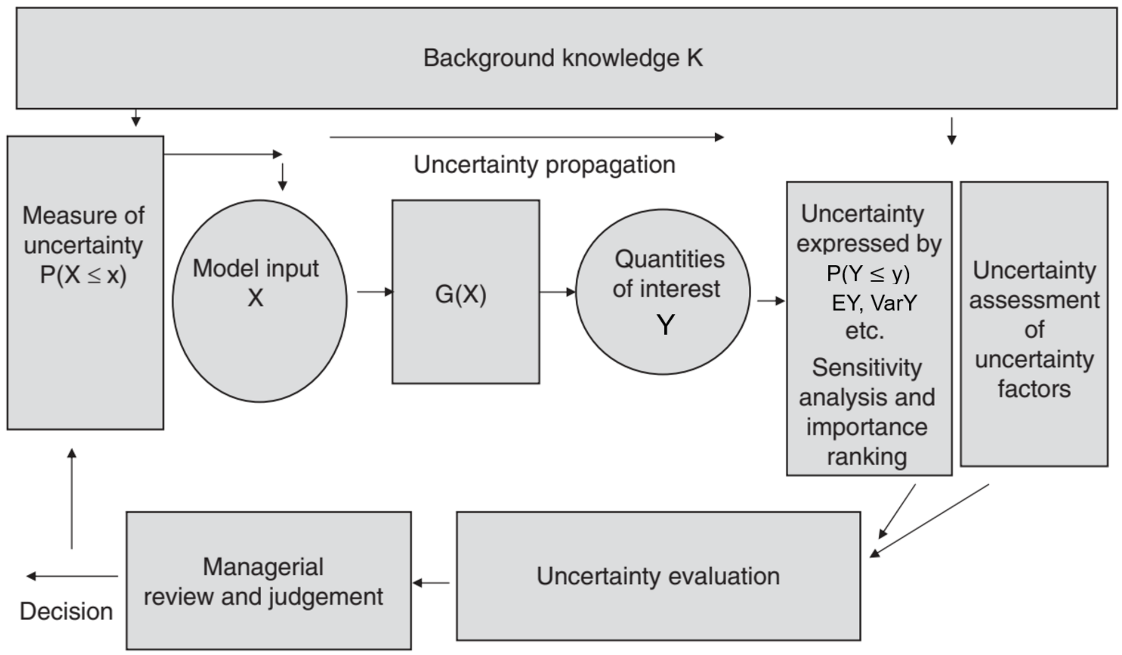


Figure 2.1: A model-based framework to risk assessment. Extracted from [14].

formulation and can be present in the model inputs X as a measure $P(X \leq x)$ (i.e., parameter uncertainty) as well as in the model $G(X)$ (i.e., structural uncertainty). These uncertainties are propagated into Y, for which a risk analyst aims to obtain insights. These insights are commonly based on sensitivity analyses and importance rankings of X, which, as seen in the right blocks of Figure 2.1, can also present uncertainties. The role of the model $G(X)$ is to represent the world, obtain insights about the phenomena being studied, and quantify uncertainties [14]. Therefore, a model and its inherent capabilities play a crucial role in the insights drawn from a system's risk.

Risk insights are linked to explanations about a phenomenon in order to inform risk management decisions about three fundamental questions: (1) How often?/How much? (2) What if? and (3) Why? Examples can be seen in predicting the likelihood of an undesired event in the future, diagnosing the state of a system and its causes, and learning from single events like accidents [8,11,19,31]. Causal reasoning, that is, the process of identifying the relationships between causes and effects, is inherent to human thinking and in the construction of explanations [18,32,42]. Therefore, any insight from a risk assessment will involve causal reasoning in one way or another [16,19].

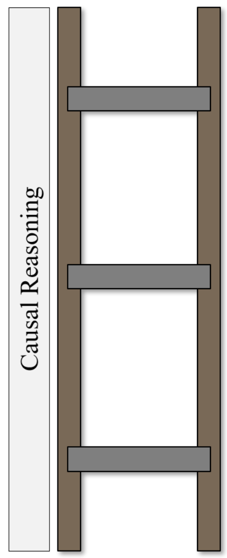
In this work we will adhere to a *mechanistic* definition of causality, which explains that [43]:

Definition 1. ***Mechanistic causation.*** *Changes in causes propagate through a network of law-like causal mechanisms to produce changes in their effects.*

A risk model $G(X)$ can have varying degrees of causal representation in its formulation and, therefore, can be used to reason and generate causal insights with varying degrees of certainty. According to Pearl [18,32], there are distinct hierarchies of causal reasoning which can be visually represented by the *ladder of causality* in Figure 2.2

Any insight about the risk of a system or process can be seen as an inquiry into the causal mechanisms and their properties encoded by the model $G(X)$. As shown in the ladder of causality of Figure 2.2, the processes to answer these inquires to a risk model can be divided into the following three types of causal reasoning [44,45]:

- **Rung 1: Associations.** Correspond to reasoning about observations and making inferences about a certain phenomenon. Associations can be assessed by answering causal queries of the form $Pr(Y = y \,|\, X = x)$ in a model $G(\bullet)$, corresponding to the probability of a variable (or set) conditioned on observa-



Rung	Type of Reasoning	Causal Query (Q)	Mathematical Expression
3	**Counterfactual** Why?	**Had $\boldsymbol{X}$ been $\boldsymbol{X = x'}$** in an event, instead of the observed $X = x$, **could $\boldsymbol{Y}$ have been $\boldsymbol{Y = y'}$,** instead of the observed $Y = y$?	$Q = Pr(Y_{X=x'} = y' \mid X = x, Y = y)$
2	**Intervention** What if?	Does **doing $\boldsymbol{X = x}$ change** the likelihood of **observing $\boldsymbol{Y = y}$**?	$Q = Pr(Y = y \mid do(X = x))$
1	**Association** How often? / How much?	If we **observe $\boldsymbol{X = x}$**, how likely is it to **observe $\boldsymbol{Y = y}$**?	$Q = Pr(Y = y \mid X = x)$

Figure 2.2: The ladder of causality. Adapted from [44].

tions in the data about a system. Therefore, associations give insight into the questions *how often?* or *how much?*

- **Rung 2: Intervention.** Correspond to reasoning about interventions (actions) to a system and making inferences about it. Interventions can be assessed by answering queries of the form $Pr(Y = y \,|\, do(X = x))$ in a model $G(\bullet)$, corresponding to the probability of a variable (or set) conditioned on data coming from a modified model of a system that accounts for the mechanisms about variables X that are replaced to a constant value x (and represented by the *do*-operator). Therefore, interventions give insight into the question *what if?*

- **Rung 3: Counterfactuals.** Correspond to the notion of "modal reasoning," that is, investigating and evaluating claims about what is possible, impossible, sufficient, and necessary. Counterfactuals can be assessed by answering queries of the form $Pr(Y_{X=x'} = y' \mid X = x, Y = y)$ in a model $G(\bullet)$, corresponding to the probability to the probability of a variable (or set) conditioned on data coming from a modified model of a system that represents what is (i.e., $X = x$ and $Y = y$) and what could have been (i.e., $Y_{X=x'} = y'$). Therefore, counterfactuals give insight into the question *why?*

These three types of causal reasoning are the basis of the risk assessment of engineering systems and processes as they support the insights generated by them [11,14,16]. The following sections will show how causal reasoning has been used (at least implicitly) in all aspects of traditional risk assessment and, in particular, QRAs.

2.2.1 Associative Reasoning

Associative reasoning is at the core of traditional statistical and probabilistic analyses performed in risk and reliability assessments. As shown in Figure 2.1, a risk model $G(X)$ maps a set of inputs X to variables of interest Y. For complex engineering systems and processes, these input variables are, in their majority, characteristics of the failure rates and reliability of the components of the studied system. Therefore, an accurate assessment of component reliability is a crucial aspect of risk assessments. Measuring, predicting, and testing reliability is mainly based on probabilistic techniques that fit parametric probability distributions to observational data (e.g., time to failure) in both field and lab settings. Regression and

Maximum Likelihood Estimation are commonly used for parameter estimation [46], which rely on the assumption that observational data is independent and identically distributed (iid). Examples of these methods can be found in Weibull analysis for reliability prediction, proportional hazards models for reliability analysis with explanatory variables, life-stress models for accelerated life testing, and estimating degradation trajectories for evaluating remaining useful life [46, 47].

Several methods described above depend, to different degrees, on empirical models for evaluating the reliability of components. These models have functional forms (e.g., exponential and power-law) that restrict reliability to follow a function that is likely to be observed in data from the field, tests, or that are based on physical concepts. As such, these functional forms allow analysts to obtain insights into a component's risk and reliability. To exemplify the use of empirical models, Guo and Liao [48] studied the degradation of the wall thickness of a chemical container and its response to the pressure applied to it. A power life model, dependent on pressure, was assumed for the shape parameter of a Weibull distribution representing the container's reliability. Fitting this model on observational time-to-failure data from multiple container units at varying pressures allowed the researchers to evaluate the dependency between life and pressure and provide reliability predictions.

However, a component's health and reliability can depend on multiple failure agents through highly complex non-linear and interdependent mechanisms. As such, "black-box" machine learning methods have gained wide popularity in the last decade to make more accurate reliability predictions at the cost of less interpretability and lack of uncertainty assessments [49, 50]. Machine learning methods do not assume a fixed functional form but rely on correlating observational data to an outcome of interest (e.g., a component's life) through composing non-linear functions and learning their parameters through optimization algorithms [51]. The idea behind these methods is to find, in large datasets, a model that entails all relationships and complex structures of the studied system. However, it should be highlighted that interpreting the results of machine learning models is challenging and can misguide decision-making [52]. For example, Moradi et al. [53] showed that interpreting the output of deep learning models (a popular subset of machine learning methods) as probabilities in risk assessment can be incorrect given that these methods are not necessarily calibrated toward measuring uncertainty (i.e., if the model results in a 0.8 probability that a system is in state $Y = y$ does not mean that 80% of the instances in which the models assigned the probability 0.8 would be, in fact, in state $Y = y$).

The methods described above are built toward correlating inputs with outputs and provide information on observational associations between variables in a component or system. The strength of these associative insights is that they can be made without strong assumptions about the causal mechanisms describing the interactions among variables in a system. As such, associative reasoning permits insights such as "untrained excavators are involved in 50% more pipeline damage incidents than trained excavators" or "if the container's thickness decreases by 10% we expect to observe a 50% reduction in its remaining useful life." Insights like these should be interpreted as an assessment of the likelihood of observing an outcome given an observation in the data and could be technically considered a causal effect if the observational data is iid and resembles a randomized experiment [18].

However, it has been widely demonstrated that the iid assumption is rarely the case in risk and reliability data, and, therefore, the models presented above struggle with producing unbiased causal insights. For instance, Hund and Schroeder [54] studied the reliability of a battery voltage by using a traditional linear life model dependent on the age of the battery (see Figure 2.3). This model resulted in low-reliability values for the battery ($\approx$ 52%), which can signify high costs in maintenance and replacement plans. However, the researchers found that a battery load could affect both the age and voltage of the batteries analyzed, violating the iid assumption. After a proper adjustment to the battery's load conditions (which will be explained in the next section), the reliability was found to be almost two times higher ($\approx$ 99%). These associative biases have been widely discussed in the literature as they can significantly affect causal insights [31], reliability predictions [55], and the design of accelerated life tests [56], among others. Furthermore, relying on associative reasoning alone can negatively affect the future of risk and reliability as a field, as dominating machine learning methods are particularly prone to associative biases [44, 49]. In fact, adjusting for these biases is a pillar of explainable machine learning research [57], which will be required for risk and reliability [49].

Even though methods relying on associative insights have been highly convenient and useful in risk assessment, it is important to distinguish what type of insights they can provide to inform decision-making, as correlation does not imply causation.

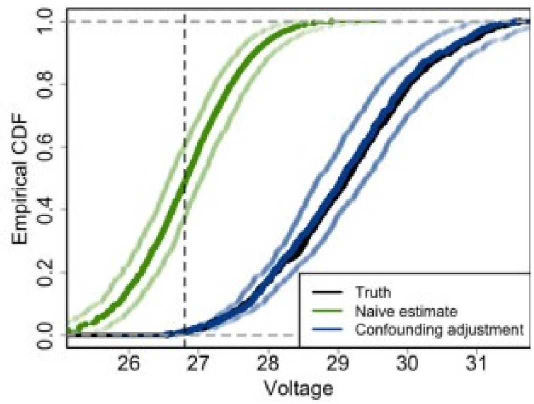


Figure 2.3: Estimated reliability of a battery. Requirements are that reliability (based on a voltage of 26.8V) should be more than 98% after 25 years of operation. Extracted from [54].

2.2.2 Intervention Reasoning

Causation is key in risk assessments. Most QRA models use causal constructs and assumptions due to the nature of the complex systems commonly under study. For instance, ISO 31010 on risk management techniques provides more than 20 methods based on explicitly identifying causes and consequences associated with an event in a system [41]. Two main factors make causality essential in risk assessments: (1) Causal insights are the core of the explanations needed to inform decision-making and risk management [19]; (2) complex engineering systems and processes usually lack enough data to accurately predict their performance [14].

Cases in which appropriate data exists for purely data-driven statistical analyses, such as the methods described in the previous section, could also enable causal insights into a system's risk. For instance, Seo and Pan [56] developed a D-optimal accelerated life test plan considering multiple sources of random effects, including test chambers and item suppliers. Under this setting, a mixed linear model was used to separate the fixed effects of stress conditions from the random effects described before that could cause associative biases on the insights provided by the test results. However, in this study, the control of random effects is specific to the testing conditions. Any changing testing conditions will likely produce associative biases in the test results. This point is highly relevant to complex engineering systems and processes, as they usually work in a dynamic environment under changing conditions [55].

Statistical methods could be able to predict the performance of a system only under static conditions (i.e., controlled and with low uncertainty) but are likely to fail under high uncertainties [16, 46]. Complex engineering systems and processes usually work under high uncertainties. The types of events that require risk management, commonly high-consequence events, are also characterized by a low frequency of occurrence. As such, data to conduct QRAs will be scarce and sparse. System analysis methods are used to deal with these scenarios, as a complex system or process can be decomposed into subsystems and components for which more information is available [14]. Information on these subsystems/components is represented as X in the framework of Figure 2.1. Causal assumptions are used to model the relationships among subsystems/components, and risk-informed insights are obtained by combining the probabilities on the subsystems/components' level with the assumed causal structure in a risk model $G(X)$.

Fault trees (FTs) and event trees (ETs) are risk assessment's most widely used causal models. These models are based on deterministic Boolean gates to relate subsystems/components and map them to a system-level outcome of interest, such as a top event in FTs or hazardous scenario path in ETs [2]. Furthermore, FTs and ETs assume that subsystem/component-level events are independent and will interact through deterministic causal chains toward an end state (composed mostly of binary TRUE/FALSE conditional dependencies). FTs and ETs are commonly used together, as FTs provide information on the likelihood of initiating and branching events in ETs. Figure 2.4 exemplifies these models by combining FTs and ETs in a bow-tie diagram for performing risk assessments under the ARAMIS methodology for process industries regulated by the European Commission [58].

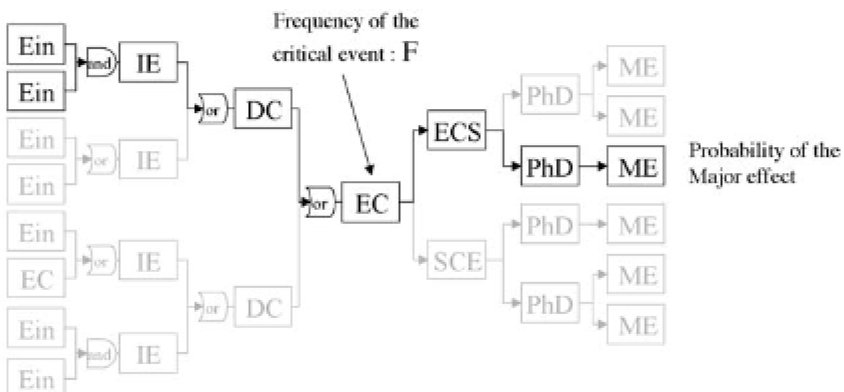


Figure 2.4: Bow-tie model for use in risk assessments under the ARAMIS methodology. Extracted from [58].

Causal models like the one shown above are a representation of the world that is introduced to quantify uncertainties and obtain insights about the risks being studied. As shown in the framework of Figure 2.1, sensitivity analysis and importance ranking on these models are the most important techniques for prioritizing risks, evaluating their contribution, and assessing their management. In fact, they are the basis of risk management guidelines for institutions like NASA [59] and the NRC [3]. These two techniques (and similar ones) are used to understand how the modeled system could change in response to external influences, such as adding redundancy for critical components or allocating additional resources to a component's maintenance. Traditionally, these insights are enabled by FTs and ETs simplistic causal assumptions, as changes in a component/subsystem level will directly provide their causal impact on the outcome of interest. For instance, industry software used for safety-critical systems like NASA's QRAS [60] and NRC's SAPHIRE [61] allow analysts to perform importance ranking and sensitivity analysis by directly modifying an FT/ET model component/subsystem level probabilities (e.g., for computing importance measures [2]) and removing/adding failure modes (e.g., for assessing sensitivity and design decisions). As can be seen, these analyses are based on reasoning about interventions to a system and obtaining insights from its response.

As shown above, intervention reasoning is at the core of risk-informed insights in traditional QRA models and techniques. However, their direct implementation in FT/ET analyses has only been possible due to the simplified (and deterministic) causal relationships encompassed in those modeling frameworks. These simplifications have become debated for assessing the risk of modern complex engineering systems and processes [16]. Modern systems are dynamic, can operate in multiple states/phases, and failures can be functionally highly dependent on each other and on human and organizational factors [9, 62]. Dealing with how systems (and associated models) change in response to external influences (i.e., interventions) is not trivial in this setting, thus limiting causal reasoning in traditional risk-informed decision support.

Several examples can be found in the literature on how reasoning about interventions, without the proper modeling tools, can misguide decision-making and risk management [11, 19, 30]. For instance, in the battery reliability example presented in the previous section (see Figure 2.3), it can be seen that considering a battery's age and voltage as independent produces incorrect reliability estimates. The reason for this misestimation is the presence of battery load as a common cause between a battery's measured age and voltage. Similarly, Yang and Bequette [63] found that field data obtained from a steel manufacturing process can suggest improvement strategies that contradict engineering principles. In their work, data suggest that adding scrap reduces clogging; however, clogs mostly consist of scrap. The reason for this spurious result comes from the fact that scrap is added in steel manufacturing to control the process temperature. As such, additional scrap will be positively correlated with a better-performing process that is also correlated with a lack of clogging, even though it is known that scrap contributes to clogging. For both examples, spurious correlations can be explained by the associative biases described in the previous section.

Although the examples above align better with performance and reliability literature, these problems persist in QRA. Common cause failures (CCFs) in FT analysis, for example, present a major challenge to the QRA of complex systems [64]. CCFs can significantly contribute to a system's risk [2], and their omission in QRAs can cause overoptimistic risk assessments. Several modeling approaches, such as the beta-factor model, are used to maintain the independence of basic events in FTs and allow traditional calculations of importance measures and sensitivity analyses. However, CCF events are often not well defined, and their data collection is scarce.

As such, more advanced causal modeling methods, such as BN models as shown in Section 2.3, are currently being used to overcome the limitations of traditional CCF modeling [9].

Reasoning about interventions, the core of risk-informed insights in QRA, can be a challenging task in modern complex engineering systems. This dissertation tackles this issue in the upcoming chapters.

2.2.3 Counterfactual Reasoning

In safety-critical systems, it is essential to learn from every event; we cannot afford to wait for data from multiple high-consequence events to become available in order to inform safety recommendations [29]. Risk management and system safety researchers and practitioners have explored different ways to tackle this issue, from which a *"learning from events"* strategy is considered fundamental to the field [65–67]. The strategy's main objective is the search for safety improvement opportunities informed by evidence on single events. Investigation reports then make recommendations on areas of improvement. These recommendations are usually based on rectifying the identified causes of the undesired event, and counterfactual reasoning is widely used to identify these causes [28, 68].

Counterfactual reasoning is a type of causal reasoning concerning hypotheses and queries on possible realizations of a past event (i.e., what could have happened, but did not). The use of counterfactual reasoning in risk management has been historically based, at least implicitly, on qualitative "but-for" investigations [29]; that is, finding the necessary causes of an event through statements of the form *"but-for this cause, the event could have been avoided."* Several examples can be found in official accident and incident reports, such as the U.S. NTSB's investigations into the Asiana Flight 214 [69] and the U.S. Air Flight 1016 accidents [70]. Both reports used counterfactuals to imply that the accidents could have been avoided *"but-for the lack of pilot training"* for the Asiana Flight 214, and *"but-for the misjudgment of the crew regarding weather conditions"* for the U.S. Air Flight 1016. In addition, counterfactuals are a central feature of several guidelines and formal analysis techniques for incidents and accidents. For example, Branford and Hopkins recommended using "but-for" counterfactuals to identify a causal structure for Accimap analysis [71]. Likewise, Ladkin's popular Why-Because Analysis method relies on a causal network constructed through "but-for" counterfactual statements [72]. Moreover, counterfactual reasoning has been recommended to safety practitioners as a heuristic for root cause analysis in manuals and guidelines [68, 73].

Some researchers, however, have criticized current counterfactual reasoning methods in risk management and system safety. These criticisms can be summarized into the following limitations: (1) *Linearity*: "but-for" counterfactuals can make investigators prone to turn complex events into linear cause-and-effect chains [74, 75]; (2) *Incompleteness*: event investigators tend to use counterfactuals to transform an event's evidence into a proof of failure to perform according to a system's procedures and norms, therefore missing additional underlying reasons for the occurrence of an event [74, 76]. We add a third limitation (3) *Uncomparable*: there is no rigorous method to compare and prioritize counterfactual hypotheses on event investigations. Consequently, current counterfactual methods limit the bounds of what can be learned from a past event.

Nevertheless, "but-for" counterfactuals are still supported by researchers who pose them as a crucial method for incorporating evidence from single events into their current knowledge of a system to learn from them and to inform safety recommendations [29, 77, 78]. In fact, "but-for" counterfactuals are currently being used to inform causal relationships on the construction of modern systemic methods for accident investigations such as Accimaps and HFACS [71, 79]. However, some authors contend that the limitations of current counterfactual methods still outweigh their benefits [74].

Traditional QRA methods such as FT and ET analysis have been crucial in informing the causality structure and associated uncertainties of past events. In early applications for risk management and system safety, these methods were used to identify malfunctions or failures of specific components of a system that explain an accident or harm (e.g., hardware, procedures, humans, among others) [30]. However, their application is limited to simple systems that are hardware-oriented and in which safety was understood as the "absence of accidents and incidents" [62, 74]. For a modern understanding of safety as "freedom from unacceptable risk," several system-oriented methods have been proposed such as HFAC, Accimap, STAMP, among others. Although these methods can represent more complex structures of causal interactions among the variables of a system, they are qualitative and do not assess uncertainty as in traditional QRA methods. Given that "unacceptable risk" is related to the uncertainty of an event [31], modern qualitative methods fail to answer what they intend to. Therefore, there is a gap in using quantitative probabilistic methods to inform risk management and system safety.

Although scarce, there is relevant literature on system safety and risk assessment that has performed quantitative counterfactual analyses as a part of their work. For instance, Lam and Cruz [80] investigated consumer-level utility gas incidents through a probabilistic network approach to represent cause-effect chains based on historical incident investigations. Here, counterfactual scenarios were modeled by removing specific sets of causes from their incident network to analyze their effect on the likelihood of relevant consequences. A similar approach can be seen in Hughes et al. [81] on a hybrid physics-based and data-driven model for power distribution infrastructure hardening and outage simulation. To analyze the effects of pole hardening on an outage probability in storm events, they set up a counterfactual study by fitting their model to historical storms. Then, they set specific pole hardening levels to estimate if an outage probability could have been decreased. Likewise, Oughton et al. [82] used a similar methodology to set up a counterfactual study on a cyber-physical attack performed in 2015 on the Ukrainian electricity distribution network. Their model represented the Ukrainian network in 2015, and multiple disruption levels were set to estimate their impact on different socioeconomic variables.

We argue that the quantitative studies mentioned above use modeling methods that, although used in a counterfactual setting, are not suited for simulating a counterfactual outcome of an event. Rather, these studies should be interpreted as intervention analyses (as explained in the previous section), in which the calculated outcome corresponds to the expected causal effect of an intervention on a system. As will be explained in Chapter 5, there is a crucial difference between counterfactuals and interventions that should be considered when modeling counterfactuals: all evidence on a past event is included in the model, which can contain information on the variable that we want to intervene in and the outcome of interest. This information is necessary to simulate the background conditions in which an event happened, bend the course of history, and assess the likelihood of an alternative outcome [83]. Interventions, on the other hand, only consider the change itself. The methods used in the works presented in the previous paragraph do not incorporate all evidence of a past event, including highly important information on the observed outcomes of the events they analyzed (e.g., in Hughes et al., some poles actually failed in past storms). As we will show in this dissertation, a quantitative assessment of counterfactuals is possible with modern BN models for risk assessment.

2.3 Bayesian Networks: An Improved Causal Model for QRA

Although useful, traditional QRA methods such as FTs and ETs perform analyses that make unrealistic modeling and causal assumptions. They oversimplify the causal relations between the components and subsystems, they are mainly hardware-oriented, and they use binary logic with deterministic cause-effect constructs that cannot thoroughly account for the causal factors involved in a system's risk scenarios.

BNs address many of the limitations of traditional QRA methods and have gained wide popularity over the last two decades in QRA and risk-informed decision support [9, 16, 84]. BNs can model multi-state systems, incorporate both hardware and organizational factors, provide a causal structure representing analysts' knowledge of a system, and, most importantly, allow causal reasoning under uncertainty.

In its most general form, BNs can be used to represent complex multi-variate joint probability distributions in an efficient way [17, 85]. However, a causal interpretation of BNs is adopted for the risk assessment of complex engineering systems and processes [86]. A BN model M represents variables $V = \{V_1, ..., V_n\}$ and their dependencies as the nodes and edges of a directed acyclic graph (DAG) G. Dependencies among variables are modeled as conditional probability distributions, $Pr(V_i \mid V_j)$, which can be quantified through discrete conditional probability tables (CPTs) or conditional probability functions (CPFs) in the case of discrete state or continuous variables, respectively. However, continuous variables must be discretized to be used in a BN setting. A BN is called "hybrid" if continuous and discrete variables are present. In this work, we will use the following nomenclature and definitions to define a BN:

- $M = \langle V, G \rangle$ corresponds to a BN model defined by a DAG G and variables V.
- $V = \{V_1, ..., V_n\} \in G$ as the set of n nodes present in a BN with DAG G. Also, $\{X, Y, Z\} \in V$ will be used later for clearer explanations.
- The set of m states of a node V_i are represented by $\{v', v'', v''', v^{(4)}, \ldots, v^{(m)}\}$.
- $pa(V_i)$ as the set of *parents* of V_i, i.e., all nodes with an outgoing edge to V_i. A causal Bayesian network model assumes that $pa(V_i)$ causes V_i.

- $de(V_i)$ as the set of *descendants* of V_i, i.e., all nodes with an in-going edge from V_i.

- $Pr(V_i \,|\, pa(V_i))$ as the CPT of a discrete variable V_i or CPF of a continuous variable V_i.

- p as a *path* between V_i and V_j, i.e., a unique set of nodes and edges that connects V_i to V_j.

The independence properties embedded in a BNs structure G are the main feature that enables BNs to reason about causality under uncertainty. These independences in BN are denoted by $Y \perp\!\!\!\perp X \,|\, Z$, (i.e., Y is conditionally independent of X given Z). Three basic junction patterns define conditional independence among nodes in a BN [87]: chains, forks, and colliders. As shown in Figure 2.5, both chains and forks embed $Y \perp\!\!\!\perp X \,|\, Z$ whereas colliders show that $Y \not\perp\!\!\!\perp X \,|\, Z$ (i.e., Y is not conditionally independent of X given Z). If multiple junction patterns are present in a BN structure, the *d-separation* criterion [88] can be used to extract node independence properties from a BN structure:

Definition 2. ***d-separation.*** *X and Y are d-separated conditional on a set of nodes Z (i.e., $Y \perp\!\!\!\perp X \,|\, Z$) if Z blocks every path p between X and Y. Z blocks a path p if either:*

- *p contains a chain $A \rightarrow B \rightarrow C$ or a fork $A \leftarrow B \rightarrow C$ such that B is in Z.*

- *p contains a collider $A \rightarrow B \leftarrow C$ such that B or its descendants are not in Z.*

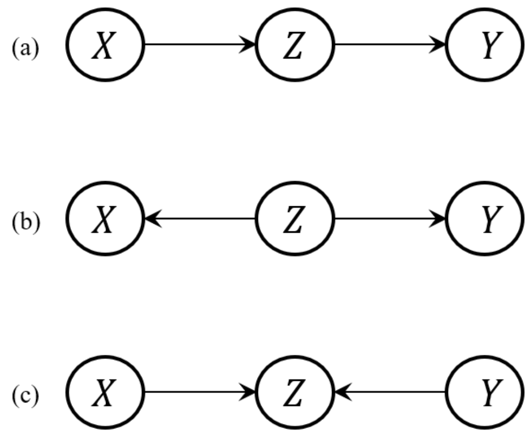


Figure 2.5: Basic junction patterns in a BN. From top to bottom: (a) Chain, (b) Fork, and (c) Collider.

d-separation enables factoring high-dimensional joint probability distributions in BNs by the product of local conditionally independent distributions that are present in the BN's structure. Any joint distribution in a BN can be expressed by the *factorization formula*:

$$Pr(V_1, ..., V_n) = \prod_{i=1}^{n} Pr(V_i \,|\, pa(V_i)) \tag{2.3}$$

Equation 2.3 is widely used for the QRA of engineering systems. An event's evidence, as well as a hypothetical scenario, can be easily incorporated into the BN as a set of specific states of the system's variables, denoted as $\{V_e = v_e\}$. Then, Bayes' theorem can be used to obtain a posterior distribution $Pr(V_1, ..., V_n \,|\, \{V_e = v_e\})$ which incorporates this new evidence [16]. Incorporating and propagating evidence through BNs is the core of associative reasoning, which is presented in the next section.

2.3.1 Reasoning About Causality with BNs

Causal BNs are meant to model a direct representation of the world. It describes the causal connections (nodes and edges in a DAG) and mechanisms (CPTs and CPFs) present in a complex engineering system or process. As such, in principle, BNs encode every possible event's probability. This feature allows BNs to express the probability of any desired scenario in terms of the conditional probabilities specified in the model. This ability is the core of associative reasoning with BNs (see Section 2.2.1). The literature on BNs application to risk and reliability is mainly based on associations. For example, Hegde et al. [89] developed a BN model to help supervisors to determine when an autonomous underwater vehicle mission should be aborted. This model connects 41 technical, organizational, and operational causal factors to a target node representing the probability of aborting a mission. Evidence is gathered on the mission's operational conditions and incorporated into the BN model using the factorization formula shown in Equation 2.3. The BN provides an updated probability of mission abortion conditioned on this evidence and is informed to supervisors to decide on continuing or aborting the mission. Another example can be found in Groth et al. [10], who modeled a reactor's failure with a BN model incorporating information on accident sequences, the health of a reactor's components, and operational parameters of the nuclear power plant. Evidence on operational parameters is gathered dynamically and integrated into the BN to diagnose the potential causes of the reactor's anomalous operations. For more examples of associative reasoning for informing operations, the reader is referred to [90–92].

As shown in the examples above, associative reasoning with BNs is highly beneficial for decision-makers when analyzing a system's operations. Another key research area in BNs for risk and reliability is finding contributors to risk, searching for action/policy opportunities, and assessing the effects of interventions on a system or process [9, 16, 90]. For instance, Zhang et al. [93] worked on identifying the main contributors to pipeline safety in tunneling construction activities. These contributors are prioritized by the results of a sensitivity analysis that considers how much the probability of a risk event changes if the contributor's observed state changes. The prioritized contributors are then informed to decision-makers to implement actions/policies to handle risks. Similarly, Bhandari et al. [94] researched the contributing factors to the risk of deepwater drilling operations using associative reasoning and evidence propagation, similar to Zhang et al. Further examples can be found in [9, 90]. However, as it will be shown in Chapter 3, associative reasoning can produce spurious results for applications concerning changing conditions like the occurrence of an event or the implementation of an action/policy.

Judea Pearl, the creator of BNs, warns about the confusion between associative reasoning (i.e., reasoning about static conditions as shown in Section 2.2) and intervention reasoning (i.e., reasoning about changing conditions) in the application of BNs [18, 95]. Similar warnings have been made for risk assessment by prominent authors such as Aven [19, 30], Zio [31], and Cox [8, 11]. A literature review was conducted by us before the results of this dissertation on the journals Risk Analysis: An International Journal (RA), Reliability Engineering and System Safety

(RESS), and Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability (JRR). Works from the year 2000 to the present were filtered using different Boolean combinations for the keywords “Causality,” “Causal Inference,” “Intervention,” “Engineering,” and “Bayesian Network.” More than 80 papers were identified as having causal reasoning with BNs in engineering systems as their main topic. Still, only two works were found that dealt with intervention reasoning. The first was a review by Langseth and Portinale [84] on the use of BNs in reliability. They dedicated a section to intervention reasoning methods and showed, qualitatively, how planned maintenance decision-making can benefit from interventions with BNs instead of influence diagrams. The second work was done by Hund and Schroeder [54], who worked on the battery performance and reliability example shown in Section 2.2. Using structural causal models (a general formulation of BNs [18]), this work shows how reliability estimations can be strongly biased if interventions are not properly addressed. These two works, however, are focused on reliability engineering and not on risk assessment. As can be seen, there is a clear gap in the development of BN-based intervention reasoning methods for risk and reliability.

Risk analyses requiring more complex causal reasoning based on counterfactuals, such as accident investigations, had not yet been studied with BNs yet. Only one paper on the literature review by Woo et al., [77] mentioned the potential of BNs to obtain insights into the causes of rare events through counterfactual analysis, but no major explanations are provided. Reasoning about the causes and potential outcomes of individual past events is a difficult task, as it requires a comprehensive understanding and manipulation of a system's causal structure, its mechanisms, and the event's evidence [83]. Associative reasoning has been used to analyze industrial accidents with BNs. However, their insights should be interpreted carefully. For instance, Wu et al. [96] used a BN model to analyze a tank farm fire accident that occurred in 1989 in Hungdao, China. Evidence from the event was propagated in the model, and associative reasoning was used to evaluate the posterior probability of potential causes of the accident. This study found that the probability of "inadequate procedures" is expected to increase from 25% to 58% in past incidents, suggesting that these procedures should be revised. This study uses a correct interpretation of associations, as no insights are provided about the magnitude that improved procedures could have had on the probability of the accident. Counterfactual reasoning is required to provide insights about "magnitudes of contribution" in past hypothetical settings [83]. However, several examples can be found in the literature that provides, through associative reasoning, insights about how much different actions/policies could have made a past accident less likely. For example, Wang et al., [97] analyzed the collision accident of the Alam Pintar bulk carrier in Singapore in 2009. They used a BN to model the accident and identified the variables whose probability increased after observing a collision of the vessel. Then, they suggested prioritizing actions/policies by estimating the probability reduction of a vessel accident given a change in the state of a variable (through evidence propagation using the factorization formula). This magnitude change, however, should not be interpreted as the potential "Probability reduction of maritime accident," as the authors suggest. The latter is a counterfactual insight and, as Chapters 3 and

5 show, controlling for associative biases and the correct integration of a BN model and an event's evidence are needed to provide these insights. Unfortunately, similar errors in interpreting associative insights as counterfactual insights can be found in several studies, such as [98].

BNs were created to reason about the three rungs of the ladder of causality [18] (see Figure 2.2). However, there is a lack of research on more complex causal reasoning in the risk and reliability analysis. Looking at the history of BNs in the field, two "eras" can be identified. In the first era, from the late '90s to 2010, BNs were mainly used to expand current risk and reliability methods (e.g., FTs and ETs) in dependability modeling and enable the analysis of multi-state systems [90]. The second era, from 2010-present, saw an increase in the use of associative reasoning methods for scenario simulation, diagnosis and prognosis, and identifying risk contributors [9]. We believe that a third era of BN research in risk and reliability will benefit by exploiting this tool's more complex causal reasoning capabilities: reasoning about interventions and counterfactuals for risk-informed decision support.

2.4 Application: Third-Party Excavation Damage to Existing Natural Gas Pipelines and Future Hydrogen Infrastructure

The causal reasoning methods established in this dissertation will be demonstrated and applied to the problem of third-party excavation damage to natural gas and hydrogen pipelines in the U.S. Third-party excavations are complex, involving physical, human, and organizational factors. Damages can lead to catastrophic consequences, and the potential of hydrogen injection into existing natural gas pipelines can lead to consequences that are not yet well understood. A BN approach is suitable for modeling the complexities of TPD, and the causal reasoning methods proposed in this dissertation can allow analysts to generate crucial insights to improve its risk assessment. The following sections introduce TPD and its risk to both natural gas and hydrogen pipelines. More detailed explanations are provided in Chapters 4 and 6.

2.4.1 Third-party Excavation Damage

A leading cause of pipeline failure in the U.S. is TPD [21,99]. According to the Pipeline and Hazardous Materials Safety Administration (PHMSA) [21], between 2016 and 2022, TPD caused more than 20% of the total number of incidents and accidents on transmission and distribution pipelines. Collectively, these incidents resulted in 11 fatalities, 36 injuries, and a total monetary loss of $144M in property damage. Figure 2.7 illustrates that the proportion of TPD incidents in the U.S. is fairly constant in time [21]. A TPD is any type of damage to a pipeline caused by an excavation activity performed by someone different from the utility company or its contractors. TPDs include damages to the external coating of the pipe, dents, gouges, or punctures directly into the pipeline itself. The following 5-step process, shown in Figure 2.6 is required by PHMSA in order to ensure safe digging [100]:

1. Make the required "One-Call" before digging;
2. Wait the required time for locations of all buried facilities (including pipelines)

to be marked;

3. Confirm that all affected utilities are accounted for;

4. Respect the location markings when digging; and

5. Dig with care, avoiding contact with underground hazards.

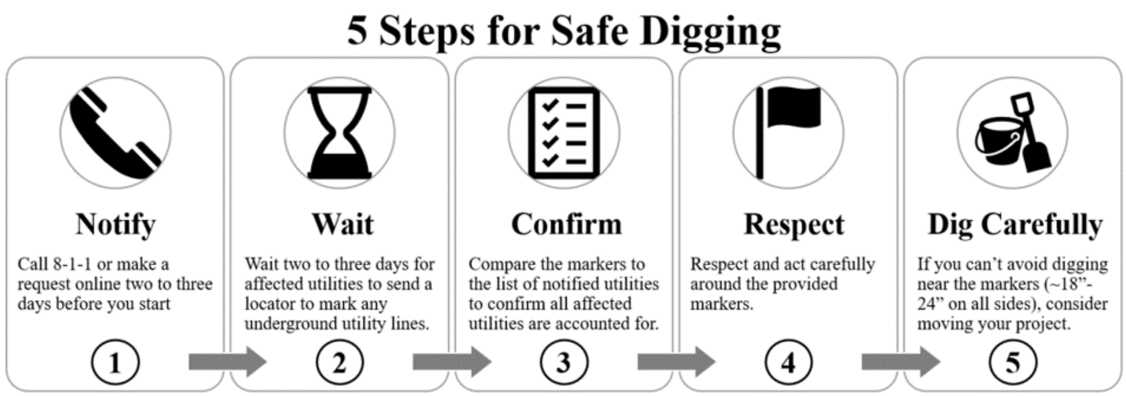


Figure 2.6: Five steps for safe digging according to PHMSA [100]

The process above and its influencing factors will be further explained in Chapter 4.

Due to the significant amount of property damage and safety risks posed by failures in excavations near pipelines, there have been a number of research projects and studies carried out that identify the root causes and contextual factors that may lead to TPD [101–106]. However, there is a noticeable lack of models that are not only comprehensive but also causally based and updated in terms of technology, practices, and regulation for modeling the probabilities of a pipe hit and subsequent damage in third-party excavations.

Across the relevant pipeline damage literature, FTs are the most common technique for modeling different failure modes of a pipeline [107–110]. In the context

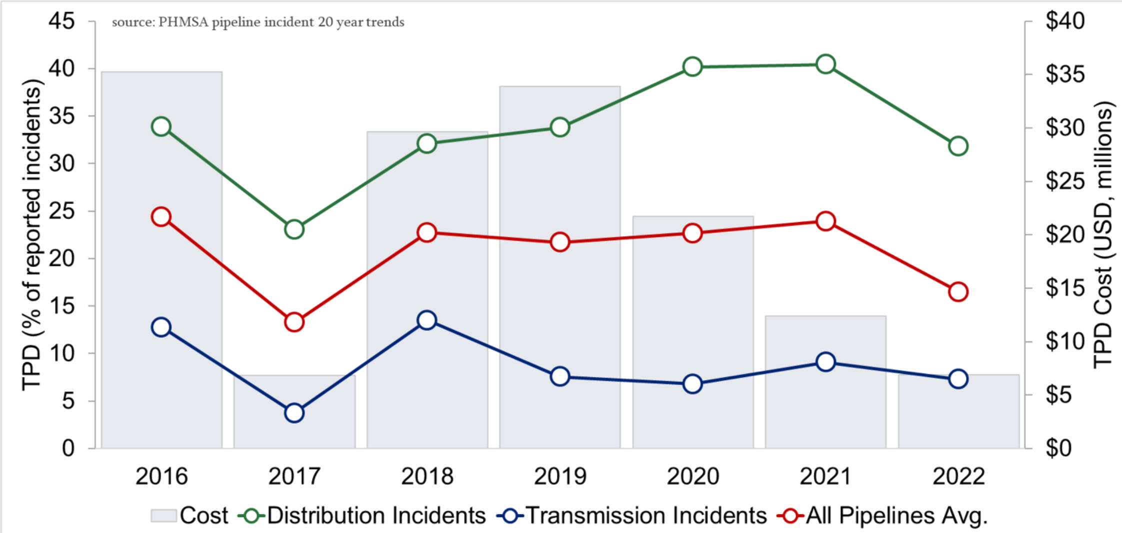


Figure 2.7: TPD constitutes a large portion of excavation incidents involving gas pipelines, leading to significant monetary losses in property damage. Source: PHMSA [21]

of TPD, these FTs model pipeline damage due to the failure of a set of safety barriers in a pipeline system. For example, in an excavation scenario, pipeline damage is more likely to occur if a pipe was marked incorrectly by a utility service representative and the excavation depth exceeds the pipeline's cover depth. Chen and Nassim's [111] FT model for TPD is widely used within the industry to predict the probability of hitting a pipeline in an excavation activity. This model served as the basis for future FT research on TPD that expanded the scope of the model with larger sets of safety barriers and damage causes [103, 112–114]. Although FTs are still a popular industry tool for addressing and evaluating the risks of TPD, they face parametrization and structural limitations. TPD data are often incomplete, which poses a significant problem for parameterizing and updating FTs. Furthermore, FTs rely on Boolean logic to describe a system's failure, which is not an adequate approach to account for the complex dependency relationships present in an excavation process [112].

The logic structure within an FT can be easily mapped to a corresponding BN; as such, FT-based BNs have been used to expand the dependency structure within FT modeling. For example, in their attempt to capture multiple pipeline failure types in a single FT-based network, Wang et al. [115] included a simplified representation of TPD. FT-based BNs have also been used to model common causes between the basic events of an FT. For instance, Xiang and Zhou [112] expanded Chen and Nessim's fault-tree model into a BN in order to include common causes between basic events, such as the public awareness level of one-call centers. Li et al. [116] proposed an FT-based BN for TPD on subsea pipelines to relax the modeling limitations of conventional FTs. Although FT-based BN models address some of the limitations of FT modeling, they still rely on Boolean logic, limiting their ability to realistically represent complex dependency relationships in pipeline risk assessments.

On their own, BNs offer both modeling and probabilistic inference capabilities for risk assessment beyond those of FT-based methods. By providing a more complete representation of the modeled system's causal dependencies and associated uncertainties, decision-makers are capable of reasoning about associations to diagnose pipeline damage root causes, predict damage probabilities, and investigate intercausal dependencies for various damage scenarios. Outside the TPD context, this approach has gained wide popularity for modeling and characterizing pipeline risks, as the context surrounding a damage scenario can vary widely. This trend is prevalent in modeling corrosion damage, one of the leading causes of pipeline damage. Smith et al. [117] used in-line inspection data as inputs to a physics-based BN to predict the corrosion growth rate for gas pipelines. Following this, Smith et al. [118] used BNs for data-driven inspection planning for oil and gas pipelines based on key pipeline parameters such as age, location, and the current state of the pipeline coating. Similarly, Zhang et al. [101] used a series of BNs to relate pipeline parameters to leakage orifice size and failure probability due to corrosion or external interference.

This trend has permeated into recent TPD research to address the complexity inherent in understanding excavation pipeline damage. Jackson and Mosleh [119] developed a BN risk model based on the decision-making process that a third-party faces during excavation activities. They model the excavation process using event-sequence diagrams and integrated both physical (e.g., pipeline diameter, material, and depth of cover) and human-related (e.g., excavator awareness of one-call systems and training) variables. However, this model is still in the initial development phase, has not been validated, and does not have details on either structure or parameterization. Guo et al. [120] model the evolution of pipeline damage risk leading up to and following TPD and pipeline failures. Their study, however, emphasizes the consequences of pipeline damage more than the causal relationships that led to TPD. In their research to model unintentional TPD, Cui et al. [121] developed a "3rd Party Disturbance Index" based on several damage indicators informed by background and fundamental factors. The disturbance index was framed as a BN. However, all damage indicators were considered independent, which does not accurately capture the complexity of an excavation activity. Although the above-mentioned models overcome the limitations of FT-based BNs, they do not capture a sufficiently comprehensive set of up-to-date root cause factors or their dependencies necessary for modeling TPD.

GTI Energy, a leading research organization focused on energy transition solutions, provided a more comprehensive model for TPD based on causal BNs. These models evaluate factors that center around different aspects of pipeline damage, including those that contribute to the puncture resistance of a pipeline [122] and the probability of a locating and marking issue [123]. These models, however, were in an initial phase and have yet to be incorporated into a comprehensive model for the risk assessment of TPD.

The studies mentioned above collectively represent a solid foundation for building a comprehensive TPD model and decision support tool through a BN approach. However, there is still a need to capture TPD risk effectively. This need is addressed in Chapter 4, where a comprehensive BN model is developed for assessing the risk of TPD in the U.S.

2.4.2 Transporting Hydrogen Through Natural Gas Pipelines

Hydrogen is an essential fuel for a decarbonized future [124]. To enable the wide adoption of hydrogen technologies, an extensive pipeline network will be needed to deliver hydrogen to facilities such as industrial sites and distribution centers. Several programs are being implemented in the U.S. to blend hydrogen into natural gas pipelines (up to 20%), which is expected to be a safe and cost-effective path toward hydrogen transportation [125, 126]. Although blending will be needed in the early stages of hydrogen delivery to facilitate economies of scale and generate demand, pipelines must transport pure hydrogen to achieve net-zero objectives by 2050 [124]. However, transporting pure hydrogen through existing natural gas pipelines poses major safety challenges concerning hydrogen embrittlement of steel pipes, higher ignition probabilities from gas releases, and unique fire and explosion behavior [127]. Therefore, the risk associated with hydrogen transportation through pipelines must be assessed.

TPD is one of the major threats to natural gas pipelines in the U.S. and is expected to be a major hazard to the safe transportation of hydrogen through pipes. Consequences from hydrogen releases have been widely studied in preliminary risk assessments of hydrogen pipelines [128–134]. However, no works have studied how hydrogen can affect the frequency and consequences of TPD.

Previous works on the risk assessment of hydrogen transportation through existing natural gas pipelines have been mainly focused on the consequences but not the causes of a hydrogen release. For example, Froeling et al. [128] performed a QRA on high-pressure transmission pipelines, focusing on the individual risk associated with a hazardous hydrogen jet fire from a rupture. Although this work accounted for the higher ignition probability of hydrogen, the pipeline's failure frequency was assumed to be the same as in a natural gas pipeline. Likewise, Witkowsi et al. [129] performed a QRA on both jet fire and explosion hazards in a ruptured transmission pipeline and focused on providing individual risk profiles for pure hydrogen, natural gas, and blend releases. Although hydrogen embrittlement was incorporated into the assessment, this was done by assuming a new independent failure cause in pipelines with a frequency equivalent to corrosion in natural gas transmission.

These and other similar works [130–134] have reduced TPD into a single frequency value for risk assessment purposes and assume that TPD rates will be equivalent to those observed in a natural gas system. However, hydrogen affects several important properties associated with resistance to TPD. Hydrogen embrittlement has been found to degrade and decrease fracture toughness, ductility, and impact resistance of carbon steels commonly used in both high- and low-pressure transmission [135–137] and distribution pipelines [138, 139]. These material property changes lead to greater vulnerability to the dents, gouges, and punctures caused by TPD. Furthermore, Zhang and Adey [140] found that for high-pressure transmission pipelines, the presence of hydrogen significantly increased the probability of delayed failures from dents from excavators. While the integrity of plastic pipelines used for natural gas distribution has been found to be unaffected by pure hydrogen transportation [141], plastic pipes are more susceptible to being punctured in excavations than metallic pipes [122], and the resulting gas releases are more likely to ignite due to static electricity buildup in plastics [142]. Considering also that hydrogen's ignition probability is consistently higher than natural gas for different gas-air mixtures [143], owing to its low ignition energy and wide flammability range, it becomes clear that TPD can pose a major hazard to plastic pipes transporting hydrogen.

Chapter 6 presents a preliminary causal BN model to assess the risk of TPD to hydrogen pipelines. This research considers, for the first time, the effect of hydrogen on both the frequency and consequence of TPD. Furthermore, we leverage the causal reasoning methods established in Chapters 3 and 5 to provide insights and recommendations for future actions and policies required to enable the safe transportation of hydrogen through existing natural gas pipelines.

Chapter 3: Exploiting the Causal Reasoning Capabilities of Bayesian Networks for the QRA of Engineering Systems: Intervention Reasoning

3.1 Introduction

Answering "what could cause an accident or harm?" is a crucial question in risk analysis. As shown in Chapter 2, intervention reasoning has been key for generating causal insights in past QRAs. Insights from popular methods in QRA, such as importance measure analysis with Fault Trees, were heavily based on interventions. However, these insights were prone to mislead decision-makers as past QRA models were known to oversimplify the causality of a system or process [62,76]. This problem was stressed by Aven and Zio [31] as a foundational issue in QRA, and they proposed causal modeling by means of graphical models, such as BNs, as a promising approach to address the need for causality in QRA. Interventions with graphical models are crucial for generating causal insights in decision-support fields such as econometrics [144], epidemiology [145], and toxicology [8], and the benefits have been promoted in the risk analysis field by authors such as Cox [11] and Fenton, and Neil [16]. Still, to the best of the author's knowledge, no works before this chapter's research have clarified the background and methods needed to enable intervention reasoning in the QRA of complex engineering systems and processes.

This chapter shows the results of RT1 and establishes, for the first time, the mathematical background and methods required for enabling intervention reasoning in QRA by means of the BN modeling approach for complex engineering systems and processes. Therefore, this research enables modern QRA to move up one rung in the ladder of causality, from generating insights based on associations and correlation (rung 1) to generating causal insights based on interventions and causal effects (rung 2). To demonstrate the applicability of interventions in QRAs, the scope of this chapter's research was narrowed to using interventions in BNs to estimate the effect of decisions on a system's risk. This chapter starts in Section 3.2 by establishing the formalisms, mathematical background, and techniques to reason about interventions in a BN-based QRA setting. Then, Section 3.3 proposes a BN-based framework that enables risk analysts to properly model actions/policies as interventions in BNs, intending to improve risk-informed decision support. This framework is exemplified through an illustrative case study on TPD for natural gas pipelines in Section 3.4, which also sheds light on the limitations of associative reasoning to generate causal insights to support risk management. Last, Section 3.5 presents concluding remarks on the benefits of performing causal reasoning through interventions in QRA.

This chapter shows the results of RT1, which is an extended version of a peer-reviewed manuscript that was published in *Risk Analysis: An International Journal* for the special issue *"Bayesian Networks for Risk Analysis and Decision*

Support" [146].

3.2 Intervention Reasoning with BNs: Mathematical Background and Methods

In QRA, causal insights from interventions differ from associations in that the latter uses evidence on already known variables and events to estimate the probability of observing an outcome in a particular scenario. In contrast, interventions explicitly calculate the causal effect that intervening (hypothetically) in a specific variable will have on the probability of occurrence of an outcome of interest. Given this fundamental difference, interventions have to be properly framed in a BN model to correctly estimate a causal effect and not an observational association.

3.2.1 Interventions and the $do()$-operator

To move from association to intervention reasoning in the ladder of causation, hypothetical interventions in BNs need to be modeled, for which the *do-operator*, $do(\bullet)$, is introduced. Given a BN model M, the effect of an intervention $do(X = x')$ on a node Y can be represented by [147]:

$$Pr_M(Y \mid do(X = x')) = Pr_{M_{X=x'}}(Y \mid X = x') \tag{3.1}$$

were $M_{x'}$ is the post-intervention BN model. In terms of the original DAG $G \in M$, the equivalent post-intervention DAG is $G' = G_{\bar{X}} \in M_{x'}$, where $G_{\bar{X}}$ corresponds to G with all in-going edges to X removed, as shown, for example, in

Figure 3.1b. In addition, CPTs (or CPFs) of M are modified in $M_{x'}$ to represent $X = x'$.

Combining equations 2.3 and 3.1, the joint distribution generated by an intervention $do(X = x')$ on a set X of nodes that represent the changes made by an action/policy in an engineering system can be expressed by the *truncated factorization formula* as:

$$Pr(V_1, ..., V_k \,|\, do(X = x')) \propto \prod_{i \,|\, V_i \notin X} Pr(V_i \,|\, pa(V_i))\big|_{X=x'} \tag{3.2}$$

where $Pr(V_i \,|\, pa(V_i))$ are pre-intervention conditional distributions present in M. Then, it can be seen that the use of Equation 3.2 to represent an intervention on a BN model M with DAG G is equivalent to applying the factorization formula 2.3 to find the joint distribution of a post-intervention BN DAG $G' = G_{\bar{X}}$, which is then conditioned on $X = x'$, resulting in $M_{X=x'}$.

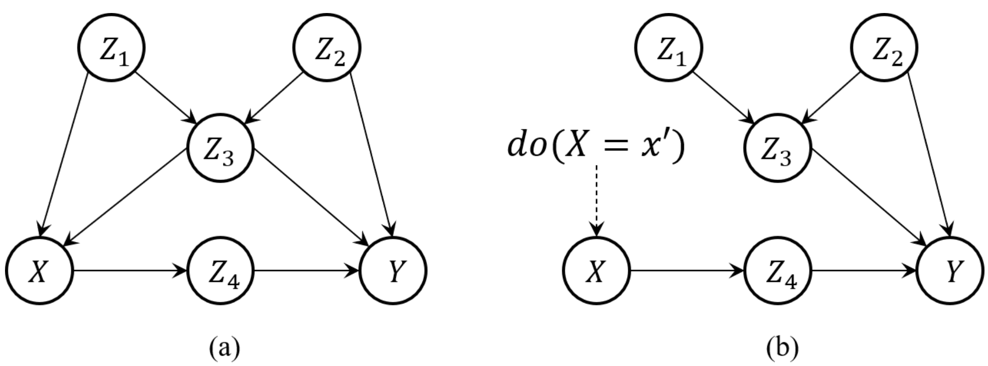


Figure 3.1: Illustrative BN DAG. (a) Pre-intervention DAG, $G \in M$. (b) Post-intervention DAG, $G \in M_{X=x'}$.

As an example, applying Equation 3.2 on Figure 3.1a BN, the post-intervention joint distribution of performing $do(X = x')$ can be computed by:

$$
\begin{aligned}
&Pr(Y, Z_4, Z_3, Z_2, Z_1 \,|\, do(X = x')) \\
&= \frac{Pr(Y \,|\, Z_4, Z_3, Z_2)Pr(Z_4 \,|\, X = x')Pr(Z_3 \,|\, Z_2, Z_1)Pr(Z_2)Pr(Z_1)Pr(X = x')}{Pr(do(X = x'))}
\end{aligned}
\tag{3.3}
$$

Then, the effect of the intervention on Y can be identified from Equation 3.3 just by averaging over Z_1, Z_2, Z_3 and Z_4 as:

$$
\begin{aligned}
&Pr(Y \,|\, do(X = x')) \\
&\propto \sum_{Z_4, Z_3, Z_2, Z_1} Pr(Y \,|\, Z_4, Z_3, Z_2)Pr(Z_4 \,|\, X = x')Pr(Z_3 \,|\, Z_2, Z_1)Pr(Z_2)Pr(Z_1)
\end{aligned}
\tag{3.4}
$$

3.2.1.1 Perfect and Imperfect Interventions

It is important to highlight that interventions evaluated using the $do()$-operator can be perfect or imperfect. A perfect intervention strictly sets a BN node to a particular value or state deterministically (i.e., $Pr(do(X = x)) = 1.0$). In contrast, an imperfect intervention sets the value or state of a node to a probability. As such, an imperfect intervention represents a manipulation of a BN that still represents a degree of uncertainty on the value or state of the intervened node. Imperfect interventions are easily implementable in discrete BN models parameterized through CPTs, as they will only require assigning a new probability to the states of the intervened node. Similarly, imperfect interventions in continuous BNs can be modeled through a new probability distribution assigned over the values the intervened variable can take after an intervention. Mathematically, using Equation 3.2, an imperfect intervention can be computed as:

$$Pr(V_1, ..., V_k \,|\, do(X = x')) \propto \prod_{i \,|\, V_i \notin X} Pr(V_i \,|\, pa(V_i))\big|_{X=x'} \times Pr(X = x') \tag{3.5}$$

For a perfect intervention, the term $Pr(X = x') = 1.0$, which makes Equations 3.2 and 3.5 equivalent to $Pr(V_1, ..., V_k \,|\, do(X = x')) = \prod_{i \,|\, V_i \notin X} Pr(V_i \,|\, pa(V_i))\big|_{X=x'}$. Unless specified differently, perfect interventions are assumed for the rest of this work.

3.2.2 Why Should We Care? The Problem of Confounding Bias

When computing intervention queries on BN models, disentangling correlational from causal relationships between the intervened variable X and the studied outcome Y due to common causes is critical. If not done, the obtained post-intervention distribution can be highly biased. In other words, this means that $Pr(Y \,|\, do(X = x'))$ usually do not coincide with $Pr(Y \,|\, X = x')$ due to the presence of common causes. For an engineering system, this phenomenon was discussed by [54], where they show that the mistreatment of common causes can lead to incorrect results and insights from a BN model. Bias-inducing common cause nodes between X and Y are called *confounders*, and are defined by [148] as:

Definition 3. ***Confounder.*** *A pre-intervention node Z is a confounder of the* *effect of X on Y if conditioning over it blocks a back-door path from X to Y.*

where a *back-door path* between X and Y is any non-causal path between them (i.e., a path with an incoming edge to Y that starts with an incoming edge to X). Evidence propagation through unblocked back-door paths is the reason for possible spurious associations between X and Y when generating causal insights, a bias called *confounding bias* [18].

The simplest graphical representation of confounding bias is shown in Fig. 3.2, where Z confounds the effect of X on Y. Then, to compute the effect of $do(X = x')$ on Y, Equation 3.2 can be used as $Pr(Z, Y \mid do(X = x')) = Pr(Y \mid X = x', Z)Pr(Z)/Pr(do(X = x'))$ and averaged over Z, obtaining:

$$Pr(Y \mid do(X = x')) = \sum_{Z} Pr(Y \mid X = x', Z)Pr(Z) \tag{3.6}$$

and thus blocking the back-door path $X \leftarrow Z \rightarrow Y$ by conditioning on Z.

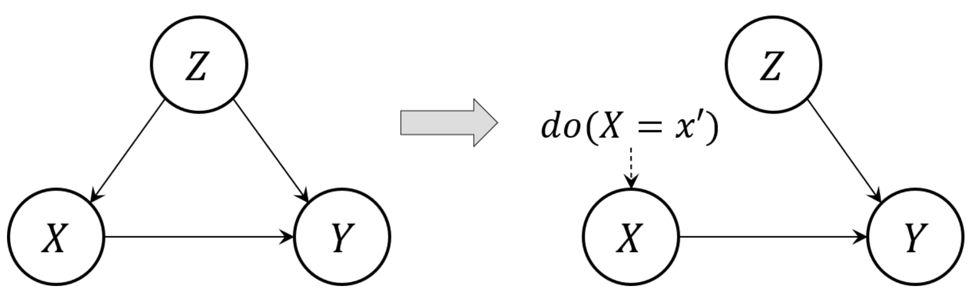


Figure 3.2: Basic confounding BN structure and its post-intervention DAG.

The process of conditioning the causal effect of X on Y over a bias-inducing confounder Z is called *adjusting for confounding*. Cases like Figure 3.2 are easy to adjust for, but real-life BN structures are usually more complex, requiring the adjustment of multiple confounders to compute post-intervention distributions properly. To address this difficulty, the first option is just to average over all non-X and non-Y nodes identified by Equation 3.2 (being not all necessarily confounders). This first option is viable only if the risk analyst has a completely parameterized BN model. A second option is to find a sufficient admissible set of confounders $Z = \{Z_1, ..., Z_l\}$ in the BN model that ensures no unblocked back-door paths exist between X and Y if they are adjusted for confounding, generating an unbiased estimation of the intervention query $Q = Pr(Y \mid do(X = x'))$ by making Y dependant on X only for the direct paths between them (which is analogous to removing the incoming edges to X as done by applying the truncated factorization formula). The set Z can be found using the *Back-door Criterion*, defined by Pearl as [95]:

Definition 4. *Back-door Criterion.* *Given an ordered pair of nodes (X, Y) in a BN model, a set of variables $Z = \{Z_1, ..., Z_l\}$ are sufficient admissible for adjustment relative to (X, Y) if:*

- *No node in Z is a descendant of X and*
- *Z blocks every back-door path between X and Y.*

Then, Equation 3.6, which we will now call as the *Back-door formula*, can be used to estimate the causal effect of intervening X on Y if Z is considered to be a set of nodes that satisfies the back-door criterion.

The back-door criterion is one of the most powerful tools present in BNs, enabling risk analysts to compute confounding-unbiased intervention queries on a model even though it is not fully parameterized. To exemplify the latter, consider an analyst that represented an engineering process by the BN shown in Figure 3.1a and has access to data on X, Z_1, Z_3 and Y. If he is interested in estimating the effect that intervening on $X = x'$ will have on Y, the intervention query $Q = Pr(Y \mid do(X = x'))$ needs to be identified. Given that there is no data on nodes Z_2 nor Z_4, the analyst uses the back-door criterion and identifies $Z = \{Z_1, Z_3\}$ as a sufficient adjustment set for confounding, enabling the use of back-door formula in Equation 3.6 as $Pr(Y \mid do(X = x')) = \sum_{Z=Z_3,Z_1} Pr(Y \mid X = x, Z)Pr(Z)$ to estimate the intervention query Q.

Sometimes, in the presence of non-parameterized nodes, it is not possible to quantify a sufficient adjustment set for confounding, and thus Equations 3.2 and 3.6 can not be used to compute an intervention query $Q = Pr(Y \mid do(X = x'))$. To generalize the identifiability of causal queries in BN models, [95] defined the *do-calculus rules* by which a post-intervention distribution in $M_{x'}$ can be computed from pre-intervention data in M. These are defined as follows:

Theorem 1. ***Do-Calculus Rules.*** *Let X, Y, Z, W be a disjoint set of nodes in $G \in M$. Let $G_{\bar{X}}$ and $G_{\underline{X}}$ be the graph obtained by deleting the incoming and outgoing edges of X, respectively. Let $G_{\bar{X}\underline{Z}}$ be the DAG resulting from deleting incoming edges to X and outgoing edges from Z. The following rules are defined for $\{X, Y, Z, W \subseteq V\} \in G$.*

Rule 1 *(Insertion / deletion of observations):*

$Pr(Y \mid do(X = x'), Z, W) = Pr(Y \mid do(X = x'), W)$

if $(Y \perp\!\!\!\perp Z \,|\, X, W)_{G_{\bar{X}}}$

Rule 2 *(Intervention / observation exchange):*

$Pr(Y \,|\, do(X = x'), do(Z = z'), W) = Pr(Y \,|\, do(X = x'), Z, W)$

if $(Y \perp\!\!\!\perp Z \,|\, X, W)_{G_{\bar{X}\underline{Z}}}$

Rule 3 *(Insertion / deletion of interventions):*

$Pr(Y \,|\, do(X = x'), do(Z = z'), W) = Pr(Y \,|\, do(X = x'), W)$

if $(Y \perp\!\!\!\perp Z \,|\, X, W)_{G_{\bar{X}Z(\bar{W})}}$, *where* $Z(W)$ *is the set of Z-nodes that are not ancestors of any* W *node in* $G_{\bar{X}}$

A sequential application of these rules guarantees the identifiability of an intervention query, if possible, by providing a *do*-free post-intervention distribution expression [149]. However, it should be noticed that in the presence of a non-fully parameterized BN model, the estimation of the intervention query $Q = Pr(Y \,|\, do(X = x'))$ by using the back-door criterion or the do-calculus rules is only possible if data on the variable of interest Y has been collected. If this is not the case, the intervention query can not be estimated.

Applying the *do*-calculus rules to identify an intervention query is not trivial and can become very demanding in large graphs. Given this, [149] automated this process, making causal identification straightforward by publicly available libraries such as the R package `causaleffect` [150].

3.3 A BN-based Framework for Intervention Reasoning in the QRA of Engineering Systems

To create a comprehensive BN model that represents the behavior of an engineering system when intervened, the full spectrum of contexts and causal factors relevant to the system should be considered. This not only enables risk analysts to express their knowledge and assumptions but also to inform decision-makers properly.

To address these requirements, we propose the three-step BN-based framework shown in Figure 3.3. This framework enables risk analysts to use the BN mathematical methods presented in Section 3.2 in a QRA setting to properly estimate and inform the effect that new action/policy can have on an engineering system.

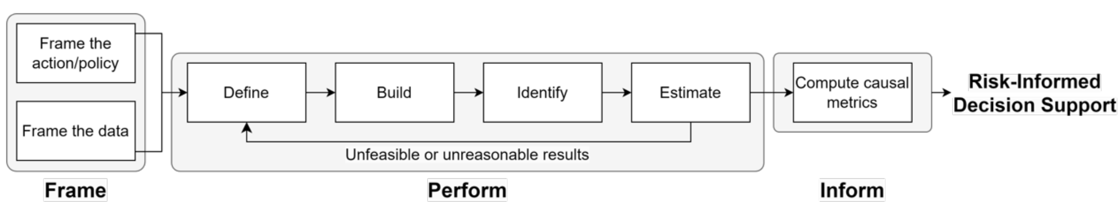


Figure 3.3: Proposed BN-based framework for causal reasoning through interventions in a risk assessment.

Risk management frameworks for risk-informed decision support, such as influence diagrams (IDs), are similar to the proposed framework in this work. Both share a node representing a decision (decision node in IDs and intervened node in the proposed framework) and an outcome of interest (utility node in IDs and outcome of interest in the proposed framework). However, IDs require a pre-defined set of decision nodes, and the model is built around assessing the utility associated with implementing the modeled decisions. On the other hand, the proposed framework differs from IDs in that the BN model used to assess the effect of decisions on risk can be built around representing a system's risk and not necessarily a decision scenario. Interventions are identified and evaluated after a BN model is built or selected. As such, the proposed framework expands the capabilities of current BN models to permit intervention reasoning without transforming the BN into a separate decision model such as an ID.

The steps of the proposed framework are defined below.

3.3.1 Step 1: Frame the Problem

First, the risk analyst frames the problem in terms of an intervention query, taking both the action/policy of study and available data into consideration. In order to do this, the following practices are suggested:

(i) *Frame the action/policy.* The risk analyst understands and contextualizes both the objective of the decision-maker and the different action/policy changes that are being considered to achieve it. To guide this process, the following questions are suggested:

- What action/policy changes are being considered? (intervention)
- What is the objective of the action/policy? (outcome)
- What is the context of the action/policy with respect to the problem?

(ii) *Frame the data.* In addition, a broad understanding of the available data to model the studied query is needed. This will help with the identification of candidate dataset variables that represent the action/policy and its objective. To guide this process, the following questions are suggested:

- What data are available?
- Is the quantity and quality of the data comprehensive enough to support modeling the action/policy?
- Is the considered action/policy change present in the data? if not, it is possible to model?
- Is the action/policy objective present in the data? if not, it is possible to model?

3.3.2 Step 2: Perform Intervention Reasoning

Second, the studied action/policy is modeled and quantified using a BN model and the tools shown in Section 3.2. This is done to estimate the intervention query outlined in the previous step so that all assumptions used by the risk analyst to compute it are explicit and properly communicated to the decision-maker. To ensure this, the following methodology is proposed:

(i) *Define the intervention query.* The intervention query is defined in a general form by:

$$Q = Pr(\text{action/policy objective} \mid do(\text{action/policy change})). \tag{3.7}$$

Equation 3.7 will explicitly express the target and intervention to be modeled in the BN.

(ii) *Build the BN model.* The risk analyst expresses his or her understanding of the system behavior as a BN model, defining its nodes, states, parameterization strategy, and quantification. This is essential to later express Equation 3.7 in terms of the available data. A new causal BN or a previously constructed model can be used for analyzing the intervention query. Previously defined BN models for QRA will be ideal as they commonly present a comprehensive causal structure representing a system. However, a new BN model can also be constructed to evaluate the effect of actions/policies on an outcome of interest.

In the case of building a new BN model, we propose the meta-BN structure shown in Figure 3.4 to guide the BN construction and to easily enable the methods presented in 3.2. Here, dashed bi-directed edges represent causal assumptions between nodes that are not imposed in this meta-model. On the other hand, directed edges should be used, as shown in the figure. Note that all nodes considered as relevant by the risk analyst should be included in the BN structure, even if they are not observed in the available data.

Given an intervention query $Q = Pr(Y \mid do(X = x'))$, the nodes in Fig. 3.4 are defined as:

- *Treatment nodes* (X). Corresponds to the set of nodes that the risk analyst wants to intervene on (applying the *do*-operator) to estimate its effect on an outcome node Y. These nodes should reflect the action/policy changes that the decision-maker wants to implement in a system.

- *Outcome nodes* (Y). Corresponds to the set of target nodes in Q that the risk analyst is interested in estimating. These nodes should represent the variables of interest that a decision-maker is taking into consideration to make a decision (i.e., indicate if the action/policy objective is achieved).

- *Confounder nodes* (Z_C). Corresponds to the set of nodes that the risk analyst considers that can confound and thus bias the estimation of Q. These nodes should be identified and informed to the decision-maker to clarify how the risk analyst interpreted the system behavior against an intervention made on it.

- *Mediation nodes* (Z_M). Corresponds to the set of nodes that the risk analyst considers that can mediate (i.e., increase or decrease) the direct effect that an intervention on X has on Y. Structurally, mediation nodes are represented in the causal path between the treatment and outcome nodes. It should be noted that mediation nodes can also be confounders if they comply with Definition 3.

- *Structural nodes* (Z_S). Corresponds to the set of nodes that are not identified by any of the definitions above but that contributes to the representation of the risk analyst knowledge of the system.

(iii) *Identify the intervention query.* The defined Q from Step 2.i is now mapped to the BN defined in Step 2.ii, where the action/policy and objective are represented by X and Y, respectively. Then, an expression for $Q = Pr(Y \mid do(X =$

x')) can be obtained using either Equation 3.2, the back-door criterion or by a systematic application of the $do()$-calculus rules.

(iv) *Estimate the intervention query.* Using the expression found in the previous step, $Q = Pr(Y \mid do(X = x'))$ is estimated using the available data. The BN structure should be revised if the implications of the obtained result are unfeasible or unreasonable.

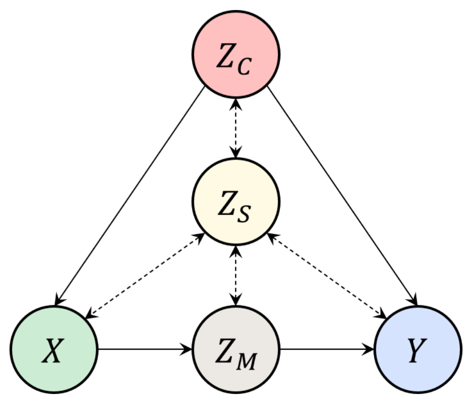


Figure 3.4: Proposed meta-BN structure for intervention reasoning.

3.3.3 Step 3: Inform Decision-Makers

Finally, the results obtained in the previous step of the framework have to be expressed in the best way possible to inform the decision-maker. The estimated value of the intervention query Q does not give much information on its own, so the following intervention-based causal metrics are suggested to provide relevant information on the causal effect that an intervened variable X has on an outcome Y:

(i) *Intervention Effect (IE).* This metric is used to express the difference in the outcome likelihood between the pre- and post-intervention distributions, given the risk analyst's knowledge of the system.

Given the pre-intervention marginal distribution of an outcome variable $Pr(Y)$, the effect of the intervention $do(X = x')$ on a specific value of $Y = y'$ can be expressed as:

$$IE = Pr(Y = y' \,|\, do(X = x')) - Pr(Y = y') \tag{3.8}$$

(ii) *Comparative Causal Effect (CCE).* This metric is used to express how an outcome changes relatively to different actions/policies, enabling the risk analyst to rigorously compare all action/policy alternatives considered by the decision-maker.

The comparative effect of two different intervention queries
$Q = Pr(Y \,|\, do(X = x'))$ and $Q = Pr(Y \,|\, do(X' = x''))$ on an specific value of an outcome variable $Y = y'$ can be expressed as:

$$CCE = Pr(Y = y' \,|\, do(X = x')) - Pr(Y = y' \,|\, do(X' = x'')) \tag{3.9}$$

(iii) *Controlled Direct Effect (CDE).* This metric provides decision-makers with insights into how much of the system's post-intervention behavior was directly affected by the performed intervention and not by its consequences. Similar values of IE and CDE suggest that the effect of the interventions is mostly direct. If not, the effect is considerably mediated.

Given the presence of a mediation node Z_M, the direct effect of the intervention $do(X = x')$ on Y can be computed by controlling for different relevant values of $Z_M = z'$ as:

$$\begin{aligned} CDE = Pr(Y = y' \mid do(X = x'), do(Z_M = z')) \\ -Pr(Y = y' \mid do(Z_M = z')) \end{aligned} \tag{3.10}$$

which requires the computation of the queries $Q = Pr(Y \mid do(X = x'),$ $do(Z_M = z'))$ and $Q = Pr(Y \mid do(Z_M = z'))$.

Both CCE and CDE metrics require the identification and estimation of multiple intervention queries on the built BN model. These can be identified and estimated by following Steps 2.iii and 2.iv from Section 3.3.2.

3.4 Illustrative Case Study: Third-Party Damage in Natural Gas Pipelines

In this section, the proposed BN-based framework for causal reasoning through interventions is demonstrated in a case study regarding TPD to natural gas pipelines (i.e., damage done by personnel not associated with the pipeline operation). In addition, the limitations of using associative reasoning to answer intervention queries are discussed. It is important to state that this case study is meant to be illustrative of the proposed framework and not conclusive in its results.

The natural gas pipeline network in the U.S. consists of more than 4 million kilometers of pipelines used to gather, transmit, and distribute natural gas to customers across the country [151]. PHMSA has collected data on natural gas pipeline incidents and accidents since 1970, which indicates that TPD is a leading cause of failures [21]. In an effort to mitigate TPD, the CGA, a non-profit organization dedicated to preventing damage to underground infrastructure, provides the following dig-in best practices steps in the excavation process [152]:

1. The excavator delineates the digging area.

2. The excavator provides notice of intent to dig to a One Call center (e.g., 811) or submits an online form providing dig and site information.

3. The utility operator locates and marks (L&M) their underground facilities in the area of the request in an accurate and timely manner. When done, One Call will notify the excavator.

4. The excavator proceeds to dig safely by exposing underground utilities using proper tools to avoid accidental damage.

TPD can occur when one of these practices fails to be properly executed.

According to a report by the CGA [153], “Notification Issues,” “Locating Issues,” and “Excavation Issues” were the leading root causes groups for excavation damage in 2018, corresponding to improper execution of steps 2, 3, and 4 respectively. Between 2015 and 2019, “Excavation Issues” represented the 44% of damage causes, with “failure to test hole (pothole)” (i.e., not exposing underground facilities using hand tools before digging) as its leading contributor [153]. Given this, one of the major recommendations provided since 2019 by the CGA report was to promote potholing as a practice for excavators.

In this illustrative case study, we will answer the question: "how does making potholing mandatory will affect the sufficiency of TPD preventive measures?" based on pipeline damage databases of a partner utility company. In order to do this, the proposed framework presented in Section 3.3 is applied as follows.

3.4.1 Step 1: Frame the Problem

First, the problem of modeling and estimating the effect that making potholing mandatory can have on TPD prevention is framed in terms of an intervention query. Both the action and available data are framed as follows.

(i) *Frame the action/policy*

- *What action/policy changes are being considered?* Evaluate making potholing a mandatory practice in excavation activities.
- *What is the objective of the action/policy?* To ensure that sufficient preventive measures are taken to avoid TPD. The idea behind this objective is that preventive measures insufficiency drops by a 90% if two out of three main risk factors for TPD (i.e., "Notification Issues," "Locating Issues," and "Excavation Issues") mitigating best practices were sufficient [122].
- *What is the context of the action/policy with respect to the problem?* Reduce TPD to natural gas pipelines. The CGA [154] recognizes that excavators have limited knowledge about best practices beyond the need to

notify before beginning work. The 2018 CGA DIRT report [153] suggests promoting potholing, given that DIRT data shows that the lack of it is highly correlated with excavation-related damages.

(ii) *Frame the data*

- *What data are available?* Natural gas pipelines third-party caused damage reports from 2015 to January 2020 from a partner utility company. To protect the proprietary information of our partner, we have created a simulated dataset using the same variables. In addition, we have access to an industry-validated fully parameterized TPD BN model, which is described in [122].

- *Is the quantity and quality of the data comprehensive enough to support modeling the action/policy?* From the data, we see that, in terms of quantity, there are approximately 9,000 damage reports available with more than 40 different context and root cause-related fields (such as "root cause" and "root cause group," "excavation tools," "dig-in description," among others). Regarding quality, only 0.4% of the data is missing, on average, in three relevant context fields. In addition, high granularity and low redundancy make the data comprehensive enough to develop criteria for quantifying possible relevant variables that are not explicitly reported.

- *Is the considered action/policy change present in the data? if not, it is possible to model?* Potholing is not reported explicitly, but it is possible to quantify it from the type of tools used in the excavation and with written dig-in descriptions. Moreover, we can compare our quantification with U.S. potholing statistics in [122] to validate our quantification.

- *Is the action/policy objective present in the data? if not, it is possible to model?* There is no field in the data that directly quantify if preventive measures were taken. However, we can use the same parameterization strategy performed in [122] to quantify this variable in our model. The details of this strategy are shown in Section 3.4.2.

3.4.2 Step 2: Perform Intervention Reasoning

The following steps are followed to model and estimate the effect of making potholing mandatory for TPD prevention.

(i) *Define the intervention query.* For this problem, the intervention query can be expressed as:

$$Q = Pr(\text{Sufficient preventive measures} \,|\, do(\text{Make potholing mandatory})) \tag{3.11}$$

(ii) *Build the BN model.* The QRA of excavation damage to natural gas pipeline considers the following root cause groups: "Notification Issues", "Locating Issues," and "Excavation Issues." The latter can be mitigated by performing the excavation best practices recommended by the CGA in [155], from which potholing is the focus of this case study. Performing sufficient excavation, L&M, and notification practices are considered the main factors preventing

TPD. Also, the specific type of third party that performed the excavation is considered to cause differences in notification practices and the likelihood of following excavation best practices. These assumptions are expressed in the discrete BN model shown in Figure 3.5. The nodes definitions (based on the meta-BN structure node types presented in Section 3.3.2) and parameterization strategy are shown below. Each node of the model was parameterized with the data from our utility partner, and CPTs can be found in Appendix A. Node states and pre-intervention probabilities are shown in Table 3.2.

- *Third-Party Excavator* (Z_1). Identifies the type of excavator that performed the excavation. These can be "Property Owner", "Contractor," or "Government Entity." This node is explicitly reported in the used damage reports dataset, so its CPT is parameterized directly from it.
- *Potholing Performed?* (Z_2). Identifies if potholing was performed in an excavation activity. Although not explicit in the dataset, we created a new field identifying potholing based on the type of tools used in the excavation (such as a shovel, air vacuum excavator, and hydro vacuum excavator) in addition to cross-validation with written dig-in descriptions for each damage report dataset entry. Given this parameterization, we found that potholing was performed in 22.07% of excavations. We validated this result with the partner company TPD BN model in [122], which shows that potholing is performed on 22% of excavations.
- *Other Excavation Best Practices Performed?* (Z_3). Identifies if any CGA

excavation best practices aside from potholing are performed. These are "maintain clearance after verifying marks," "protect/shore/support marked and exposed facilities," "proper backfilling practices," and "maintain marks or request re-marking if faded" [155]. This node is explicitly reported in the used damage reports dataset, so its CPT is parameterized directly from it.

- *Sufficient Excavation Practices?* (Z_4). Identifies if there were insufficient excavation practices in general, becoming a contributor to TPD in the damage report. In addition to CGA excavation best practices, this node includes all other excavation issues that are not directly dependent on the specific excavator, such as failing to dig with care. This node is explicitly reported in the used damage reports dataset, so its CPT is parameterized directly from it.

- *Sufficient L&M Practices?* (Z_5). Identifies if there were insufficient L&M practices by part of the utility operator, becoming a contributor for TPD. Inaccurate and lack of marks are considered bad L&M practices. This node is explicitly reported in the used damage reports dataset, so its CPT is parameterized directly from it.

- *Sufficient Notification Practices?* (Z_6). Identifies if there were insufficient notification practices by part of the excavator, becoming a contributor for TPD. Excavation activities without notification or made different from the information notified are considered bad notification practices. This node is explicitly reported in the used damage reports dataset, so its CPT is parameterized directly from it.

- *Sufficient Preventive Measures?* (Z_7). Indicates if sufficient preventive measures were taken to avoid TPD, which depends on the sufficiency of three factors: excavation, L&M, and notification practices. This node is not explicitly reported in the dataset. To parameterize its CPT, we used the same parameterization strategy proposed by [122], which is shown in Table 3.1. The three unknown parameters α, β and γ in the CPT represent a drop of 90% in preventive measures insufficiency if two out of three factors were sufficient. These parameters can be found using damage reports mapping of Z_4 and Z_5 to $Z_6 = No$.

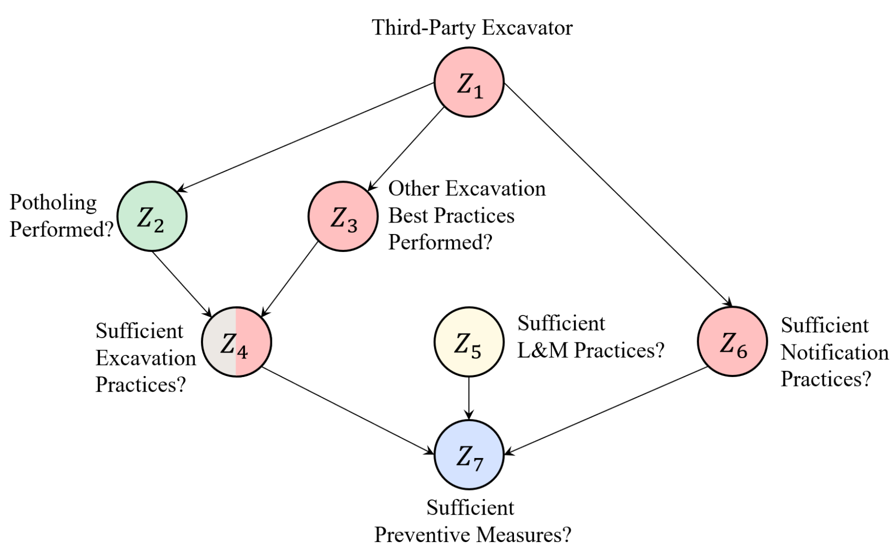


Figure 3.5: Built BN model for TPD with focus on potholing practices. Colors denote node types as done in Figure 3.4.

Given Section 3.4.1 identified action/policy changes and objectives, Z_2 : "Potholing Performed?" and Z_7 : "Sufficient Preventive Measures?" are considered as

Z_4	Yes				No			
Z_5	Yes		No		Yes		No	
Z_6	Yes	No	Yes	No	Yes	No	Yes	No
Yes	1.0	1-α	1-0.1β	1-β	1-0.1γ	1-γ	0.9	0.0
No	0.0	α	0.1β	β	0.1γ	γ	0.1	1.0

Table 3.1: CPT formulation for node Z_7: "Sufficient Preventive Measures?"

Node	Node State	Pre-Intervention Probability [%]	Post-Intervention Probability [%]	Node Type
Z_1: Third-Party Excavator	Property Owner	19.41	19.41	Z_C
	Contractor	74.81	74.81	
	Government Entity	5.78	5.78	
Z_2: Potholing Performed?	Yes	22.07	100.00	X
	No	77.93	0.00	
Z_3: Other Excavation Best Practices Performed?	Yes	71.60	71.60	Z_C
	No	28.40	28.40	
Z_4: Sufficient Excavation Practices?	Yes	54.57	94.30	Z_C, Z_M
	No	45.43	5.70	
Z_5: Sufficient L&M Practices?	Yes	87.66	87.66	Z_S
	No	12.34	12.34	
Z_6: Sufficient Notification Practices?	Yes	54.54	53.02	Z_C
	No	45.46	46.98	
Z_7: Sufficient Preventive Measures?	Yes	77.17	78.35	Y
	No	22.83	21.65	

Table 3.2: TPD BN model nodes, states, pre- and post-intervention probabilities and node types.

treatment and outcome nodes respectively in the proposed meta-BN structure shown in Fig. 3.4 (i.e., $X = \{Z_2\}$ and $Y = \{Z_7\}$). Then, given the context of the action/policy, $Z_C = \{Z_1, Z_3, Z_4, Z_6\}$ are identified as confounding nodes and $Z_S = \{Z_5\}$ as a structural node. In addition, $Z_M = \{Z_4\}$ is also identified as a mediation node of the effect of Z_2 on Z_7.

(iii) *Identify the intervention query.* Mapping the intervention query defined in (i) to the BN built in (ii), we can express Q as:

$$Q = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes})) \tag{3.12}$$

with subsequent post-intervention BN DAG shown Figure 3.6.

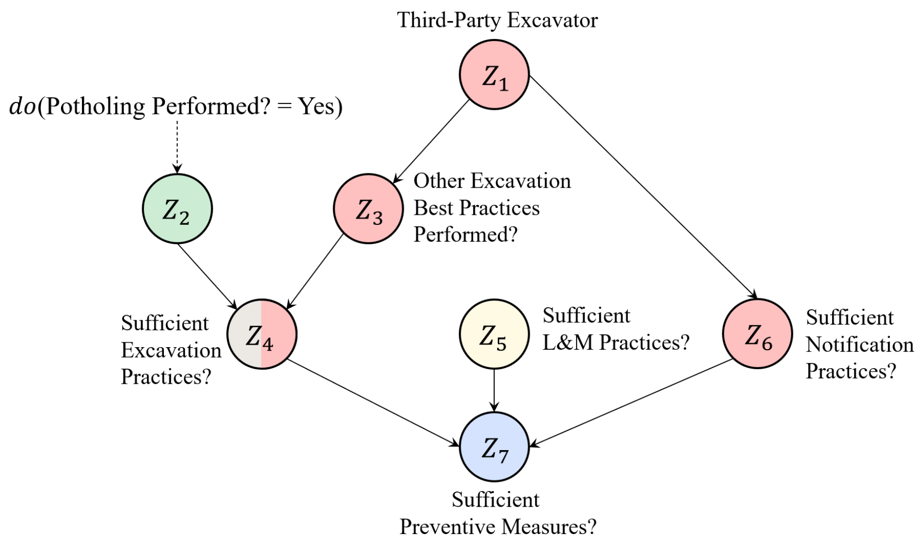


Figure 3.6: Post-intervention DAG for the TPD BN model shown in Figure 3.5.

Given that we have a fully parameterized BN model for the case study, the truncated factorization formula in Equation 3.2 is applied to identify the intervention query Q in Equation 3.12. Averaging over the nodes Z_7, Z_6, Z_5, Z_4, Z_3 and Z_1, Q can be computed by:

$$\begin{aligned} Q = \sum_{Z_6, Z_5, Z_4, Z_3, Z_1} & Pr(Z_7 \,|\, Z_6, Z_5, Z_4) Pr(Z_6 \,|\, Z_1) \\ & Pr(Z_5) Pr(Z_4 \,|\, Z_3, Z_2 = \text{Yes}) Pr(Z_3, \,|\, Z_1) Pr(Z_1) \end{aligned} \tag{3.13}$$

It is important to highlight that Equation 3.13 is possible to be used only because we have a fully parameterized BN model. However, a more common scenario is one in which a risk analyst has data on the treatment and outcome nodes but not for the other variables that compose their BN model. Nevertheless, as shown in Section 3.2.2, it is still possible to identify an intervention query using the back-door criterion or a sequential application of the *do*-calculus rules. This scenario will be further discussed in Section 3.4.4.

(iv) *Estimate the intervention query.* Using the available data for this case study, Equation 3.13 is estimated as:

$$Q = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes})) = 78.35\% \tag{3.14}$$

which shows the post-intervention probability of sufficiency of preventive measures to mitigate TPD in an excavation activity if potholing is made mandatory. The post-intervention probabilities for all nodes of the BN model are shown in Table 3.2.

Even though we can model the effect of making potholing mandatory on the probability of sufficient TPD preventive measures, not all excavators will likely follow this action. Therefore, we will also evaluate the results of Equation 3.13 for the imperfect intervention $Pr(do(Z_2 = \text{Yes})) = [40\%, 60\%, 80\%]$. All of these imperfect interventions consider that the probability of potholing will increase in an excavation if it is made mandatory. As such, the imperfect intervention represents uncertainty on the evaluated new action/policy. Figure

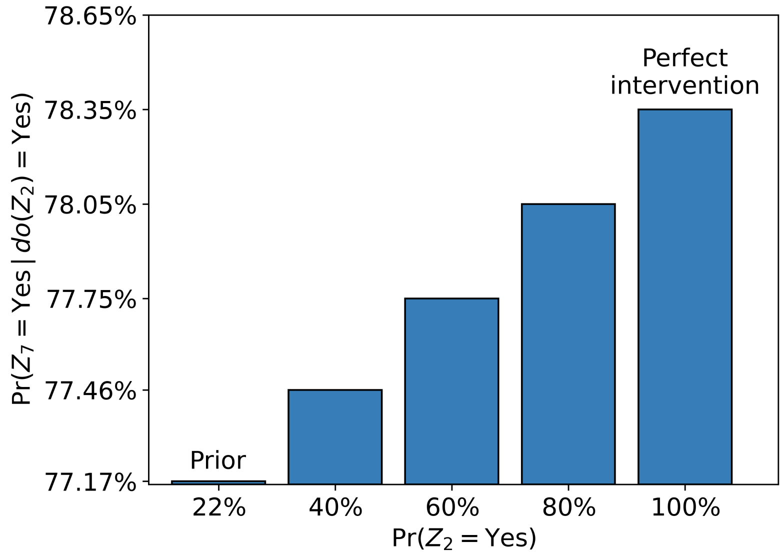


Figure 3.7: Effect of imperfect interventions on potholing (Z_2) on the probability of sufficient TPD preventive measures (Z_7).

3.7 shows the results of assuming an imperfect intervention on potholing.

3.4.3 Step 3: Inform Decision-Makers

To better inform a decision-maker about the effect of making potholing mandatory on the sufficiency of TPD preventive measures, the following IE, CCE, and CDE metrics are calculated.

(i) *Intervention Effect (IE).* Using the estimated intervention query in Equation 3.14 and the pre-intervention probability of $Y = Z_7$ shown in Table 3.2, IE can be estimated as:

$$IE = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes})) - Pr(Z_7 = \text{Yes}) = 1.18\% \tag{3.15}$$

This result informs decision-makers that making potholing mandatory can increase the likelihood of sufficient TPD preventive measures by 1.18% compared to previous conditions.

(ii) *Comparative Causal Effect (CCE).* To put the studied policy in perspective for decision-makers, the CCE of making potholing mandatory over the rest of CGA excavation best practices is calculated. In order to do this, the intervention query $Q = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes}))$ is going to be compared with $Q' = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes}),\, do(Z_3 = \text{Yes}))$. The latter can be estimated using Equation 3.2, obtaining:

$$Q' = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes}),\, do(Z_3 = \text{Yes})) = 78.52\% \tag{3.16}$$

Then, the CCE can be computed by:

$$\begin{aligned} CCE = Q' - Q = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes}),\, do(Z_3 = \text{Yes})) \\ - Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes})) = 0.17\% \end{aligned} \tag{3.17}$$

This result shows that making all CGA excavation best practices mandatory can increase the likelihood of preventive measures sufficiency by 0.17% over the effect of just making potholing mandatory with respect to pre-policy conditions. Decision-makers can use this result to put their decisions in perspective by comparing the effect of different policy alternatives.

(iii) *Comparative Direct Effect (CDE).* Last, to understand how much the effect of making potholing mandatory is mediated by performing good excavation practices in general, we can control for $Z_M = \{Z_4 = \text{Yes}\}$ and calculate the CDE as:

$$\begin{aligned} CDE = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes}),\, do(Z_4 = \text{Yes})) \\ -Pr(Z_7 = \text{Yes} \,|\, do(Z_4 = \text{Yes})) \end{aligned} \tag{3.18}$$

Then, the CDE is estimated using Equation 3.2 for $Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes}),$

$do(Z_4 = \text{Yes}))$ and the back-door criterion for $Pr(Z_7 = \text{Yes} \,|\, do(Z_4 = \text{Yes}))$ (using Z_1 as a sufficient admissible set for adjusting for confounding), obtaining:

$$CDE = 78.52\% - 78.52\% = 0.00\% \tag{3.19}$$

Comparing this value to $IE = 1.18\%$, it shows that the effect of making potholing mandatory over pre-policy conditions is completely mediated by performing proper excavation practices in general. This result shows that performing sufficient excavation practices is what mostly affects the sufficiency of preventive measures in the context of excavation practices.

As shown above, IE, CCE, and CDE metrics give the decision-maker information on the magnitude of the effect that making potholing mandatory can have on the sufficiency of preventive measures. These metrics also enable decision-makers to put this policy in perspective to other possible actions, such as making all CGA excavation best practices mandatory. Additionally, it is possible to inform decision-makers that performing sufficient excavation practices is what mostly affects the sufficiency of preventive measures in the context of excavation practices.

3.4.4 The Limitations of Relying on Associative Reasoning to Answer Intervention Queries

As stated in Section 3.2.2, usually $Pr(Y \,|\, do(X = x')) \neq Pr(Y \,|\, X = x')$ due to the presence of confounders between X and Y. This shows that intervention queries can not be properly estimated using associative reasoning by evidence propagation, which is the main BN tool used in engineering risk assessment. This limitation not only affects the estimation of causal effects but also affects how results from BNs are interpreted in risk analyses. The latter can significantly affect translating popular methods from past QRA models to BNs, such as importance measure analysis. This section used the TPD case study to provide two examples of the limitations of relying on associative reasoning to answer intervention queries.

3.4.4.1 Example 1: Estimating Causal Effects

Going back in the case study to Step 2.iv, we can try to answer the intervention query defined in Section 3.4.2 (i) through associative reasoning. To do this, we can use the BN model of Figure 3.5 to estimate $Pr(Z_7 = \text{Yes} \,|\, Z_2 = \text{Yes})$ using the factorization formula in Equation 2.3. The prior and posterior probabilities of the sufficiency of TPD preventive measures are presented in Table 3.3.

Table 3.3 shows a 0.77% decrease in the probability of sufficiency of preventive measures if potholing is performed, which is counterintuitive given that potholing is a recommended excavation best practice. However, an associative reasoning ap-

Sufficient Preventive Measures?	
$Pr(Z_7 = \text{Yes})$ [%]	77.17
$Pr(Z_7 = \text{Yes} \mid Z_2 = \text{Yes})$ [%]	76.40
Δ	-0.77

Table 3.3: Prior and posterior probabilities of "Sufficient Preventive Measures?" by associative reasoning with evidence on "Potholing Performed? = Yes".

proach was employed, so it is essential to interpret this result as just the evidential association that potholing has on TPD preventive measures' sufficiency in the used dataset, given the assumed BN model structure.

If we want to estimate the effect that potholing has on TPD preventive measures' sufficiency (which is an intervention query), spurious associations due to confounding between these two variables need to be adjusted to obtain an unbiased estimation. Based on the assumed BN structure shown in Figure 3.5, an unbiased estimate of the effect that "$Z_2 = \text{Yes}$" has on "$Z_7 = \text{Yes}$" can be obtained by blocking all confounding-inducing back-door paths between these two variables. Using the back-door criterion on the BN of Figure 3.5, we found that conditioning on the type of third-party excavator Z_1 blocks all back-door paths between Z_2 and Z_7. The latter means that $Z_2 \perp\!\!\!\perp Z_7$ for the paths $Z_2 \leftarrow Z_1 \rightarrow Z_6 \rightarrow Z_7$ and $Z_2 \leftarrow Z_1 \rightarrow Z_3 \rightarrow Z_4 \rightarrow Z_7$, which is equivalent to eliminating the incoming edges to Z_2 as done when the truncated factorization formula is used for intervention reasoning (see Equation 3.2).

Then, $Pr(Z_7 = \text{Yes} \mid Z_2 = \text{Yes}, Z_1)$ is computed using the factorization formula on Figure 3.5 and shown in Table 3.4.

Table 3.4 shows, as expected, an increase in the probability of sufficiency of

	Z_1: Third-Party Excavator		
Sufficient Preventive Measures?	Property Owner	Contractor	Government Entity
$Pr(Z_7 = \text{Yes} \,|\, , Z_1)$ [%]	61.20	80.40	89.00
$Pr(Z_7 = \text{Yes} \,|\, Z_2 = \text{Yes}\,,\, Z_1)$ [%]	61.75	81.69	90.86
Δ	+0.55	+1.29	+1.86

Table 3.4: Prior and posterior probability of "Sufficient Preventive Measures?" by associative reasoning with evidence on "Potholing Performed? = Yes" and conditioned on each specific type of third-party excavator.

preventive measures if potholing is performed for each type of third-party excavator. Now, for the aggregated population of excavators, the causal effect that potholing has on the sufficiency of TPD preventive measures can be calculated using intervention reasoning by applying the back-door formula (see Equation 3.6) as:

$$Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes})) = \sum_{Z_1} Pr(Z_2 = \text{Yes} \,|\, Z_2 = \text{Yes}, Z_1) Pr(Z_1) = 78.35\% \tag{3.20}$$

which is analogous to the result obtained using the truncated factorization formula in Section 3.4.2 (iii). It is important to note that the difference between the result shown in Equation 3.20 and the one in Table 3.3 is that the former contains only information on the causal path $Z_2 \to Z_4 \to Z_7$, whereas the latter also contains information on the spurious backdoor paths between Z_2 and Z_7.

Misleading causal effect estimations based on associative reasoning as the one encountered between Tables 3.3 and 3.4, where a positive association between the studied intervention and outcome for each type of excavator reverses for the combined population, is a widely studied phenomenon called "Simpson's Reversal" [156]. Pearl [157] used this phenomenon as an extreme example of the type of biases that improper handling of confounding variables can have on estimating an intervention query, demonstrating that it is possible to remediate it by adjusting for confounders that can bias the causal effect being estimated. For the Simpson's Reversal observed in this case study, it can be explained by the fact that, even though a pothole is performed, excavators tend to fail in preventive measures by not notifying their intent of excavation (see Appendix A). This spurious confounding association must be blocked to obtain an unbiased causal effect.

Misleading causal effect estimations based on associative reasoning as the one encountered between Tables 3.3 and 3.4, where a positive association between the studied intervention and outcome for each type of excavator reverses for the combined population, is a widely studied phenomenon called "Simpson's Reversal" [156]. Pearl [157] used this phenomenon as an extreme example of the type of biases that improper handling of confounding variables can have on estimating an intervention query, demonstrating that it is possible to remediate it by adjusting for confounders that can bias the causal effect being estimated. The Simpson's Reversal observed in the case study can be explained by the fact that, even though a pothole is performed, excavators tend to fail in preventive measures by not notifying their intent of excavation (see Appendix A). This spurious confounding association must be blocked to obtain an unbiased causal effect.

Confounding adjustment is easily performed in our case study by using the truncated factorization formula as done in Section 3.4.2 (iii). However, in the much more common scenario where a dataset does not contain information on all of the considered BN model variables, an intervention query can still be estimated using the back-door criterion or the do-calculus rules. For example, our case study query $Q = Pr(Z_7 = \text{Yes} \,|\, do(Z_2 = \text{Yes}))$ can still be adequately estimated using the back-door criterion as shown in Equation 3.20, where only data on the treatment and outcome variables Z_7, Z_2 and the confounder Z_1 is needed to compute it.

3.4.4.2 Example 2: Importance Measures Analysis

Importance measures (IM) are a popular set of indices used in QRAs for risk-informed decision support [2, 46]. They allow prioritizing variables within a system or process for prevention, safety, and maintenance actions. IMs quantify a degree of change of the failure (or success) probability of an outcome of interest, conditioned on making the occurrence (or non-occurrence) of a basic event certain (e.g., assigning perfect reliability to a component or assuming that a damage prevention measure is perfect). IMs are traditionally computed within a Fault Tree analysis; however, an increasing interest has been observed in applying IM analysis to BN models. A Fault Tree assumes an independent set of basic events that leads, through a Boolean combination of these, to an outcome of interest. As such, an IM calculated on a basic event will be equivalent to “intervening” on the basic event to force it to take the desired state (e.g., perfect reliability) and propagate that information throughout the Fault Tree. Thus, IM analysis is akin to an intervention query. BNs have been used in QRA to model more complex causal interactions among variables in a system, which will thus result in the addition of common causes and dependencies between basic events. As such, the calculation of an IM in a BN is not as straightforward as setting evidence on a basic event and then propagating it to an outcome of interest. This practice, unfortunately, has been observed in modern QRA due to a lack of understanding of the differences between association and intervention reasoning in BNs.

The TPD case study will be used to exemplify the misconceptions associated with translating IM analysis to BN-based QRAs. Four basic events are considered for an IM analysis in the BN of Figure 3.5: Z_2, Z_5, Z_6. To illustrate, the Risk Achievement Worth (RAW) IM will be used to inform which of these three basic events should be prioritized for preventing TPD. The RAW IM indicates the basic event that most contributes to the probability of an undesired outcome of interest, i.e., which basic event is most responsible for "achieving" the risk. Reducing the occurrence probability of undesirable states of events Z_2, Z_5, Z_6 is considered to reduce the probability of insufficient preventive measures in excavations, so RAW will be applied to prioritize actions in these variables.

The BN translation of the RAW IM proposed in Noroozian et al. [158] for a two-state node in a BN is implemented and will be calculated for the BN of Figure 3.5 as:

$$I_{RAW} = \frac{Pr(Z_7 = \text{No} \mid Z_i = \text{No})}{Pr(Z_7 = \text{No})} \tag{3.21}$$

As can be seen, the value of I_{RAW} in Equation 3.21 is expected to be higher for basic events whose failure will cause the highest increase in the probability of insufficient preventive measures in excavations. Furthermore, it is expected that $I_{RAW} \geq 1.0$.

These statements are valid for a Fault Tree, as a basic event is independent of other basic events, and a change in its failure probability will be directly propagated to an outcome of interest. However, the BN model of Figure 3.5 presents two back-door paths between Z_2 and Z_7 and two other paths between Z_6 and Z_7. Therefore, confounding bias will limit the interpretation of the IM. In order to keep the definition and interpretability of RAW, intervention reasoning should be used as follows:

$$I_{RAW}^{corrected} = \frac{Pr(Z_7 = \text{No} \mid do(Z_i = \text{No}))}{Pr(Z_7 = \text{No})} \tag{3.22}$$

If Equations 3.21 and 3.22 are used on the variables Z_2, Z_5, Z_6 for computing RAW, the results of Table 3.5 are obtained. As can be seen, I_{RAW} for "Z_2 : Pot-hole performed?" produces an invalid value, 0.990, for the metric. This is caused by the back-door path $Z_2 \leftarrow Z_1 \rightarrow Z_6 \rightarrow Z_7$; the same path that produced the Simpson's reversal in the example of Section 3.4.4.1. This error is not observed for the value of $I_{RAW}^{corrected}$, as confounding bias is properly handled by the $do()$-operator. Additionally, we can see discrepancies in the values of RAW for Z_6, where I_{RAW} slightly underestimates the effect of a failure of Z_6 on Z_7. Lastly, it is important to see that $I_{RAW} = I_{RAW}^{corrected}$ for Z_5, as it is an independent basic event in the model. This last result shows that importance measures calculated for independent basic events in a BN are expected to present the same interpretation as in a Fault Tree model. This has been the case in works such as [158–160] in which the used BNs are a direct translation of a Fault Tree model. However, additional precautions should be taken in BNs that account for more complex causal dependencies between variables in a system, which can hinder the correct calculation and interpretation of IMs if adequate causal reasoning methods are not considered. Unfortunately, the lack of distinction between associative and intervention reasoning has led authors to suggest potentially erroneous formulations of IMs analysis with BN models, such as in [158, 161, 162].

Node	I_{RAW}	$I_{RAW}^{corrected}$
Z_2	$0.990 \ngeq 1.0$	1.014
Z_5	1.690	1.690
Z_6	2.059	2.061

Table 3.5: RAW calculated with and without considering methods for intervention reasoning in BNs.

As demonstrated by the two examples in this section, intervention queries cannot be adequately estimated nor informed to decision-makers in a risk management phase by using associative reasoning, highlighting the importance of using intervention reasoning methods if an intervention query wants to be answered appropriately.

3.5 Concluding Remarks

Modern QRAs based on BNs have uniquely relied on associative reasoning to support decision-making and risk management. However, QRA could highly benefit from expanded causal insights based on intervention reasoning. The latter is a type of causal reasoning that is admissible within a BN approach to QRA, which could solve complex problems such as estimating the effect of policies and actions before being implemented. In this chapter, we established the mathematical background and methods required to model interventions with BNs, showing how to properly express their post-intervention joint probability distribution by means of the *do*-operator. Furthermore, this research shows that the methods established for intervention reasoning ensure unbiased estimations of the causal effect of interventions on an outcome of interest, even in the presence of confounding. This last feature not only expands the type of causal insights that a QRA can generate to support risk management but also provide the means to translate crucial methods, such as IMs, to the analysis of complex engineering systems and processes. Lastly, a BN-based framework was developed and proposed to inform decision-makers of the effect that their action can have on a system risk. The applicability of this framework was shown in an illustrative case study on the risk of TPD to natural gas pipelines. This case study demonstrated that intervention reasoning is readily available for risk analysts to expand causal insights on the risk of a system, which is the main goal of QRA. The results of RA1 shown in this Chapter made the following contributions to the risk analysis field:

- Established the mathematics to enable intervention reasoning in modern QRA models based on BNs.
- Established the requirements to enable intervention reasoning in QRA models based on BNs.
- Developed a framework and a set of causal metrics to calculate and inform the causal effect of a decision on a system's risk.
- Demonstrated that associative reasoning could not support insights based on interventions.

- Demonstrated the applicability of interventions in an illustrative case study on TPD to natural gas pipelines.

To the best of the authors' knowledge, this research is the first attempt to expand the causal reasoning capabilities of BN-based QRAs by means of interventions to support risk management. This work's results enabled us to achieve the first objective of this research and serve as a baseline toward achieving O2: establish the mathematical background and methods needed to perform counterfactual reasoning with BNs in risk assessment. Intervention reasoning can be applied to a BN model of a system or process if the model includes all confounding variables and back-door paths between the intervened variables and the outcome of interest. However, as explained in Chapter 2, reasoning about counterfactuals requires a comprehensive causal model of a system or process, as the background conditions of a past event must be simulated to generate causal insights into it. Therefore, the next chapter of this dissertation will be focused on developing a comprehensive BN model of the problem of TPD to natural gas pipelines in the U.S. This model will then be used in Chapter 5 to enable counterfactual reasoning in modern QRA models.

Chapter 4: BaNTERA: A Bayesian Network for Third-Party Excavation Risk Assessment

4.1 Introduction

This chapter presents the process and results of building BaNTERA (Bayesian Network for Third-Party Excavation Risk Assessment), a comprehensive probabilistic model for assessing natural gas pipeline hit and subsequent damage probabilities in excavation activities in the U.S. The results of this chapter correspond to RT2.1, which is executed toward achieving O2 of this dissertation as explained in Chapter 1. BaNTERA is built with the objective of presenting a comprehensive causal model representing an analyst's knowledge of an excavation scenario. This comprehensive causal model will include not only causes and consequences of interest but also the background information in which they occur. To do this, BaNTERA combines a wide variety of data and causal information from multiple information sources. Additionally, BaNTERA's associative reasoning capabilities (rung 1 in the ladder of causality) will be used to expand upon insights gained from historical statistics, which can then be used to mitigate potential excavation-specific risks or drive more appropriate risk management actions.

The rest of this chapter is structured as follows. Section 4.2 presents the necessary background information on TPD and BNs to develop this work. This is followed by a description of the methodology used to develop BaNTERA in Section 4.3. Then, in Sections 4.4-4.6, further information is provided about building BaNTERA, including the formulation of a TPD-specific taxonomy and high-level model architecture, the construction of BaNTERA's structure, and the strategies behind data selection and model parameterization. After presenting BaNTERA's parameterization results in Section 4.7, its probabilistic inference capabilities are applied to a number of different case study excavation scenarios, as well as a real-world industry application, in Section 4.8. A discussion of the model, its implications, and limitations are then included in Section 4.10, followed by concluding remarks in Section 4.11.

The results of RT2.1 were peer-reviewed and published in the journal *Reliability Engineering and System Safety* [23].

4.2 Background

This section provides general background information relevant to the development of BaNTERA.

4.2.1 Third-party Excavation Process and Site Description

The general excavation process considered for BaNTERA is structured on the best dig-in practices provided by the Common Ground Alliance (CGA; an organization that promotes damage prevention across underground utilities in the U.S.) around the use of the 811 One Call system [20,152]. These practices were considered to use the most generic damage prevention data-collection process in the U.S. In particular, the CGA's Damage Information Report Tool (DIRT) [20] incorporates data from multiple stakeholders from the natural gas industry, making it a comprehensive summary of data collection and reporting practices on excavation damage to U.S. underground pipelines.

According to the CGA, prior to the excavation, the third-party excavator marks the anticipated digging area. He or she then calls the local 811 One Call center or submits an online form providing dig and site information at least two working days before digging. If the location contains underground facilities, the responsible utilities are notified, and a utility service representative locates and marks the utility's facilities. Following this, the One Call center notifies the excavator to begin the excavation. The excavator proceeds to dig safely by exposing underground utilities using hand tools to avoid accidental damage. Failure to appropriately execute any of these steps may result in a gas pipeline getting hit. If the hit occurs with a strong enough force to overcome the pipe's inherent resistance, TPD in the form of a puncture failure can occur.

Figure 4.1 shows a conceptual diagram of the primary entities, both actors and objects, at an excavation site; this defines the scope of the excavation context within which BaNTERA could be applied. The three main actors are the third-party excavators, the owner or manager of the pipeline, and the One Call notification center. Each has characteristics that pertain to on-site actions as well as broader

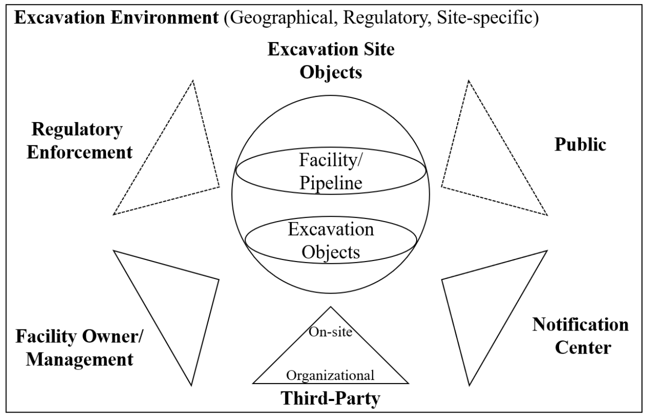


Figure 4.1: Conceptual diagram of an excavation site. The context of an excavation site consists of actors (triangles) interacting with site objects (ovals) within a larger excavation context.

organizational practices. Although the public and regulatory enforcement could be considered key stakeholders, they do not directly interact with the excavation site. As such, when building BaNTERA, they were considered part of the environment where the excavation occurs.

In addition to the actors mentioned above, the excavation site also consists of the facility and other objects pertaining to the excavation, including the excavator, facility maps and records, and the notification ticket. The environment in which the interactions between the actors and the objects occur relates to the specific circumstances present at the excavation site. Example environments include geographical, regulatory, and site-specific contexts.

4.2.2 Bayesian Networks

An excavation process is comprised of multiple factors that may impact the likelihood of TPD. Moreover, these factors are usually causally dependent on each other. These features must be considered when choosing a modeling approach.

Therefore, a BN approach (as described in Section 2.3) was selected to address the difficulty of modeling TPD by representing how risk influencing and other causal factors interact at an excavation site and lead to damage.

As explained in Section 2.3, associative reasoning with BNs can provide the joint probability distribution of a system's variables when updated with new knowledge about the system via evidence on the model's nodes. The factorization formula (see Equation 2.3) can be used in addition to Bayes's theorem to reason about the associations between a BN's nodes. The following three types of probabilistic inferences based on associative reasoning with BNs will be used in this study:

- *Forward inference:* primarily predictive, this inference type propagates evidence from causes to effects. Figure 4.2a shows how evidence $V_1 = v_1$ can affect $Pr(V_3 \mid V_1 = v_1, V_2)$.
- *Backward inference:* primarily diagnostic, this inference type propagates evidence from effects to causes. Figure 4.2b shows how evidence $V_3 = v_3$ can diagnose $Pr(V_1 \mid V_3 = v_3)$.
- *Intercausal inference:* primarily explanatory, this inference type propagates evidence between causes with a common effect. Figure 4.2c shows how evidence $\{V_1 = v_1, V_3 = v_3\}$ can explain away $Pr(V_2 \mid V_1 = v_1, V_3 = v_3)$.

4.3 Methodology for Developing BaNTERA

A multi-step process was used to build BaNTERA as a comprehensive decision-support tool for third-party excavation. This section presents the methodology used

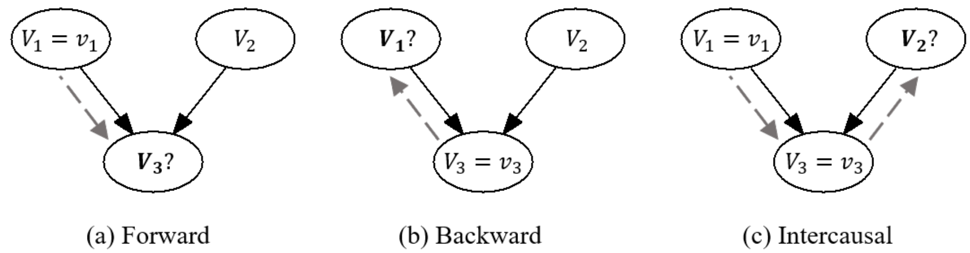


Figure 4.2: Visual representation of BN probabilistic inference reasoning methods. Dashed lines indicate the flow of knowledge towards the specific node of interest, marked with a question mark.

to create the model, including building a taxonomy for the excavation context, developing the model's structure, and parameterizing the model nodes. These features were obtained by performing the following 3 steps:

1. *Define a taxonomy and high-level model architecture.* Prior to model development, a high-level description of the current processes, actors, and objects involved in a third-party excavation scenario was outlined to bound the scope of BaNTERA and to determine the dependencies among the variable's considered for its construction. This was done by first developing a structured hierarchical taxonomy and then identifying the model's architecture and dependencies of the risk-influencing factors (RIFs) involved in such a scenario.

2. *Define BaNTERA's structure.* The model's structure was subsequently defined by placing the different elements from the taxonomy into the model's high-level dependency architecture.

3. *Process the data and parameterize BaNTERA.* Formulating the conditional probability dependencies between BaNTERA's nodes requires the integration of different and distinct data sources. A multi-step parameterization methodology was used to parameterize BaNTERA. This methodology's input is a set

of data sources and a BN structure. Then, through data processing and node parameterization steps, a parameterized BN model is obtained.

Each portion of the abovementioned methodology is described in Sections 4.4-4.6.

4.4 Taxonomy and High-level Model Architecture

4.4.1 Third-party Damage Excavation Taxonomy

Based on the processes and conceptual excavation site described in Section 4.2.1, a detailed hierarchical taxonomy of the third-party excavation scenario was built. This taxonomy is separated into the three large categories identified in the conceptual excavation site ("Actor," "Objects," and "Environment") and provides both broad and specific terms. The full taxonomy is provided in B.1, and a summary overview is provided below.

The "Actor" section in the taxonomy provides terms and concepts associated with the third-party excavator, the facility owner and manager, and the One Call notification center. The subsections within the third-party and facility owner terms focus around on-site actors and practices, while the terms associated with the notification center are related to its performance quality.

There are two main groupings within the taxonomy's "Objects" category, which captures the equipment and physical items found at the excavation site. The first group is associated with the characteristics and nature of the pipeline, while the second group comprises the excavator machinery and other physical objects associated with the excavation site. The pipeline is described in more detail than other objects at the site, as understanding the pipeline's physical features can lead to a better understanding of the pipe's resistance to potential damages, such as a puncture.

The final category, "Environment," provides further details about the context and background in which the excavation occurs. This is further split into three smaller context groups. The "physical environment" describes the geography and other natural hazards present at the site, while the "legal environment" covers regional and regulatory oversight. Last, "excavation environment" describes the excavation site with respect to the current and previous excavations.

4.4.2 Relevant Factors, Hierarchical Structure, and Associated Dependencies

The third-party excavation taxonomy illustrates a wide range of contextual factors that were considered when creating BaNTERA. The likelihood of pipeline damage at these sites is impacted by these different factors and should be captured in a meaningful structure within the model. Figure 4.3 presents the conceptual diagram of the relevant factors, hierarchical structure, and associated dependencies used to model TPD within BaNTERA. This diagram is based on expert knowledge, PHMSA incident reports causal factors, and CGA incident reports causal factors. As shown in Figure 4.3, at the bottom of BaNTERA are *target variables*, which represent scenario summary information, such as the sufficiency of pipe hit preventive measures, the

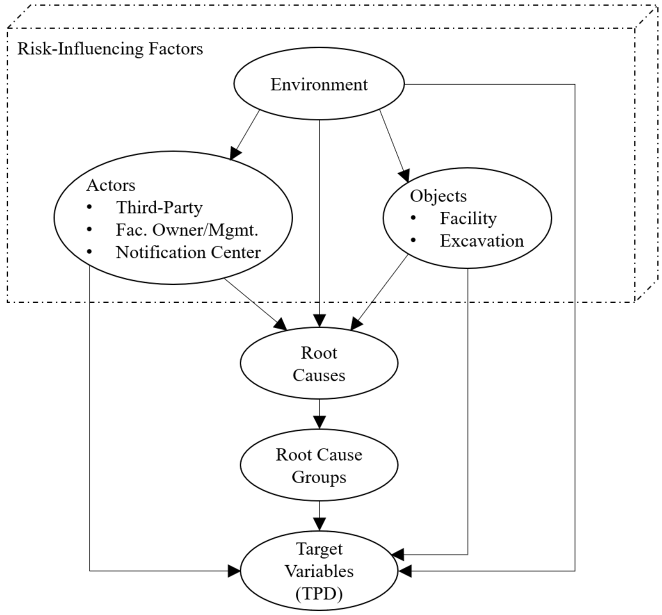


Figure 4.3: Conceptual diagram of third-party damage model outlining the causal flow from excavation context nodes and risk-influencing factors context nodes to root causes and target variables.

pipe hit itself, and subsequent damage. Target variables depend on the other node types existing in the excavation process space; one such type consists of the groups of root causes that contribute to TPD. These groups capture accumulated relationships of similar pipeline damage root causes. These groups are similar to those expressed by the CGA's DIRT [20], which are based on distinct excavation steps (notifying, locating, and excavating). The CGA DIRT Data Committee identified six different root cause groups to provide a high-level perspective of what can go wrong when following CGA-proposed best excavation practices. These groups are "excavation practices," "invalid use of request," "locating practices," "miscellaneous," and "no locate request."

The next level of model nodes in BaNTERA to consider are the root causes and the context variables that affect them. This part of BaNTERA has the greatest number of variables that can vary during the excavation process, making this part of the model the most interactive. Root causes defined by the CGA DIRT are considered for its construction. This decision was taken due to the wide use of CGA DIRT root causes by utility companies. Moreover, the number and definitions of root causes considered by the CGA DIRT are more comprehensive (i.e., granular) than other sources such as PHMSA's. However, CGA DIRT's root causes are mainly focused on the excavation process itself and not on its context. To expand the range of acknowledged factors that can contribute to TPD target variables and, in particular, their root causes, a set of risk-influencing factors (RIFs) were defined based on the context of an excavation. According to the taxonomy presented in Section 4.4.1, RIFs were distinguished into two categories by what kind of measures could be adopted to mitigate them: actor-based RIFs (e.g., encouraging excavation best practices, increasing awareness on One Call centers, among others) and object-based RIFs (e.g., improving pipeline materials, maintenance of locating technologies, among others). These factors are affected by a third RIF, the environment where the excavation occurs. According to the taxonomy defined in Section 4.4.1, this environment can be physical (e.g., geographical features), legal (e.g., regulatory oversight), or related to the excavation activities themselves (e.g., excavation depth).

4.5 BaNTERA Structure

BaNTERA was structured using the defined nodes and arcs reflecting node relationships and dependencies identified in Figure 4.3. The result is a comprehensive model that captures the root causes of TPD and provides a method of expressing failure through context described within the taxonomy and the overarching RIFs. Figure 4.4 shows BaNTERA's structure.

Similar to the high-level generalized conceptual model shown in Figure 4.3, BaNTERA's structure consists of nodes that provide different functions within the model and different types of information. At the bottom of the structure are the summary target nodes that combine root cause information and excavation-related context. These provide information into potential occurrences of pipe contact, puncture, and damage during a third-party excavation. The parents of the target nodes are the large root cause sub-group and group nodes, which capture accumulated relationships of similar root causes. These groups are similar to those expressed in the DIRT report, which are based on distinct steps of the excavation process (i.e., notifying, locating, excavating). The next step up in the structure are specific root causes and the risk-influencing nodes that affect them. This part of the structure has the greatest number of nodes that can vary during the excavation process, making it highly interactive. Lastly, the top of the structure consists of overarching context nodes. These nodes represent the environment in which the excavation is taking place.

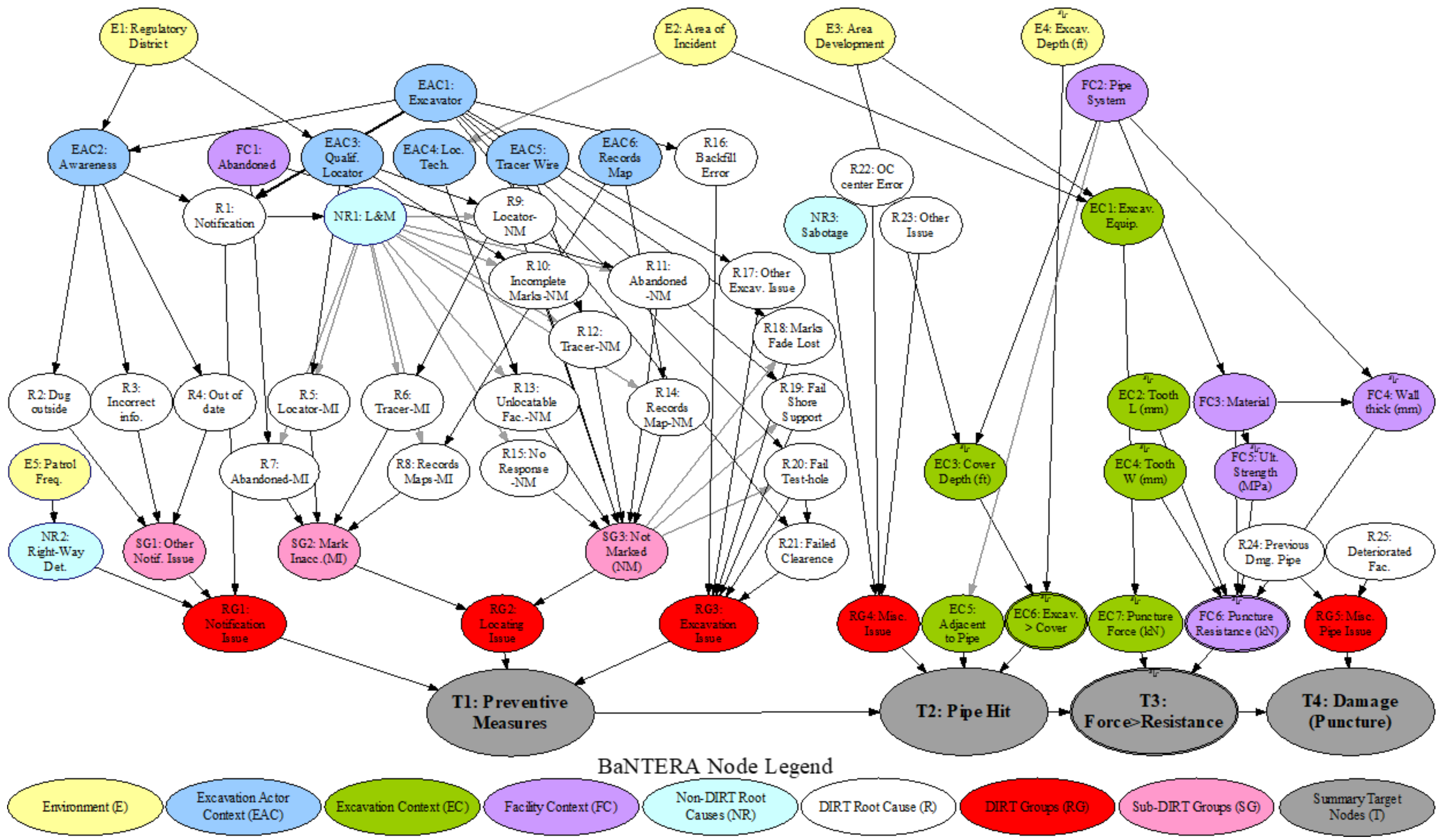


Figure 4.4: BaNTERA's structure. A detailed list of the nodes and node characteristics is presented in B.2. Visualization made in GeNIe. Color legend: gray - target summary nodes; red - DIRT root cause group; pink - DIRT root cause subgroup; white - DIRT root cause; light blue - non-DIRT root cause; blue - third-party excavation actor context; yellow - environment context; purple - facility context; green - excavation context.

The sixty-four nodes captured in BaNTERA's structure represent different causal contexts and root causes for TPD. The color of the nodes reflects what kind of information they provide. As shown in Figure 4.4's legend, white, light blue, red, pink, and gray nodes describe specific root and general causes for TPD during an excavation. There are forty of these root cause nodes within BaNTERA. White and light blue nodes are individually described root causes of damage, while red nodes indicate the formation of a root group based on a number of similar root causes. Pink nodes are used to provide an initial amalgamation of root causes into smaller root cause groups, which are then joined to form a root cause group. Gray nodes describe the summary target nodes that BaNTERA is trying to represent. The other twenty-four nodes represented in BaNTERA's structure describe the context in which the TPD occurs. Yellow nodes describe the overall excavation environment nodes that are specific to the context of the excavation site, while the blue, green, and purple nodes relate to context factors associated with the specific excavation site scenario. Blue nodes describe the context associated with the third-party excavation actors, which includes both the excavator itself and the utility locator. Green nodes describe the context of the excavation site, while purple nodes are used to represent pipeline characteristics. These causal context factors can also be placed within the overarching RIFs described in Section 4.4.2.

According to PHMSA [163], there are four pipeline damage types due to excavation activities: dents, gouges, coating damage, and puncture. In this current version BaNTERA, puncture damage is the only damage type considered for modeling TPD. This decision stemmed from the fact that, to the best of the authors' knowledge, Brooker's puncture damage model [164–166] was the only one that provided a level of granularity sufficient enough to be included in BaNTERA's complex dependency structure. The implications of this decision are discussed in Section 4.10.

4.6 Data Processing and BaNTERA Parameterization

Formulating the conditional probability relationships between BaNTERA's nodes requires the integration of different and distinct data sources. A conceptual structure to the multi-step methodology used to parameterize BaNTERA, consisting of data processing and node parameterization, is represented in the diagram in Figure 4.5. Each portion of the parameterization methodology is described in the following sections.

4.6.1 Data Sources for Parameterizing BaNTERA

After defining BaNTERA's structure in Section 4.5, different data sources were selected to parameterize its CPTs and CPFs. These data sources have varying amounts of data granularity, redundancy, and quality. Data was limited to those that contained conditional dependency information between BaNTERA's nodes. In order to reflect the present conditions of pipeline damage risk, only data from 2016-present was used during the parameterization process.

The following four data sources were used to parameterize BaNTERA:

- *Partner utility company damage reports (Source A).* This data source con-

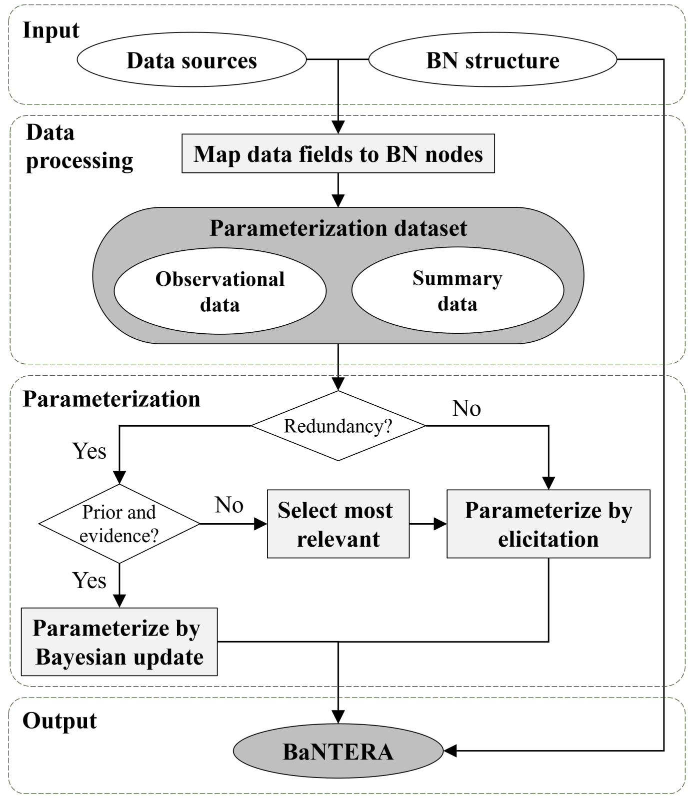


Figure 4.5: Graphical representation of BaNTERA's parameterization methodology.

tains more than 7,000 TPD reports from a U.S. partner utility company between 2016-2020. These reports provide more than forty different context and damage cause-related fields, including "root cause," "excavation tools," and "excavator type," among others. A semiotic data quality assessment was successfully performed to make sure that these data represent an excavation activity as needed for modeling TPD risk [167]. Source A's high granularity, low data redundancy, and limited missing data make it an ideal data source to inform BaNTERA.

- *PHMSA's gas distribution and transmission incident data. (Source B).* Since 1970, PHMSA has maintained records of incident data [168]. Due to Source B's high quality, granularity, and public availability, numerous TPD models include it in its parameterization process [120–122]. However, the amount of useful entries in this data is relatively small, in the order of 200. In addition, this source has low granularity on specific root causes. Given these issues, this data source is only used on nodes that are not present on more comprehensive sources.

- *CGA's 2018 and 2019 DIRT database (Source C).* CGA's annual DIRT database [20, 153] provides an extensive set of summary statistics on the different aspects of excavation and its risks based on voluntarily provided damage reports between 2018-2019. Each year, approximately 35,000 reports are used to estimate these summary statistics. As the damage reports themselves are not available for public use, this data source is useful for node validation and as a starting point for parameterizing node relationships not present in more comprehensive sources.
- *GTI Energy previous models (Source D).* In the development of BaNTERA, GTI Energy provided three BN models developed in 2019 designed to address TPD and locating and marking risks during excavation activities [122, 123]. These BNs were parameterized by aggregating damage reports data from 2016-2018 provided by multiple utility companies throughout the U.S. Consequently, these models are useful sources of validation and prior information for nodes that can be updated with more recent data. These BNs also provide information on complex variable constructions, such as "sufficiency of preventive measures," that must be defined by experts in the field. Collectively, these

models provide information on 74 different variables relevant to an excavation process.

4.6.2 Data Processing

Data from the selected data sources described above were processed into a format that is consistent with BaNTERA's structure. To do this, each data source field was first mapped to a node in the network structure. Most nodes in BaNTERA could be linked to relevant fields across the different data sources, making the mapping process a straightforward one. Field redundancy between data sources was reduced by standardizing names using CGA's DIRT reporting field names; for example, the root cause field in Source A called "Excavated outside delineated area" was renamed "Excavator dug outside area described on ticket" to align with CGA DIRT's reporting scheme.

The end result is a comprehensive parameterization dataset for BaNTERA consisting of observational data as well as summary data. The observational portion consists of a collection of features from TPD incidents on natural gas pipelines. This dataset includes the collections of damage reports found in Sources A and B, as well as data concerning the relative counts of any instances of interest, such as the CPTs and CPFs of the BNs provided by Source D. In contrast, the summary portion of the parameterization dataset includes data that represents a summary statistic relevant to a specific node in BaNTERA's structure, such as the percentages of TPDs across U.S. regulatory districts provided in Source C. Both the observational and summary portions of the dataset were used to parameterize different parts of BaNTERA based on the nature of the node of interest. For instance, nodes representing specific root causes and their conditional dependencies were parameterized as a CPT using the observational dataset, while nodes representing broad generalizations of factors were parameterized using the summary dataset.

4.6.3 Node Parameterization Strategies

Two different types of parameterization strategies were used based on whether there was information redundancy or previously defined causal relationships available: elicitation and Bayesian update.

4.6.3.1 Parameterization by Elicitation

Since four different independent data sources were used to parameterize the nodes within BaNTERA, information redundancy exists across the summary and observational data in the parameterization dataset. As shown in Figure 4.5, parameterization by elicitation was used if only one piece of redundant data is selected as relevant or if only one piece of information is found on the parameterization dataset. The elicitation methods used vary whether a node represents a discrete or continuous variable. Discrete nodes were parameterized by the following methods:

- *Relative count of events.* CPT parameters derived from this approach were based on the frequency of occurrence for each particular node state in a dataset. For example, node "E2: Area of Incident" CPT was parameterized by counting the number of excavations in the dataset that were performed on exposed ground, soil, pavement, or other, between 2016-present.

- *Conditional assignment.* Discrete nodes representing a Boolean relationship between its parents were parameterized by conditional assignment. For example, node "EC6: Excav. > Cover" was parameterized by counting the instances in which the node's parent "E4: Excav. Depth (ft)" values were greater than its other parent "EC3: Cover Depth (ft)."

For continuous nodes, the following elicitation methods were used:

- *Distribution fitting.* Nodes presenting continuous data were modeled by fitting a probability distribution over the data. For example, node "EC3: Cover Depth (ft)" was parameterized by fitting a truncated Normal distribution to the depth of cover of pipelines involved on excavations performed between 2016-present.
- *Histogram.* In some instances, a clearly defined probability distribution does not match the values observed in the data. In those cases, like node "E4: Excav. Depth (ft)," a histogram was defined by experts over ranges of observed excavation's depth values.
- *Functional equation.* Nodes that represent physical models were parameterized through functional equations. For example, BaNTERA's "FC6: Puncture Resistance (kN)" node was parameterized using Brooker's puncture model [164–166].

4.6.3.2 Parameterization by Bayesian Update

As shown in Figure 4.5, this parameterization strategy was used for nodes in which redundant information on them not only existed across the parameterization dataset, but also where one piece of information could be considered as prior information. Any remaining information was then treated as new evidence. This parameterization is used, for example, when some of Source D's previous BNs node's CPTs are updated with more recent data from Sources A and B.

As this condition only held for discrete nodes, they were parameterized through a Beta-Binomial Bayesian update procedure. By assuming that a node's CPT entries are independent and represent the expected success probability p of independent binomial trials of hypothetical excavation scenarios, the node's uncertainty can be modeled as $p \sim Beta(\alpha, \beta)$. If the trust in the prior data is high, the method of moments [169] was used to define a Beta distribution prior for $p \sim Beta(\alpha_0, \beta_0)$ as:

$$\begin{aligned} \alpha_0 &= \frac{\mu^2 - \mu^3}{\sigma^2} - \mu^2 \\ \beta_0 &= \frac{\alpha_0(1-\mu)}{\mu}, \end{aligned} \tag{4.1}$$

where μ and σ are the mean and standard deviation of the prior data. On the other hand, if the trust in the prior data was low, the following weak prior for α_0 and β_0 was used [170]:

$$\begin{aligned} \alpha_0 &= 0.5 \\ \beta_0 &= \frac{1-\mu}{2\mu} \end{aligned} \tag{4.2}$$

The probability p was then updated using the data considered "new evidence" as:

$$p = \frac{\alpha_0 + k}{\alpha_0 + \beta_0 + n}, \tag{4.3}$$

where k is the number of occurrences of a node's CPT entry and n is the total number of observations.

4.7 BaNTERA's Parameterization Results

The result of Section 4.6 is a fully parameterized version of BaNTERA's structure presented in Section 4.5. The obtained prior probabilities of BaNTERA's target nodes are calculated by simulating 10^6 excavations and can be found in Table 4.1. An example excavation simulation can be found in B.3. These simulations were run using the GeNIe software through a Python-based PySMILE wrapper [1]. For the rest of this work, this same simulation process will be used to report the probability results on a node in BaNTERA as an expected value. To capture the variability introduced by the continuous nodes in BaNTERA, 95% equal-tailed credible intervals are reported for each expected probability (i.e., the interval which contains 95% of the simulated expected probabilities).

In order to verify that BaNTERA reflects the damage rates present in the data sources used for its parameterization, we compare its output to Sources A and D average TPD rates per 1,000 notified excavations. In addition, a 10% spread

Target Nodes	States	Expected Probability	95% CI [Lower, Upper]
T1: Preventive Measures	Sufficient	0.88832	[0.88340, 0.89725]
	Insufficient	0.11168	[0.10275, 0.11660]
T2: Pipe Hit	No	0.99689	[0.99590, 0.99830]
	Yes	0.00311	[0.00170, 0.00410]
T3: Punc. Force > Resistance	No	0.99747	[0.99670, 0.99870]
	Yes	0.00253	[0.00130, 0.00330]
T4: Damage (Puncture)	No	0.99746	[0.99660, 0.99880]
	Yes	0.00254	[0.00120, 0.00340]

Table 4.1: Prior expected probabilities and 95% equal-tailed credible intervals for BaNTERA's target nodes based on 10^6 simulated excavations. Five significant digits are used to present these results. This is done to show summation to one for probability values; however, this feature should not be considered as the precision of the BaNTERA.

local sensitivity analysis was performed to account for model uncertainty [1]. These comparisons are shown in Table 4.2. It is important to highlight that Sources A and D damage rates are for all types of excavation damages, not just puncture (as BaNTERA's output). In addition, although Sources A and D damage rates can be seen as valuable sources of information for BaNTERA's validation, they do not necessarily correspond to nationwide statistics (e.g., source A only considers data from one state). Given this, only Source A and D are used for verification purposes.

BaNTERA was validated by checking its consistency with historical industry data. The natural gas industry generally maintains records on excavation damage rates per 1,000 notified excavations. PHMSA, in particular, maintains public his-

[1]Model uncertainty can be addressed through sensitivity analyses to assess the impact of alternate but credible assumptions [171]. In this work, BaNTERA's sensitivity was analyzed by assuming a local X% spread on each parameter of the model; that is, each CPT entry was increased and decreased on X% and the overall reachable range of the outcome of interest, recorded. Both 10% and 20% spreads were considered in this work to account for possible parameter variability not present in the data used to parameterize BaNTERA. For information on the sensitivity analysis algorithm, refer to [172].

	TPDs per 1,000 Notified Excavations
Source A average (2017-2020)	1.410
Source D average (2016-2018)	1.620
BaNTERA (third-party puncture damage)	$\boldsymbol{E = 1.531}$ **95% CIs = [1.490, 1.567]** **10% spread: [0.74, 2.36]**

Table 4.2: Model verification comparing BaNTERA's third-party puncture damage rates per 1,000 notified excavations to the average damage rates of Sources A and D. BaNTERA's results are shown in terms of: expected value (E), 95% equal-tailed credible intervals, and local sensitivity analysis results over a 10% parameter spread.

torical records on these rates for all U.S. states distribution lines [173]. Even though these nationwide statistics are not limited to TPD or puncture damage as BaNTERA is, they provide a relevant source of information for model validation. Therefore, BaNTERA's third-party puncture damage rate is compared to PHMSA's 2020 nationwide average, and region-level minimum, and maximum, excavation damage rates per 1,000 notified excavations in distribution lines. This comparison can be found in Table 4.3. BaNTERA predicts an expected puncture damage rate of 1.540; however, since puncture is less likely to occur than other excavation damages (such as dents and gouges), this prediction should be considered as a lower bound for overall excavation damage rate statistics. On the other hand, since pipe hit rates involve both damaged and undamaged pipelines, BaNTERA's expected pipe hit probabilities are used as an upper bound of comparison with PHMSA's benchmark overall damage rate. Furthermore, in order to take into account the statistical discrepancies between BaNTERA and PHMSA damage rate formulation (i.e., BaNTERA will consider third-party puncture and pipe hit probability as damages, whereas

PHMSA considers overall excavation damages), local sensitivity analysis was performed. Table 4.3 shows this analysis's results as a minimum and maximum bounds on BaNTERA's outputs considering a 10% and 20% parameter spread.

	Damage Rate of Distribution Lines per 1,000 Notified Excavations
PHMSA's nationwide 2020 data (general excav. damage)	$E = 2.712$ min, max = [1.980, 3.556]
BaNTERA Third-Party Puncture Damage	$\boldsymbol{E} = \mathbf{1.540}$ **95% CIs = [1.508, 1.570]** **10% spread: [0.744, 2.366]** **20% spread: [9.825e-3, 3.178]**
BaNTERA Third-Party Pipe Hit	$\boldsymbol{E} = \mathbf{2.518}$ **95% CIs = [2.461, 2.563]** **10% spread: [1.207, 3.876]** **20% spread: [0.000, 5.210]**

Table 4.3: Model validation comparing BaNTERA's third-party puncture damage rates of distribution lines per 1,000 notified excavations to PHMSA's 2020 excavation damage rate. PHMSA's rates are shown in terms of an overall nationwide average (E) and its region-level minimum and maximum values. BaNTERA's results are shown in terms of: Expected value (E), 95% equal-tailed credible intervals, and local sensitivity analysis results over a 10% and 20% parameter spread.

4.8 Case Studies

Two case studies are provided in this work to illustrate BaNTERA's capabilities as a holistic decision-support tool for addressing the third-party excavation damage problem. The studies illustrate different types of inference that can be used to answer excavation process-related queries of interest to various stakeholders. The first case study shows how a natural gas utility operator could estimate the risk of an 811 One Call ticket by using BaNTERA's predictive capabilities, while the second study shows how regulatory agencies could utilize BaNTERA's diagnostic and explanatory capabilities for policy risk-informed decision support.

4.8.1 Predictive Capabilities of BaNTERA: Estimating the Risk of an 811 One Call Ticket

One of the safety barriers against TPD in an excavation is the notification process. Excavators notify an intent to excavate through the 811 One Call system and provide information on different aspects of the excavation activity. This information is processed in the form of a notification ticket, which is provided to utility operators. A notification ticket provides valuable information about an expected excavation activity, including location, and excavator type, which can be used as predictors of potential locating and marking errors and improper excavation practices. As such, this type of information is highly relevant to utility operators, as they would like to predict the risk of pipe damage and identify whether further safety actions should be performed in potentially high-risk excavations.

In this case study, information from a hypothetical 811 One Call ticket for a fencing job is used to predict the probability of observing TPD if we observe the information present in the ticket. The ticket indicates that homeowners are notifying their intent to excavate in a suburban house lot located in College Park, Maryland. The ticket also highlights the presence of a gas pipeline permanent mark on the site and provides information on the specific excavation equipment that will be used.

This case study consists of two similar scenarios related to the hypothetical excavation scene described above. As shown in Table 4.4, the ticket information for both scenarios is assigned as evidence for different node states within the model. These scenarios were each simulated 10^6 times through BaNTERA using the GeNIe software and PySMILE wrapper to determine the expected likelihood of pipe damage due to puncture failure; the resulting estimated pipeline hit and damage probabilities for the simulated excavations are shown in Table 4.5.

Scenario	**Node Name**	**Evidence**
	EAC1: Excavator	General_public
	EAC2: Awareness	Aware
	R1: Notification	Yes
1,2	E1: Regulatory District	South_atlantic
	E2: Area of Incident	Under_soil
	E3: Area Development	Class_3_location
	FC2: Pipe System	Distribution
	EC5: Adjacent to Pipe	Yes
1	EC1: Excav. Equip.	Hand_tools
2	EC1: Excav. Equip.	Excavator_backhoe_trackhoe

Table 4.4: Shared and distinct evidence used in both scenarios for estimating the risk of an 811 One Call ticket.

- *Scenario 1.* The 811 One Call ticket information on the fence work specifies the use of a low-impact hand tool like a shovel. In this scenario, BaNTERA estimates that the expected probability of a pipe hit is 0.673%, 2.16 times more likely than in a general excavation (0.311% according to Table 4.1). Despite this large increase, the expected probability of damage is estimated to be just 0.275%, only 1.08 times more likely than is expected in a general excavation (0.254% according to Table 4.1).
- *Scenario 2.* The 811 One Call ticket information on the fence work specifies the use of a high-impact mechanized equipment, such as a track hoe. In this scenario, BaNTERA estimates that the expected probability of a pipe hit is

similar to the one obtained in Scenario 1 (0.672% for the former and 0.673% for the latter, according to Table 4.1); however, the use of high impact excavation equipment results in a 0.423% expected probability of damage; 1.67 times more likely than the prior of 0.275% in a general excavation.

Scenario	**Target Nodes**	**Node State**	**Expected Probability**	**95% CI [Lower, Upper]**
1	T2: Pipe Hit	Yes	0.00673	[0.00647, 0.00691]
	T4: Damage (Puncture)	Yes	0.00275	[0.00264, 0.00282]
2	T2: Pipe Hit	Yes	0.00672	[0.00650, 0.00693]
	T4: Damage (Puncture)	Yes	0.00423	[0.00409, 0.00436]

Table 4.5: BaNTERA's results for estimated pipe hit and puncture damage probability for both types of 811 One Call ticket scenarios. The values reported are the expected values based on 10^6 simulated excavations.

These two scenarios illustrate how BaNTERA's values for pipe hit and pipe damage probabilities can vary given differing evidence for other nodes in the model.

4.8.2 Diagnostic and Explanatory Capabilities of BaNTERA: Informing Policy Decisions

The previous case study considered how BaNTERA could evaluate the risk of excavation based on its notification ticket; however, a common issue is the lack of notification prior to an excavation activity [168, 174]. According to the CGA DIRT databases [20, 153], at least a third of reported TPDs had a lack of notification as a root cause of the incident. Both PHMSA and the CGA have identified that further policy actions must be made regarding the notification problem.

This case study shows how BaNTERA can be used to help inform policy decisions, such as the notification problem mentioned above. By identifying the main factors and causal dependencies that lead to TPD given a notification issue, the model can provide insight into potential actions that can reduce potential third-party damage. In order to do this, the following two scenarios were each simulated 10^6 times based on an observed instance of pipeline damage:

- *Scenario 3.* A policymaker is interested in evaluating the probability of observing a lack of notification given damage on a pipeline. Using the related scenario evidence specified in Table 4.6, BaNTERA indicates that 66.13% of damaged pipeline incidents are expected to be observed on an excavation that did not have a notification.
- *Scenario 4.* A policymaker is interested in calculating the probability of observing an excavator that is unaware of the required notification practices damaging a pipeline. Using the results from Scenario 3, further analysis is made into the sources of these excavation incidents. The scenario evidence and simulation results from BaNTERA presented in Table 4.6 show that pipelines damaged in excavations that were not disclosed are expected to be caused 81.32% of the time by unaware excavators.

The results from Scenario 4 highlight a major issue for addressing the TPD notification problem: there are some excavations that were not reported by excavators who were aware of the region's notification practices. The following two scenarios were studied to gain further insight into this issue:

- *Scenario 5.* A policymaker is interested in determining the probability of observing an excavation that was not reported given the presence of a damaged pipeline and an excavator that was aware of required notification practices. The scenario evidence and simulation results from BaNTERA presented in Table 4.6 show that this instance of failing to observe notification procedures is expected to be observed with a 26.73% probability.

- *Scenario 6.* To further narrow down potential risk-informed policy actions, a policymaker is interested in identifying which specific type of excavator (i.e., professional or non-professional) is most likely to be involved in the type of event described in scenario 5. The scenario evidence and BaNTERA's simulation results presented in Table 4.6 indicate that in 68.72% of excavations that resulted in pipeline damage and where the excavator did not follow notification procedures even though he or she was aware of notification practices, the excavator is expected to be a professional excavator.

BaNTERA's results from Scenario 6 corroborate a major issue in pipeline safety identified in both the 2018 and 2019 CGA DIRT reports [20,153], namely, that some professional excavators, who are aware of the necessary notification procedures, are choosing not submit a notification for the intent of excavation to a One Call center. This has been a contingent problem and in 2021 CGA stressed the need for professional excavators to notify a One Call center. CGA showed that, in 2021, 60% of the damages caused by a lack of notification were associated with professional excavators; this statistic further validates the use of BaNTERA.

Scenario	Scenario Set Up: Evidence and Target Node (Bold)	Node State	Expected Probability	95% CI [Lower, Upper]
3	T4: Damage (Puncture)	Yes	-	-
	R1: Notification	**No**	**0.66133**	**[0.66131, 0.66135]**
4	T4: Damage (Puncture)	Yes	-	-
	R1: Notification	No	-	-
	EAC2: Awareness	**Non aware**	**0.81315**	**[0.81315, 0.81315]**
5	T4: Damage (Puncture)	Yes	-	-
	EAC2: Awareness	Aware	-	-
	R1: Notification	**No**	**0.26733**	**[0.26731, 0.26735]**
6	T4: Damage (Puncture)	Yes	-	-
	EAC2: Awareness	Aware	-	-
	R1: Notification	No	-	-
	EAC1: Excavator	**Professional**	**0.68719**	**[0.68719, 0.68720]**

Table 4.6: Scenario setup, node evidence, and results of applying BaNTERA towards risk-informed decision making. The target node's expected probability and 95% equal-tailed credible interval (bold) are based on 10^6 simulated excavations.

4.9 Industry Application: A Cloud GIS Software for Identifying Pipelines at Risk of TPD

After the successful development of BaNTERA we partnered with GTI Energy, AgenaRisk (a software and consulting company on risk-informed decision support. Developers of the AgenaRisk BN modeling software [175]), and a utility company in the U.S. to deploy BaNTERA as a pipeline integrity management software. This software was hosted in a cloud server, and its objective was to identify a utility company's natural gas pipelines that are at risk of TPD given historical incident information. Data from the pipe segments can be propagated as evidence in the BaNTERA model to calculate the probability of a puncture caused by third-party excavation activities, which will be used by our utility partner to prioritize damage prevention works in their pipelines. The input and output data from this software will contain Geographical Information System (GIS) data on the pipe segments.

An application programming interface (API) was developed with AgenaRisk to enable the transfer of information between our utility partner's data and BaNTERA through a cloud server. The API and information transfer process is summarized in Figure 4.6 and is explained as follows:

1. A stakeholder, such as a utility company, performs a POST request to the cloud server with their data on a pipeline system to assess the risk of TPD. The input data is a *.csv* file containing a pipeline system's historical incidents and current characteristics divided into multiple pipe segments of 1 mile in length. Each pipe segment is identified geographically with point coordinates in a World Geodetic System 1984 (WGS84) projection [176].

2. The stakeholder then performs a GET request to the server, which is capable of performing the following three tasks:

    (a) Simulate scenarios by propagating input data as evidence in BaNTERA.

    (b) Update BaNTERA's parameters with the input data.

    (c) Test interventions to the system given a scenario defined with the input data.

3. The API requests access to the BaNTERA model and executes one of the tasks requested in the previous step.

4. The results of the previous step are loaded to the cloud server and then provided as a response to the stakeholder's GET request. The output from this process is a *.csv* file containing the result of the requested task in Step 3 and the corresponding pipe segment and GIS coordinates.

The API was built with Python and JavaScript codes and used the *.json* format and AgenaRisk's cloud platform to perform calculations. The BaNTERA model running on the server is an AgenaRisk version of the model. Additionally, the output of the process shown above can be visualized in any GIS software of preference using a WGS84 projection, such as QGIS [177].

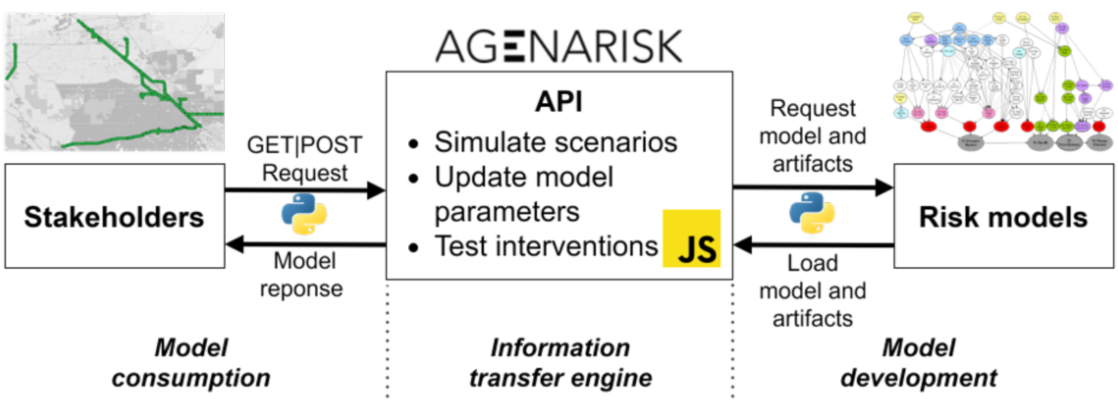


Figure 4.6: API and information transfer process for the TPD integrity management of natural gas pipelines.

A preliminary version of the software was developed in 2022 with the ability to simulate excavation scenarios conditioned on the characteristics of a natural gas pipeline system. This version of the software was tested on 8 miles of a transmission pipeline from our utility partner, and the results are shown in Figure 4.7. As seen in the figure, a multi-modal distribution was found for the probability of punctures per 10,000 excavations. As such, the system's pipe segments could be classified into three TPD risk levels: [0-4), [4-6), and [6-inf] punctures per 10,000 excavations. Furthermore, the risk levels of the pipeline's segments were visualized in GIS software, which provides more granular results on the pipeline's risk of TPD. Currently, it is not possible to provide details on the data and model results. However, an analysis of these results showed that BaNTERA could predict pipe segments that have been historically observed at risk of TPD.

Both the software and BaNTERA are still in development; however, they show high potential in informing the integrity management of our partner's pipeline systems. Currently, the risk of TPD is not assessed by our partner beyond historical failure rates. The problem with using failure rates for a system's pipe segments is that TPD has been demonstrated to be a quasi-random process [167]. Therefore, a model-based approach such as BaNTERA is needed to provide further insights into the causes of a pipe segment failure. The results shown in Figure 4.7 can be used to prioritize prevention strategies and practices in high-risk pipe segments. Furthermore, BaNTERA's causal inference capabilities could be used in the future to identify the major risk drivers for each pipe segment and the effect of implementing specific risk prevention policies by implementing intervention reasoning as shown in Chapter 3.

4.10 Discussion

As a holistic decision support tool, BaNTERA provides a number of benefits for addressing third-party excavation damage in U.S. natural gas pipelines. The first

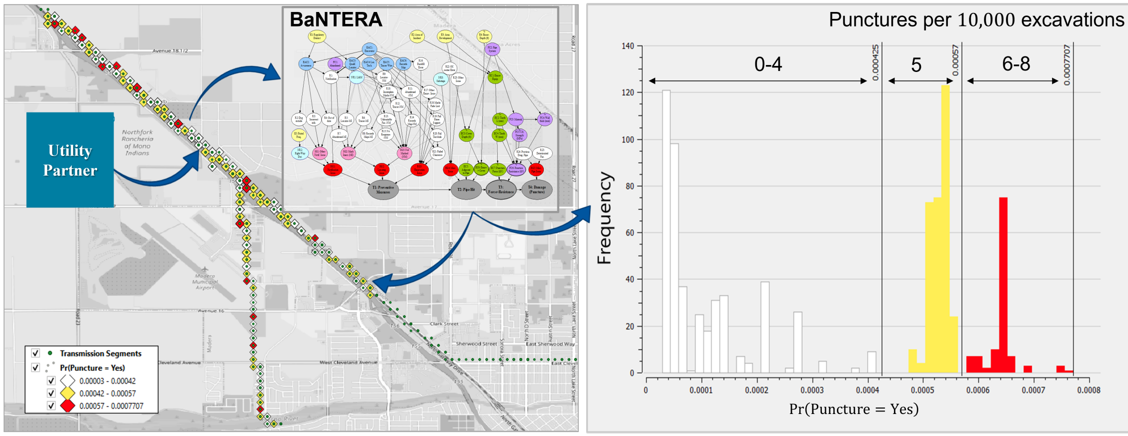


Figure 4.7: Cloud GIS software results for an illustrative natural gas transmission system.

advantage of utilizing BaNTERA is its comprehensiveness with respect to the variety of variables and dependency relationships normally considered to model TPD. This was achieved by integrating the best available data from multiple well-recognized industry and government sources with previous models on different causes of TPD within a BN. Although the CGA DIRT data-collection process was used as the baseline for BaNTERA's construction and data integration, its datasets, which are standardized across the industry, are suitable for incorporating damage prevention data as they are continually updated in terms of definitions, technology, and regulation. In fact, CGA continuously adapts its data-collection practices in order to provide relevant and up-to-date damage prevention insight and recommendations to stakeholders throughout the U.S. Also, it is worth noting that prevention practices are fairly similar across continents [120], making BaNTERA easily adaptable for applications outside the U.S.

Another advantage of BaNTERA is that it is constructed for use over a wide range of third-party excavation contexts and locations; this greatly expands its applicability to excavation sites across the U.S. To the best of the authors' knowledge, this is the first TPD model built with this objective in mind.

As expected in a complex system such as a third-party excavation, the data sources used to parameterize BaNTERA (see Section 4.6.1) do not consist of the same data fields nor have the same levels of granularity. For example, Source C considers twenty-six distinct root causes for TPD, whereas Source B considers only twenty. Additionally, Sources B and C capture nationwide statistics, while Sources A and D provide region-level information. Differences among data sources pose a challenge when they are integrated into a single model and require further assumptions to be made about variable dependencies. Ultimately, this can lead to a model output that may not accurately reflect the TPD rates present in the data used for its parameterization. To verify BaNTERA's performance, a comparison was made between the model's predicted damage rates with those from Sources A and D. BaNTERA's expected third-party puncture damage rate (1.531 damages per 1,000 notified excavations) successfully falls between the values captured in Sources A and D (1.410 and 1.620, respectively); however, the variability introduced by the model's continuous nodes, shown in the 95% credible intervals ($[1.490, 1.567]$), is not enough to account for the differences in damage rates. Performing a local sensitivity analysis by applying a 10% parameter spread accounts for possible parameter variability in BaNTERA. As the resulting damage rate spread from this sensitivity analysis ($[0.74, 2.36]$) includes the average damage rates from Sources A and D, BaNTERA is shown to have the potential of integrating data from disparate sources to provide a comprehensive assessment of third-party excavation damage.

To validate BaNTERA as a trustworthy decision support tool for nationwide use, model outcomes were compared to industry data for outcome consistency; that is, whether if BaNTERA is able to predict the 2020 PHMSA excavation damage rate statistics for natural gas distribution lines. The results from this analysis indicate that BaNTERA's predicted puncture damage rate (1.540 damages per 1,000 notified excavations) is approximately half of what PHMSA had reported (2.712). A lower value in damage rates is expected from BaNTERA as puncture damage is the least likely damage type to result from an excavation work; other damage types, such as dents, gouges, and coating damage, require less force to be inflicted on a pipe [122, 163]. As such, a more appropriate comparison would be between BaNTERA's outcome and PHMSA's minimum distribution line damage rate (1.980), as puncture damage can serve as a lower bound for damage rates.

Although those two values are more comparable, it is not possible to objectively determine whether the difference between BaNTERA's outcome and PHMSA's minimum distribution line damage rate is statistically significant; as such, local sensitivity analyses were conducted by applying both a 10% and 20% parameter spread to account for possible parameter variability in BaNTERA. The results from these analyses indicate that BaNTERA's damage rate range from a 10% parameter spread ($[0.744, 2.366]$) includes PHMSA's minimum value, while a 20% parameter spread ($[9.825\text{E} - 3, 3.178]$) is needed to generate a large enough damage rate range to include PHMSA's 2020 average.

The result presented above shows that, under the current assumptions of the model, BaNTERA is unable to generate a spread of damage rate values that includes the maximum distribution line damage rate value reported by PHMSA (3.556 damages per 1,000 notified excavations). This is because puncture damage is, at most, a lower bound for damage rates; there are other potential sources of damage. A similar validation analysis exercise can be performed, however, using the probability of a pipe hit as an upper bound. By assuming that every hit results in pipe damage, more likely damage types, such as dents and gouges, are then included in BaNTERA's output damage rates. Using this assumption, BaNTERA predicts a TPD rate (2.518) much closer to PHMSA's general excavator's average damage rate. However, it still does not capture PHMSA's maximum damage rate value. A local sensitivity analysis was performed by applying a 10% parameter spread to account for possible parameter variability. BaNTERA's resulting spread of damage rate predictions ($[1.207, 3.876]$) includes all of PHMSA's 2020 industry damage rate values, showing that the model has a high potential for becoming a trustworthy decision support tool for addressing damage to natural gas pipelines in third-party excavations.

The final advantage of using BaNTERA lies in its associative-based predictive, diagnostic, and explanatory capabilities for risk-informed decision support. The different case study results presented in Section 4.8 demonstrate how different stakeholders can use BaNTERA to answer specific TPD-related queries of interest over a number of distinct scenarios. Utility managers interested in maintaining facility operations and asset safety can use BaNTERA to predict the risk of damage to their pipelines within specific excavations or pipe segments. Regulatory organizations and mutual benefit associations, such as PHMSA and the CGA, respectively, can take advantage of BaNTERA's probabilistic inference capabilities to support policy decisions and recommendations by diagnosing the most likely causes of TPD and by explaining under which excavation conditions they are more likely to occur. By addressing the interests of multiple parties, BaNTERA provides a single decision support platform that is not only usable for quantifying TPD risk and identifying how to reduce it, but also enhances information and knowledge transfer between natural gas pipeline stakeholders.

4.10.1 Improvement Opportunities

BaNTERA is built as a proof-of-concept model for a comprehensive probabilistic third-party excavation decision support tool; as such, there are a number of opportunities to improve the model structure, parameters, and functionality.

The first area of improvement is addressing the necessary complexity of the model. Excavation scenarios are significantly complex to model and can vary between different regions and companies. As BaNTERA was designed for a broad range of potential third-party excavation scenarios, some nodes are relatively broad in scope. Nevertheless, BaNTERA still serves as a base for more specific excavation problems and applications, as BNs nodes can be tuned for specific data, and its structure expanded or condensed as needed.

A current limitation of BaNTERA is that the only pipeline damage type represented in the model is puncture damage. Although BaNTERA's damage rate predictions based on puncture are a good approximation of damage rates based on all types of pipeline damage, it should realistically serve as a lower bound value for damage rate. If a pipe is hit in an excavation, it is likely that coating damage, dents, and gouges would occur rather than a puncture. Pipe hit rates predicted by BaNTERA were shown to serve as a good upper bound approximation for damage rates; however, to thoroughly represent TPD, further damage-type models should be incorporated into the model.

Lastly, a recurring challenge throughout any TPD model, including BaNTERA, is the limited availability of success space data [119]; that is, information about excavations that did not result in damage. This appears in a number of forms; ticket datasets typically do not specify the party responsible in the ticket, and utilities are not collecting granular information on excavations without incidents. On top of this, a large number of excavation activities and damages are not reported [20,178]. Collectively, these data limitations are a detriment to accurately modeling TPD. While the research and industry communities are addressing this issue by combining damage data with expert judgment [101,119], limited data sources and data availability can nevertheless restrict the usefulness of TPD models. This can be seen in BaNTERA's node "T1: Preventive Measures." Although the node represents both sufficient and insufficient damage prevention practices in an excavation activity, its parent nodes are parameterized primarily with historical damage data. Additional expert judgment was needed to construct this node's CPT outside of a "failed excavation" space in which damage has occurred to one that also includes excavations without TPD. The limited availability of success space data makes TPD models highly sensitive to the expert judgment used in parameterizing the variables that represent both damage and success spaces. Further efforts should be made by both the industry and regulatory agencies to obtain third-party-specific information on successful excavation activities. By doing so, TPD models such as BaNTERA can incorporate this data and become more robust for real-world applications.

4.11 Concluding Remarks

This chapter presented BaNTERA, a comprehensive probabilistic model for third-party excavation risk assessment of natural gas pipeline hits and subsequent damage. The causal BN structure inherent in the model provides opportunities for decision-makers to gain insight into third-party excavation practices and outcomes at varying granularity levels using different causal reasoning methods. The results from BaNTERA's parameterization and application to a range of case study scenarios indicate a successful and validated integration of historical data from multiple data sources, modern guidelines and root causes, expert judgment-derived causal relationships, and previous TPD models. This distinguishes BaNTERA from previous models, which lack a comprehensive set of up-to-date risk-influencing factors, or their dependencies, needed for assessing third-party damage risk. Although it is the first step in providing a more comprehensive model for TPD risk assessments, the research and methodology behind developing BaNTERA, as well as its preliminary results, are promising for expanding the capabilities of third-party excavation management and providing more effective policies aimed at improving pipeline safety. The contributions of this work are:

- Developed BaNTERA, an up-to-date and comprehensive Bayesian Network for Third-party Excavation Risk Assessment
- Developed a cloud GIS software application for assessing the risk of TPD to transmission pipelines from a utility company in the U.S.
- Informed, using BaNTERA, the risk management of TPDs to natural gas pipelines in the U.S.

In the next chapter of this dissertation, BaNTERA will be used to demonstrate the methods established for enabling counterfactual reasoning in modern QRA.

Chapter 5: Exploiting the Causal Reasoning Capabilities of Bayesian Networks for the QRA of Engineering Systems: Counterfactual Reasoning

5.1 Introduction

Answering "What could have happened?" is a crucial question in risk analysis. As shown in Chapter 2, counterfactual reasoning has been key for generating causal insights on a "learning from events" strategy to risk management [28]. In particular, "but-for"-based counterfactual analysis has been the baseline of several methods for accident and incident investigation [29]. However, as shown in Section 2.2.3, current counterfactual reasoning methods have been criticized for their *linearity*, *incompleteness*, and for being *uncomparable*.

Although we agree that the aforementioned limitations of counterfactuals are well justified, we argue that these criticisms do not apply to counterfactual reasoning *per se*, but specifically for their use on qualitative "but-for"-based accident and incident investigations. The main issue with the "but-for" approach to counterfactual reasoning in system safety is that it is typically used in terms of learning through *attribution*, that is, in trying to infer what caused an event's outcome [179]. Instead, we hold that counterfactual reasoning should be used to enable investigators to learn from a single event by generating and analyzing counterfactual scenarios that are motivated by their hypotheses on why the event happened the way it did. In order to do this, counterfactual reasoning must explicitly acknowledge the stochastic nature of history; that is, a past event is only one of the many possible realizations that could have unfolded. As such, we hold that counterfactual hypotheses can only be rigorously studied through the use of modern causal QRA techniques since they provide a comprehensive characterization of the risks of a system, its associated uncertainties, and different scenarios and events evaluations that can be studied through probabilistic inference methods [30].

This chapter presents the results of RT2.2, which was performed to modernize counterfactual reasoning in QRA and move it up to the third and last rung of the ladder of causality. To do this, this work establishes a new perspective on counterfactual reasoning for system safety by going beyond "but-for" investigations toward a probabilistic "possible worlds" approach. This approach aims to formally integrate the knowledge from a risk assessment model of a system with the best available evidence on a past event to simulate counterfactual scenarios of the event quantitatively. By doing so, "possible worlds" counterfactual reasoning enhances a *"learning from events"* strategy for a system safety approach to risk management by enabling investigators to assess their current qualitative counterfactual hypotheses on a past event through a rigorous quantitative evaluation.

This work uses the BaNTERA model developed in the previous RT2.1 (see Chapter 4) to demonstrate the proposed "possible worlds" approach to counterfactual reasoning through a case study on the 2018 Sun Prairie gas explosion in the U.S. [33], showing how “possible worlds” counterfactuals can be used to inform safety recommendations that overcome the limitations of current counterfactual analyses.

The rest of this chapter is structured as follows: Section 5.2 presents background on counterfactuals and the relevant methods used in this work. Section 5.3 describes the proposed methodology for enabling a “possible world” approach to counterfactual reasoning in system safety. This methodology is applied to a real-world case study on the Sun Prairie gas explosion in Section 5.4. Section 5.5 discusses the implications of “possible world” counterfactuals in system safety. This work ends with concluding remarks in Section 5.6.

The results of RT2.2 were peer-reviewed and published in the journal *Reliability Engineering and System Safety* [180].

5.2 Relevant Methods

This section presents the relevant definitions and methods needed to develop a methodology to perform “possible worlds” counterfactual reasoning in system safety. First, we provide a review of the constitutive elements of counterfactual hypotheses and queries that will be used throughout this work. Second, causal BNs models and associated intervention reasoning methods, the backbones that support counterfactual reasoning, are presented. Last, we explain the graphical twin network approach to counterfactual modeling with BNs, which enables the probabilistic assessment of counterfactual hypotheses and queries.

5.2.1 Counterfactuals

Take an event V_y, in which $V_y = v_y$ occurred. A counterfactual hypothesis looks for an answer to queries of the type *"Could the consequent state of V_y be v'_y in an event, instead of the observed v_y, had the antecedent state of V_x been v'_x, instead of the observed v_x?"* Two "worlds" can be identified in this query. First, a *"factual world"* that represents what actually happened in the event, which is composed of the states of the antecedent $V_x = v_x$ and consequent $V_y = v_y$. Second, a *"possible world"* (also called *counterfactual world*) that indicates what could have happened to the "factual world" consequent, namely $V_y = v'_y$ instead of $V_y = v_y$, had the antecedent been different, namely $V_x = v'_x$ instead of $V_x = v_x$. The factual and possible worlds of a counterfactual query share a common set of background conditions; that is, the variables in the event that are assumed to be causally independent of the factual and possible worlds [83, 181]. This notation will be used for the rest of this work.

Counterfactual reasoning in system safety has traditionally relied on a "but-for" approach. This approach is used to identify the causes of an event by testing the validity of a counterfactual hypothesis of the form: *but-for the antecedent $V_x = v_x$, the consequent $V_y = v_y$ would have been $V_y = v'_y$ in a "possible world".* For technical and hardware causes, logic is often used to test the validity of the "but-for" statement (e.g., informing it by Fault and Event Tree models of a system), while human-related causes are often assessed through expert judgment [29]. An example of "but-for" counterfactuals can be found in many applications, such as in the use of the Swiss cheese model of accident causation, which holds that *but-for* a failure of a safety barrier, an event would not have happened.

The problem with using logic or expert judgment to assess the validity of a "but-for" statement is that it requires a thoroughly understood causal mechanism to support that the consequent state $V_y = v'_y$ would be true in an event had the antecedent been $V_x = v'_x$ in a "possible world." A probabilistic approach is required to address this issue in counterfactual reasoning. We hold that it is valid to state that the consequent $V_y = v'_y$ is a *possible* counterfactual outcome of an event had the antecedent been $V_x = v'_x$ in a "possible world" [182]. Therefore, probabilistic inference methods must be used to assess a counterfactual hypothesis's probability, rather than its certainty. Assessing uncertainty is key for evaluating counterfactuals [30].

To ensure the feasibility of a possible counterfactual outcome, it is necessary to assess its compliance with the constraints entailed by the causal model that characterizes the system in which the event of interest happened. As such, it is crucial that counterfactuals are grounded on the current knowledge of this system and its risks. Therefore, a QRA model serves as an ideal framework for studying counterfactuals. A QRA model is built towards the thorough characterization of the scenarios, consequences, and uncertainties that are involved in a system [14,46], with cause-and-effect relationships as the basis of the modeling methods that have been used for this purpose (e.g., Fault Trees, Event Trees, and currently, BNs) [16, 46]. Fault and Event Trees have been used in the past to support counterfactual claims and hypotheses on the causes of past events and include an assessment of uncertainty. However, they have been criticized and drawn away from modern system safety-based risk management due to their oversimplified modeling of the causality of an event [27, 62].

The definitions and assumptions presented above are the basis of what we propose as a probabilistic "possible world" approach to counterfactual reasoning in risk analysis. To enable this approach, the following methods are used.

5.2.2 Causal Bayesian Networks

Causal BNs are selected to enable the probabilistic "possible world" approach to counterfactual reasoning in risk analysis. Causal BNs for modern QRA have become widely popular in both the system safety [24, 183, 184] and risk assessment [9, 84] literature, which stems from their ability to integrate multiple sources of information and probabilistically model a complex system's risk-influencing factors and causal dependencies. In addition, as shown in previous chapters, BNs can perform probabilistic reasoning under uncertainty [16, 18, 146], making them an ideal framework for capturing an event's complexity and reason about its potential counterfactual realizations.

The first two rungs of the ladder of causality, namely, associative and intervention reasoning, will be used to establish the mathematical background and methods to enable counterfactual reasoning in risk analysis. Associative reasoning will be used through the factorization formula (see Equation 2.3) to incorporate evidence obtained after the investigation of a particular system event (e.g., an incident or accident) into a BN model of a system or process. The posterior probability distribution of the BN model will simulate the background conditions of the event of interest. Then, the intervention reasoning methods established in Chapter 3 (specifically, the *do*-operator and the truncated factorization formula in Equation 3.2) will allow the simulation of the "possible world" in a counterfactual model by forcing the antecedent variables to have a different state in a "possible world" from the ones observed in the "factual world." The next section shows how rungs 1 and 2 of the ladder of causality are incorporated in a BN model to enable the probabilistic "possible world" approach to counterfactuals in risk analysis.

As shown in the next section, interventions allow the simulation of the "possible world" in a counterfactual model by forcing the antecedent variables to have a different state in a "possible world" from the ones observed in the "factual world."

5.2.3 The Twin Network Representation of Counterfactuals

In order to represent and evaluate counterfactual hypotheses and queries on a system's event, Pearl's twin network graphical approach to counterfactuals [18, 185] will be used in the context of BN models. Twin networks graphically represent Pearl's perspective on counterfactual reasoning to predict the possible world consequent of an event given a hypothetical possible world antecedent. This perspective can be explained in three steps [18]. First, the factual world evidence on an event's antecedent and consequent is used to update the past information on the event's background variables. Second, the course of history is bent to comply with the hypothetical possible world antecedent. Third, the possible world consequent can be predicted based on the new understanding of the past information on the event's background variables and the newly established possible world antecedent.

To illustrate the twin network approach, consider the BN model of Figure 5.1a. The nodes of this network are "V_1: Maintenance," "V_2: Safety barrier," and "V_3: System failure." Consider also a system event in which a system failure and a safety barrier failure were observed (i.e., "V_3: System failure = yes" and "V_2: Safety barrier= fail" in Figure 5.1a). In an event investigation, a counterfactual hypothesis could be "had the safety barrier worked, the failure could have been avoided." This hypothesis looks for an answer to the query "*Could the system's failure have been avoided had the safety barrier worked?*" Figure 5.1b shows a twin network representation of the query. Given that we are interested in the effect that "V_2 = work" could have had on the outcome of "V_3," a copy of these two variables, identified with a hat accent, are added to the model. In Figure 5.1b, the portion of the network that includes the antecedent and consequent $\{V_2, V_3\}$ represents the "factual world," while $\{\hat{V}_2, \hat{V}_3\}$ represents the "possible world." In the "possible world" portion of the network, all incoming edges to the antecedent $\hat{V}_2$ are removed, thereby emulating the action of the *do*-operator. This is done to force the antecedent of the "possible world" to take a different value to the one observed in the "factual world" and "bend the course of history" (i.e., "$\hat{V}_2$ = work" instead of the observed "V_2 = fail"). Last, the variables whose posterior probabilities remain invariant in both worlds, such as V_1 in Figure 5.1a, represent the background conditions of the event. For the rest of this work, a twin network model will be denoted by $M^c_{\hat{V}_x = v'_x}$, where $\hat{V}_x = v'_x$ represents the "possible world" antecedent of interest.

It is important to note that twin network theory has been mainly developed for functional models [18]. Nevertheless, works by Fenton and Neil have successfully explored its use in BNs for health risks applications [16, 186, 187]. Although their research has demonstrated the value of performing counterfactual reasoning with BNs for decision-making support, their work is unrelated to system safety and does not provide the requirements or guidelines needed to construct a causally-sound twin network model from a BN model; two gaps that are tackled in this work.

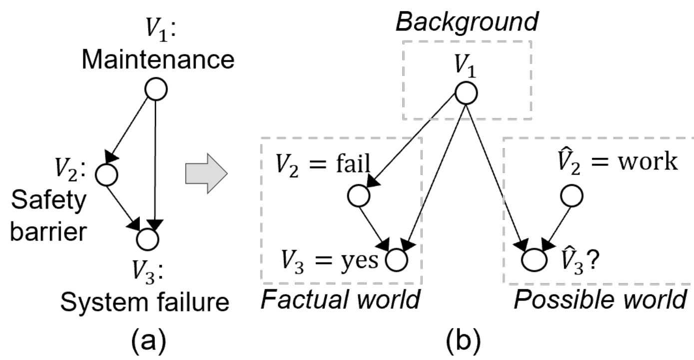


Figure 5.1: Twin network model construction [16]. (a) Original BN model M. (b) Twin network model $M^c_{\hat{V}_3=v'_3}$.

5.3 A Methodology to Enable a "Possible World" Approach to Counterfactual Reasoning in System Safety

To enable a "possible world" approach to counterfactual reasoning in system safety, we propose the methodological framework presented in Figure 5.2. The framework's objective is to assess the likelihood of counterfactual hypotheses on a past event in a system. To do this, the framework uses the best available evidence about an event together with a causal BN-based risk assessment model of the system in which the event occurred. The elements of the framework are described as follows.

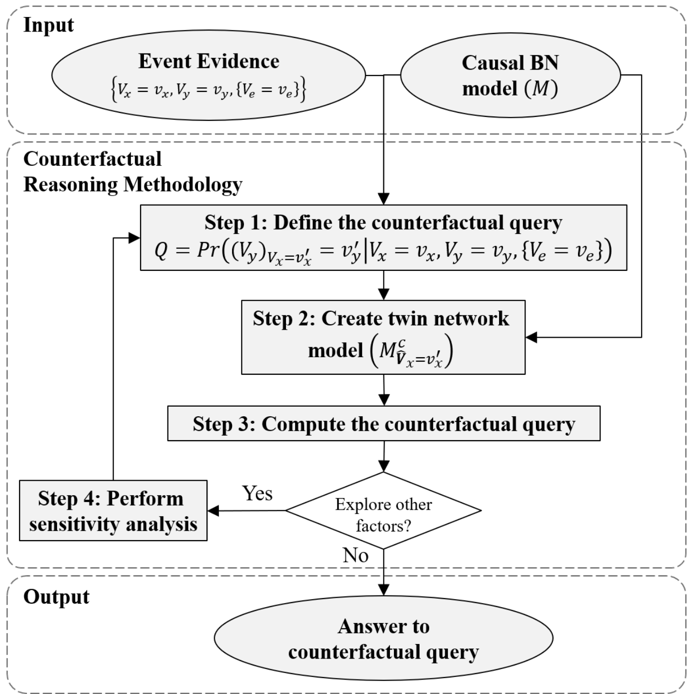


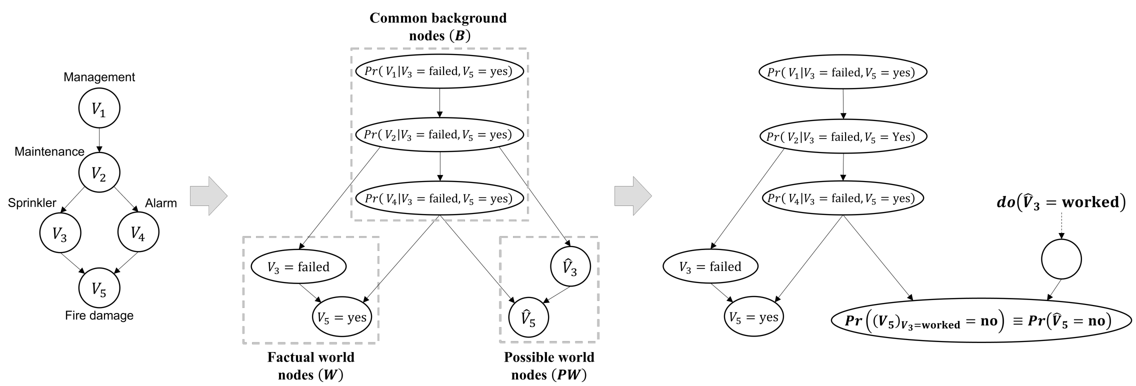
Figure 5.2: Proposed methodological framework for counterfactual reasoning.

5.3.1 Input: Event Evidence and Causal BN Model

A counterfactual hypothesis studies a possible outcome that a system event could have had but did not (i.e., an outcome in a "possible world"). Therefore, in order to assess the likelihood of a counterfactual hypothesis, an analyst only has access to the evidence gathered from the event (i.e., the "factual world") and their knowledge of the system in which the event happened. As explained in Section 5.2, a BN-based risk assessment model of a system is a suitable model for expressing an analyst's expert knowledge of a system and performing reasoning under uncertainty when evidence becomes available. As such, the input to the counterfactual reasoning methodology of Figure 5.2 consists of the tuple:

$$\text{Input} = \langle M, \{V_x = v_x, V_y = v_y, \{V_e = v_e\}\}\rangle \tag{5.1}$$

where M corresponds to a system's BN-based risk assessment model and $\{V_x = v_x, V_y = v_y, \{V_e = v_e\}\}$ corresponds to the set of evidence gathered from a past event in terms of the nodes in M. The specific elements of the evidence set $\{V_x = v_x, V_y = v_y, \{V_e = v_e\}\}$ are explained in Section 5.3.2.1.



(a) System BN model - M (b) Pre-intervention twin network model - M^c (c) Intervened twin network model - $M^c_{\hat{V}_3=\text{worked}}$

Figure 5.3: Simple illustrative example of a fire event (a) BN, (b) pre-intervention twin network, and (c) twin network model for the fire event example. In this example, the following counterfactual query is solved: "Could the fire damage state be "No," instead of the observed "Yes," had the sprinkler state been "worked," instead of the observed "failed," in the event?"

5.3.2 Counterfactual Reasoning Methodology

Using the input tuple in Equation 5.1, we propose the following four-step methodology to assess the likelihood of counterfactual hypotheses in a system safety context. It is important to note that the explanations presented below are written to answer a counterfactual query regarding a single antecedent variable V_x. This methodology, however, also holds for a query on a set $\{V_x\}$ of antecedent variables.

To exemplify the steps of the proposed counterfactual reasoning methodology, consider a hypothetical event in which a fire damaged a facility. In this example, an investigation found that the sprinkler system did not work at the time of the event. In addition, the event investigators have access to a BN-based risk assessment model of the facility's fire protection system, M, as shown in Figure 5.3a. Therefore, the tuple $\langle M, \{V_5 = \text{yes}, V_3 = \text{failed}, \{V_e = \emptyset\}\}\rangle$ will be used as an input to exemplify the proposed methodological steps for counterfactual reasoning.

5.3.2.1 Step 1: Define the Counterfactual Query

First, an analyst transforms their counterfactual hypothesis into a counterfactual query and expresses it in probabilistic terms. The goal is to express the query in a form that can be answered by setting evidence on a system's causal BN model. The evidence is gathered from a past event.

To clearly express a counterfactual's antecedent and consequent in a query, this must be structured as: *"Given evidence $\{V_e = v_e\}$, could the consequent state of V_y be v'_y, instead of the observed v_y, had the antecedent state of V_X been v'_x, instead of the observed v_x, in the event?"*

In this query, the "factual world" antecedent and consequent variables and states of interest are $V_x = v_x$ and $V_y = v_y$, respectively. Likewise, the "possible world" antecedent and consequent are $V_x = v'_x$ and $V_y = v'_y$, respectively. In addition, all evidence on the "factual world" regarding another set of variables relevant to the event but different from V_x and V_y are expressed as $\{V_e = v_e\}$. Therefore, a counterfactual query can be translated into a probabilistic expression as:

$$Q = Pr\left((V_y)_{V_x=v'_x} = v'_y \,\middle|\, V_x = v_x, V_y = v_y, \{V_e = v_e\}\right) \tag{5.2}$$

where Q is the query of interest and $(V_y)_{V_x=v'_x} = v'_y$ is the possible outcome state of the consequent V_y had $V_x = v'_x$ in the event.

To illustrate this step through the fire damage example, consider that, after gathering evidence on the event, investigators hypothesized that the fire could have been mitigated had the sprinklers worked. Following Step 1 of the proposed counterfactual reasoning methodology, this hypothesis can be transformed into the counterfactual query: "Could the fire damage state be "No," instead of the observed "Yes," had the sprinkler state been "worked," instead of the observed "failed," in the event?" Then, the input tuple $\langle M, \{V_5 = \text{yes}, V_3 = \text{failed}, \{V_e = \emptyset\}\}\rangle$ can be used with Equation 5.2 to generate the counterfactual query $Q = Pr((V_5)_{V_3=\text{worked}} = \text{no} \,|\, V_3 = \text{failed}, V_5 = \text{yes})$.

5.3.2.2 Step 2: Create a Twin Network Model

The term $(V_y)_{V_x=v'_x} = v'_y$ in Equation 5.2 corresponds to an unobserved counterfactual outcome of a past event. Thus, computing Equation 5.2 through traditional probabilistic methods is not straightforward. This computation is done with a twin network model as shown in Section 5.2.3. Therefore, the input causal BN model of the system, M, is transformed into a twin network model, $M^c_{\hat{V}_x=v'_x}$, of the counterfactual query obtained in Step 1.

The main feature of a twin network model is that it simulates the same background conditions of the "factual world" in an event and allows the transfer of this knowledge to a "possible world" scenario. As explained at the beginning of Section 5.2, this is a key requirement for assessing the likelihood of a counterfactual hypothesis.

To guide the construction of a twin network model $M^c_{\hat{V}_x=v'_x}$ from an initial BN M, the following three node types are defined:

- *Factual world nodes* (W). Corresponds to the nodes in the system's BN model M representing the counterfactual query antecedents $\{V_x\}$, consequents $\{V_y\}$, and all of their direct descendants (i.e., all nodes in M connected by a direct path starting with an outgoing edge from a node in $\{V_x\}$ or $\{V_y\}$). In addition, the "factual world" event evidence on these nodes must be instantiated. If the evidence on the "factual world" antecedent is uncertain or unavailable from an event's investigation, no evidence is instantiated on the node representing it.
- *Common background nodes* (B). Corresponds to all other nodes in M different from those identified as factual world nodes. The "factual world" evidence on these nodes must be instantiated. These nodes represent the common background conditions between the factual and possible worlds in the counterfactual. As such, these nodes will enable the transfer of knowledge gathered in the "factual world" to a counterfactual "possible world."

- *Possible world nodes* (PW). Corresponds to a copy of the nodes in W (and their causal dependencies) without instantiating the event's evidence. These nodes are differentiated from the ones in W by a hat accent (e.g., if $V_x \in W$, then $\hat{V}_x \in PW$).

The node types defined above are sufficient to construct a "pre-intervention twin network" of a counterfactual query, which will be denoted by M^c. This network encodes all the information needed to perform counterfactual reasoning on a past event. However, the "possible world" antecedent has yet to be forced to be different from what it was in the "factual world." In order to do this, the pre-intervention model M^c is intervened to generate the final twin network model $M^c_{\hat{V}_x=v'_x}$. This is done through the *do*-operator by forcing $do(\hat{V}_x = v'_x) \in PW$ in a possible world, instead of $V_x = v_x \in W$ as it was observed in the event. Consequently, estimating the probability of $(V_y)_{V_x=v'_x} = v'_y$ in the counterfactual query of Equation 5.2 is equivalent to estimating the probability of $\hat{V}_y = v'_y$ in the twin network model $M^c_{\hat{V}_x=v'_x}$.

To illustrate this step, consider the input tuple of the fire event example, $\langle M, \{V_5 = \text{yes}, V_3 = \text{failed}, \{V_e = \emptyset\}\}\rangle$, and the corresponding counterfactual query $Q = Pr((V_5)_{V_3=\text{worked}} = \text{no} \,|\, V_3 = \text{failed}, V_5 = \text{yes})$ constructed in Step 1. Analysts can identify the factual world nodes as $\{V_3 = \text{failed}, V_5 = \text{yes}\} \in W$; the common background nodes as $\{V_1, V_2, V_4\} \in B$; and the possible world nodes as $\{\hat{V}_3, \hat{V}_5\} \in PW$. Then, the pre-intervention twin network M^c shown in Figure 5.3b can be constructed. Last, the intervention $do(\hat{V}_3 = \text{worked})$ is applied to evaluate what could have happened if the sprinkler had worked. The intervention eliminates all incoming edges to $\hat{V}_3 \in PW$, thus generating the finalized twin network model $M^c_{\hat{V}_3=\text{worked}}$ shown in Figure 5.3c.

5.3.2.3 Step 3: Compute the Counterfactual Query

As shown in the previous step, a counterfactual's twin network model $M^c_{\hat{V}_x=v'_x}$ is a post-intervention distribution of M^c in Figure 5.3. As such, its joint probability distribution can be expressed through the truncated factorization formula (see Equation 3.2) as:

$$\begin{gathered} Pr\left(V_i \in \{W \cup B\}, \hat{V}_{i\neq x} \in PW \,\middle|\, do(\hat{V}_x = v'_x)\right) = \\ \prod_{i\,|\,\{V_i,\hat{V}_i\}\neq\hat{V}_x} Pr\left(\{V_i,\hat{V}_i\} \,\middle|\, pa\left(\{V_i,\hat{V}_i\}\right)\right)\big|_{\hat{V}_x=v'_x} \end{gathered} \tag{5.3}$$

Then, if the network variables are discrete, Equation 5.3 can be averaged over $\{V_i, \hat{V}_i\} \neq \{\hat{V}_x, \hat{V}_y\}$ to compute the counterfactual query Q of Equation 5.2 as:

$$Q \propto \sum_{i\,|\,\{V_i,\hat{V}_i\}\neq\{\hat{V}_x,\hat{V}_y\}} \left(\prod_{i\,|\,\{V_i,\hat{V}_i\}\neq\hat{V}_x} Pr\left(\{V_i,\hat{V}_i\} \,\middle|\, pa\left(\{V_i,\hat{V}_i\}\right)\right)\big|_E \right) \tag{5.4}$$

where $E = \left\{V_x = v_x, V_y = v_y, \{V_e = v_e\}, \hat{V}_x = v'_x\right\}$, that is, all the event's evidence gathered by an investigator and the "possible world" antecedent of interest. Given that Equation 5.4 can be computed from a single twin network model $M^c_{\hat{V}_x=v'_x}$, BN software (e.g., GeNIe [1]) or programming libraries (e.g., Python's `pgmpy` [188]) can be used for this task. Additionally, Equation 5.4 is extendable to continuous nodes through integration instead of summation. However, its computation becomes difficult, requiring specialized algorithms such as dynamic discretization for hybrid BNs [16].

This step can be illustrated in the fire event example, in which estimating the counterfactual query $Q = Pr((V_5)_{V_3=\text{worked}} = \text{no} \,|\, V_3 = \text{failed}, V_5 = \text{yes})$ is equivalent to estimating the probability of "$\hat{V}_5 = \text{no}$" in the twin network model $M^c_{\hat{V}_3=\text{worked}}$ of Figure 5.3c.

5.3.2.4 Decision: Explore Other Factors?

In this step, an analyst decides if the estimated answer to the counterfactual query is enough to inform their analysis and safety recommendations. This decision is not trivial and should align with the objectives of the analysis. If no further information is needed from counterfactual queries, the analyst proceeds to provide the generated knowledge to risk managers as decision support. An example of this can be illustrated using the fire event example. If the objective of the analysis was only to determine if a working sprinkler could have helped to avoid fire damage, the outcome of Step 3 of the proposed methodology provides enough information. However, a simplistic objective like this is rarely the goal of an event investigation.

To learn the most and inform the best decisions from past events, an analyst is encouraged to include in their analysis objectives the exploration of potential safety improvement opportunities and to expand their knowledge on a past event through the analysis of multiple counterfactual scenarios. This additional objective has the potential to reveal a number of factors that could have had a relevant influence on the outcome of a past event that might have been overlooked at first glance. The importance of this additional objective has been stressed by authors such as Woo et al. [77, 189] and Oughton et al. [82], which have proposed a downward counterfactual searching approach (that is, an analysis on how an event could have had a worse consequence, such as higher failure probabilities or more severe losses) to elucidate the potential of black swans and extreme events based on past hazardous experiences (we refer the reader to [189] for further information on the counterfactual search approach). Leveraging this objective into quantitative system safety, Step 4 provides a sensitivity analysis-based method for exploring additional factors that could have a relevant effect on the outcome of a past event.

5.3.2.5 Step 4: Perform a Sensitivity Analysis

In order to explore other factors that could have had a relevant effect on the outcome of a past event in a counterfactual setting, we propose a "downward" approach to counterfactual reasoning. In particular, we focus on which variables of the system could have increased the probability of an undesired consequence state in a "possible world" scenario. This is done through a BN-based local sensitivity analysis over the "possible world" consequent $\hat{V}_y = v'_y \in PW$. Then, the most influential "possible world" nodes on the consequent's sensitivity are selected to be further analyzed through counterfactual reasoning; that is, going back to Step 1 and analyze how they could have affected the outcome of the event of interest.

This step is illustrated in the fire damage example as follows: after computing the result of the counterfactual query "Could the fire damage state be "No," instead of the observed "Yes," had the sprinkler state been "worked," instead of the observed "failed," in the event?" the analysts found that a working sprinkler would have not significantly reduced the probability of fire damage. Consequently, the analysts decide to explore if other safety barriers, such as the correct functionality of the alarm system and the maintenance schedule, could have further reduced the likelihood of the event. Using the twin network $M^c_{\hat{V}_3=\text{worked}}$ built in step 2, in which $\{V_3, V_5\} \in W$, $\{\hat{V}_3, \hat{V}_5\} \in PW$, and $\{V_1, V_2, V_3\} \in B$, a sensitivity analysis is performed over $\hat{V}_5 = \text{yes}$. Imagine that the sensitivity analysis results show that "V_4 : Alarm" is the most influential factor on the counterfactual outcome. As such, analysts can decide on going back to Step 1 of this methodology to answer the counterfactual $Q = Pr((V_5)_{\{V_3=\text{worked}, V_4=\text{worked}\}} = \text{No} \,|\, V_3 = \text{failed}, V_5 = \text{Yes})$.

5.3.3 Output: Answer to Counterfactual Query

The output of the counterfactual reasoning methodology presented above corresponds to the probability of a possible outcome of a past event on a system. Depending on how the BN model of the system is specified, this result can be in the form of an expected value or a probability distribution. However, we discourage investigators from using these specific values as the sole basis of their safety improvement recommendations. Rather, the two following points must be taken into consideration:

- The power of counterfactual reasoning in system safety lies in assessing the contrasts between an event and other possible realizations of it [190]. Consequently, the output value of a counterfactual query should not be studied on its own but rather in comparison to other possible outcomes of a past event.

- The output value of a counterfactual query should be studied as a complement to the results obtained through traditional *"learning from events"* safety strategies. For instance, "possible worlds" counterfactual reasoning can be used to assess the distance between models and practice. This can be done by evaluating the differences between the answer to counterfactual queries estimated through the proposed methodology of Section 5.3.2 and the narrative and recommendations provided in an accident investigation report [191].

These points will be illustrated in the case study presented in Section 5.4 and further discussed in Section 5.5.

5.4 A Case Study on Excavation Damage of Natural Gas Pipelines

Thousands of TPD incidents are recorded every year in the country [20,22,153], resulting in major property losses, harm, and fatalities. The is a large amount of qualitative information on past TPD incidents in PHMSA's investigation reports [192]. However, these data are overlooked for informing safety recommendations as they commonly come from reports based on historical statistics (e.g., the CGA's

DIRT report [22]). Extended incident narratives and evidence are mainly used in courtrooms for attributing blame in past events, which also results in simplistic safety recommendations. Therefore, probabilistic "possible world" counterfactuals represent an opportunity to integrate these data into risk assessments to learn from past events caused by TPD and inform safety recommendations.

The BaNTERA model developed in Chapter 3 presents a comprehensive causal structure for assessing the risk of TPDs in the U.S. As such, it can be used for implementing the counterfactual reasoning methods established in 5.3. This section presents a case study that uses BaNTERA and a real-world accident report narrative to demonstrate the capabilities of "possible world" counterfactuals for learning from a past event.

5.4.1 The 2018 Sun Prairie Gas Explosion

Figure 5.4: The 2018 Sun Prairie gas explosion. Source: The Daily Reporter [33]

Around 6 p.m. on July 10, 2018, in Sun Prairie, Wisconsin, U.S., a third-party excavator from VC Tech struck a natural gas main while performing drilling activities to install fiber-optic lines for Bear Communications, LLC. Gas was released,

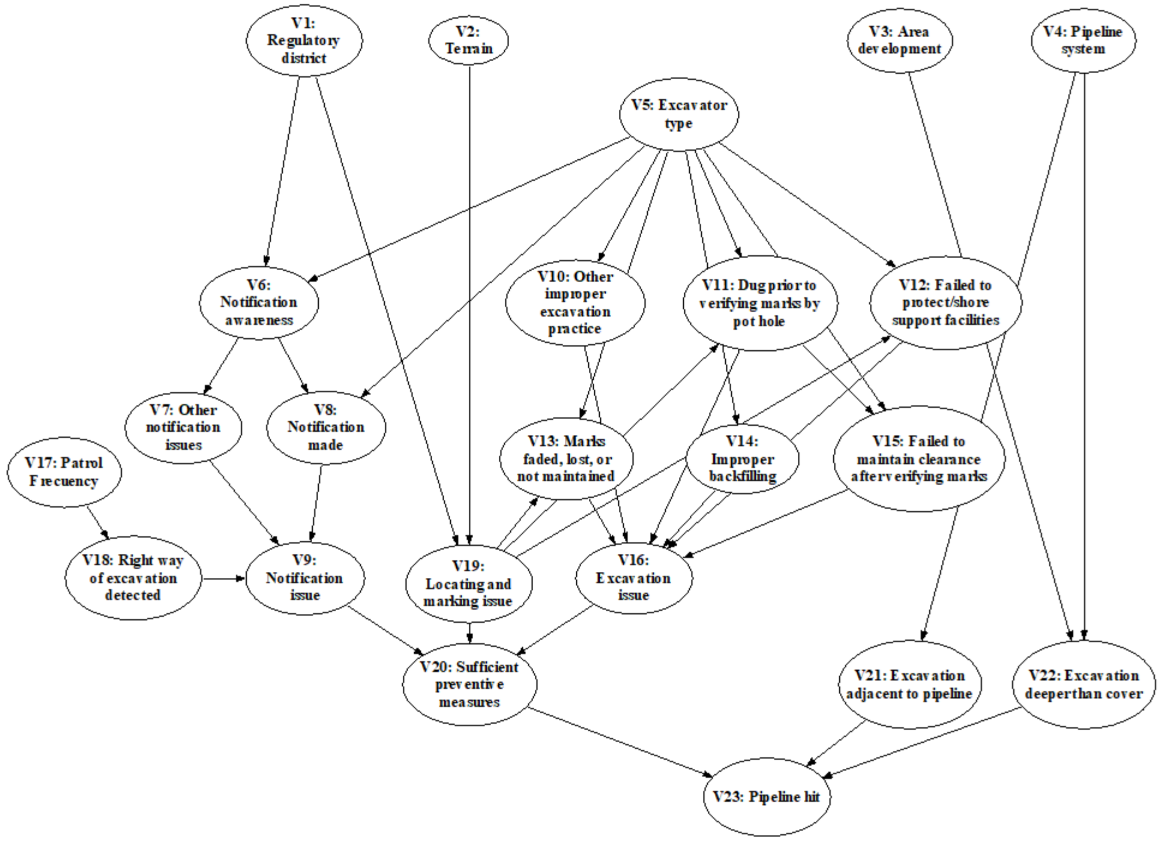


Figure 5.5: Marginalized version of BaNTERA adapted from [23]. This model will be identified as M.

migrating below ground into nearby buildings. One hour later, the gas was ignited by an unknown source, provoking an explosion that resulted in 1 fatality, 8 injures, and approximately \$20M in property damage [192]. A picture of the event is shown in Figure 5.4.

An investigation by the U.S. Department of Labor's Occupational Safety and Health Administration (OSHA) [193] determined, through an administrative trial, that both Bear Communications, LLC and its contractor, VC Tech, were liable for this event. Each company failed to notify an intent to excavate, which is a mandatory practice by Wisconsin's state law. However, the event was more complicated than the judgment entails. The following points summarize the event's story-line presented by OSHA during the administrative trial, which has been widely covered by specialized media [33]:

- Spring of 2018 - Bear Communications, LLC contracted Jet Underground Drilling Company to perform an excavation for fiber-optic cable installations.
- June 2018 - Jet Underground provided a locate request through a notification to Diggers Hotline, Wisconsin's 811 One Call center. USIC Locating Service performed the location and marking of underground utilities on the site.
- July 2018 - Due to timing issues, Bear Communications LLC decided that a competing drilling company, VC Tech, would proceed with the excavation activities. VC Tech had not notified their intent to excavate to an 811 One Call center, relying on previous markings made for Jet Underground in June. While performing drilling operations, VC Tech struck a natural gas main, causing an explosion.
- Public hearings and event's aftermath - In public hearings, VC Tech acknowledged that they were knowingly violating state law by excavating without a locate request notification. However, they blamed USIC Locating Service for the event, arguing that the marks performed for Jet Underground were incorrect. This information could not be verified nor refuted. USIC Locating Service answered that the purpose of the hearing was not to blame them but rather to determine whether VC Tech notified their intent of excavation or not. Moreover, USIC Locating Service argued that, in addition to a lack of notification, VC Tech did not follow dig-in best practices, which could have

avoided the event even though the site was mismarked.

The investigation by OSHA indicates clear liability; the excavators did not comply with state law by failing to notify their intent to excavate. Additionally, a separate investigation by PHMSA determined that the root cause of the event was a lack of notification (see event #20180073 in [192]); the lesson seems clear, *"but-for the lack of notification, the event could have been avoided."* However, the storyline presented above shows that multiple event precursors could have influenced the outcome of the accident. As system safety practitioners, we need to learn from this event as thoroughly as possible to prevent future accidents. To expand OSHA's and PHMSA's investigations on the Sun Prairie gas explosion, we analyzed it through the "possible world" counterfactual reasoning methodological framework presented in Section 5.3.

5.4.1.1 Input: BaNTERA and Accident Narrative

Following the guidelines presented in Section 5.3.1, the following input tuple (see Equation 5.1) is selected for the "possible world" counterfactual reasoning case study on the Sun Prairie gas explosion:

- *Causal BN model (M)*: Since BaNTERA [23] represents a comprehensive causal model for assessing the risk of TPD, a marginalized version of it is used for this case study. This model is presented in Figure 5.5 and will be identified as M. A marginalized version of BaNTERA is selected to clarify the case study in this work. Originally, BaNTERA has 64 nodes and 101 edges, representing a level of complexity that could hinder the presentation of the case study results. The model in Figure 5.5 was marginalized to only 23 nodes and 36 edges. Additionally, it must be highlighted that BaNTERA was marginalized only to model a pipe hit probability, not pipe damage. This was done to focus this study on the excavation process. Additionally, the marginalization of BaNTERA was done carefully to not marginalize any node representing the background conditions of the Sun Prairie gas explosion. We refer the reader to C.1 for a list of the nodes in the model of Figure 5.5, their states, and their prior marginal probabilities.

- *Event evidence*: The evidence on the Sun Prairie gas explosion is obtained from PHMSA's event report #20180073 in [192]. This evidence was mapped to the nodes of the BN model M, resulting in the node states presented in the "Prior" column of Table 5.1. An additional piece of evidence not explicitly shown in this column is that a pipe was hit in the excavation. In terms of the model M, this evidence is equivalent to "$V_{23} = \text{Yes}$."

A value of interest that can be obtained using the information presented above is the prior expected probability of a pipe hit by VC Tech, given the conditions in which the excavation occurred. This value can be calculated through Equation 2.3 using the input model M of Figure 5.5 and the event's evidence presented in Table 5.1. As a result, a prior pipe hit probability of $Pr(V_{23} = \text{Yes}) = 25.75\%$ is obtained for the case study's excavation activity. In frequentist terms, this probability can be interpreted as expecting 25.75 pipe hits in 100 excavations in which the same event evidence is observed. This value will be used for comparison with the outcomes of the counterfactual scenarios analyzed in the next section.

5.4.1.2 Scenario Analysis Through Counterfactual Reasoning

The probabilistic "possible world" counterfactual reasoning methodology presented in Section 5.3.2 is applied to four distinct scenarios regarding the Sun Prairie gas explosion. These scenarios were formulated to thoroughly study the different event precursors presented in the trial storyline shown in Section 5.4.1. Scenarios A and C are used to study the effectiveness that a notification could have had on impacting the probability of a pipe hit by VC Tech. Additionally, scenarios B and D are used to evaluate if there were any opportunities to decrease the likelihood of a pipe hit at the time of the event had the previous marks made by the USIC Locating Service, and used by VC Tech, been incorrect.

- *Scenario A.* First, we will analyze the lack of notification present in the excavation process, which was hypothesized by authorities as the root cause of the Sun Prairie gas explosion. To do this, the following query is studied through probabilistic "possible world" counterfactuals: *Given evidence* $\{V_i = v_i\}$ *gathered from the event, could the consequent state of "*V_{23} *: Pipeline hit" be "No," instead of the observed "Yes," had the antecedent state of "*V_8 *: Notification made" been "Yes," instead of the observed "No," in the event?* Here, and in the subsequent scenario analyses, $\{V_i = v_i\}$ corresponds to all evidence gathered from the event that is different from the antecedent and

Node (Figure 5.5)	Prior	Scenario A	Scenario B	Scenario C	Scenario D
V_1: Regulatory district	East north central	B : East north central	B : East north central	B : East north central	B : East north central
V_2: Terrain	Under pavement	B : Under pavement	B : Under pavement	B : Under pavement	B : Under pavement
V_3: Area development	Class 3	B : Class 3	B : Class 3	B : Class 3	B : Class 3
V_4: Pipeline system	Distribution	B : Distribution	B : Distribution	B : Distribution	B : Distribution
V_5: Excavator type	Professional	B : Professional	B : Professional	B : Professional	B : Professional
V_6: Notification awareness	Aware	B : Aware	B : Aware	B : Aware	B : Aware
V_7: Other notification issues	-	-	-	-	-
V_8: Notification made	No	W**: No;** PW**:** do(**Yes**)	W : No	W**: No;** PW**:** do(**Yes**)	W : No
V_9: Notification issue	-	-	-	-	-
V_{10}: Other improper excavation practice	-	-	-	-	-
V_{11}: Dug prior to verifying marks by pot hole	-	-	-	-	W**: -;** PW**:** do(**No**)
V_{12}: Failed to protect/shore/ support facilities	-	-	-	-	-
V_{13}: Marks faded, lost, or not maintained	-	-	-	-	-
V_{14}: Improper backfilling	-	-	-	-	-
V_{15}: Failed to maintain clearance after verifying marks	-	-	-	-	W**: -;** PW**:** do(**No**)
V_{16}: Excavation issue	-	-	-	-	-
V_{17}: Patrol frequency	-	-	-	-	-
V_{18}: Right way of excavation detected	No	B : No	B : No	B : No	B : No
V_{19}: Locating & marking issue	-	-	W**: -;** PW**:** do(**Yes**)	W**: -;** PW**:** do(**Yes**)	W**: -;** PW**:** do(**Yes**)
V_{20}: Sufficient preventive measures	-	-	-	-	-
V_{21}: Excavation adjacent to pipeline	Yes	B : Yes	B : Yes	B : Yes	B : Yes
V_{22}: Excavation deeper than cover	Yes	B : Yes	B : Yes	B : Yes	B : Yes
V_{23}**: Pipeline hit**	Pr(**No**) $= 0.7425$ Pr(**Yes**) $= 0.2575$	W: Pr(**Yes**) $= 1$; PW: Pr(**No**) $= 0.8961$ Pr(**Yes**) $= 0.1039$	W: Pr(**Yes**) $= 1$; PW: Pr(**No**) $= 0.3197$ Pr(**Yes**) $= 0.6803$	W: Pr(**Yes**) $= 1$; PW: Pr(**No**) $= 0.8815$ Pr(**Yes**) $= 0.1185$	W: Pr(**Yes**) $= 1$; PW: Pr(**No**) $= 0.3925$ Pr(**Yes**) $= 0.6075$

Table 5.1: Summary of the inputs and results of Sun Praire gas explosion case study. The "-" symbol indicates that no information was instantiated in the BN model used in the study. For counterfactual scenarios, W, PW, and B indicate information that was instantiated in the "factual world," "possible world," and background nodes, respectively, of the scenario's twin network. $do(\cdot)$ indicates the use of the do-operator in the scenario's twin network "possible world" nodes.

consequent of interest (V_8 and V_{23} respectively for this scenario). This evidence is presented in Table 5.1 under the "Scenario A" column. Following Step 1 of the methodology, this query can be expressed mathematically as:

$$Q_A = Pr\left((V_{23})_{V_8=\text{Yes}} = \text{No} \,\middle|\, V_8 = \text{No}, V_{23} = \text{Yes}, \{V_i = v_i\}\right) \tag{5.5}$$

Then, following Step 2 of the methodology, the twin network presented in Figure C.2 is created to compute the answer to the query in Equation 5.5. Last, following Step 3 of the methodology, the answer to the counterfactual query is calculated, obtaining $Q_A = 89.61\%$. As such, a notification by VC Tech could have reduced the expected probability of a pipeline hit from 25.75% in the "factual world" to 10.39% in a "possible world" (see Table 5.1).

- *Scenario B.* The result obtained in Scenario A aligns well with the conclusions provided by the OSHA and PHMSA investigations; that is, if VC Tech had provided a notification, the incident could have been less likely to occur. VC Tech, however, argued that the actual root cause of the event was that the prior locating and marking works performed by USIC Locating Service were wrong. In order to assess this hypothesis, the following counterfactual query is studied: *Given evidence* $\{V_i = v_i\}$ *gathered from the event, how likely is that the consequent state of "*V_{23} *: Pipeline hit" been "No," instead of the observed "Yes," had the antecedent state of "*V_{19} *: Locating & marking issue" been "Yes" in the event?* Here, the "factual world" antecedent state on V_{19} is unknown; that is, there was no clear evidence supporting a locating and marking issue at the time of the event. Following Step 1 of the methodology, this query can be expressed mathematically as:

$$Q_B = Pr\left((V_{23})_{V_{19}=\text{Yes}} = \text{No} \,\middle|\, V_{23} = \text{Yes}, \{V_i = v_i\}\right) \tag{5.6}$$

The twin network shown in Figure C.3 is used to compute Equation 5.6, obtaining $Q_B = 31.97\%$. Therefore, the "possible world" counterfactual suggests that if the marks used by VC Tech to perform their excavation activities had been incorrect, there was 68.03% probability of a pipeline hit at the time of the event (see Table 5.1).

- *Scenario C.* Even though the potential negative effect of locating and marking issues were considered in OSHA's investigation, the focus remained on a lack of locate request notification. To assess the effect of a notification on the probability of a pipe hit in the scenario in which there were incorrect locating and marking works at the excavation site, the following counterfactual query is studied: *Given evidence $\{V_i = v_i\}$ gathered from the event, how likely is that the consequent state of "V_{23} : Pipeline hit" been "No," instead of the observed "Yes," had the antecedent state of "V_{19} : Locating & marking issue" been "Yes" and the antecedent state of "V_8 : Notification made" been "Yes," instead of the observed "No," in the event?* Mathematically, this query is equivalent to answering:

$$Q_C = Pr\left((V_{23})_{V_8=\text{Yes},V_{19}=\text{Yes}} = \text{No} \,\middle|\, V_8 = \text{No}, V_{23} = \text{Yes}, \{V_i = v_i\}\right) \tag{5.7}$$

  The twin network shown in Figure C.4 is used to compute Equation 5.7, obtaining $Q_C = 11.85\%$. Therefore, a notification by VC Tech could have reduced the expected probability of a pipeline hit from a 25.75% in the "factual world" to a 11.85% in a "possible world" in which there was a locating and marking issue in USIC Locating Service works at the site (see Table 5.1).

- *Scenario D.* The result obtained for scenario B shows that a pipe hit could have been highly probable (68.03%) if the marks used by VC Tech would have been incorrect. Nevertheless, Scenario C showed that this probability could have still been significantly reduced (11.85%) had VC Tech provided a locate request notification. In public hearings, however, USIC Locating Service claimed that VC Tech did not follow dig-in best practices that could have avoided the event even in the scenario in which a locating and marking issue was present. This claim motivated us to study if there were any dig-in best practices that VC tech could have followed to effectively avoid a pipe hit in this scenario.

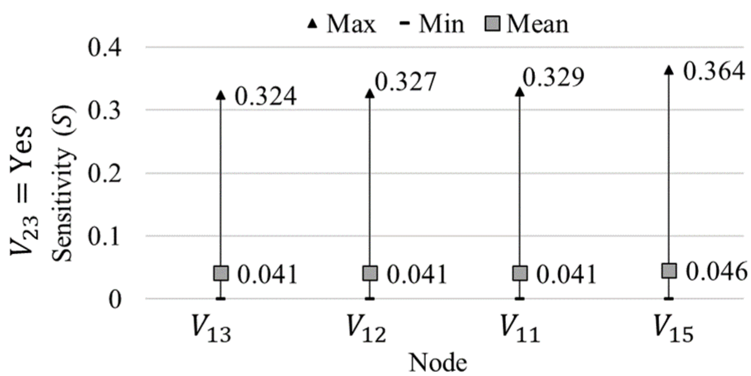


Figure 5.6: Min, max, and mean sensitivity derivative parameter (S) of the nodes V_{11}, V_{12}, V_{13}, and V_{15} on the target node V_{23}.

In order to perform the abovementioned study, Scenario B will be expanded to include the fourth step of the proposed counterfactual reasoning methodology; that is, explore other risk-influencing factors through a sensitivity analysis. The same evidence used in Scenario B is instantiated on the twin network presented in Figure C.3, and a sensitivity analysis is performed over the "possible world" outcomes of the node "V_{23} : Pipeline hit". In particular, this is done by calculating Kjærulff and Van Der Gaag [172] sensitivity derivative parameter, S, for all "possible world" dig-in best practices errors that depend on a marked excavation site (that is, V_{11}, V_{12}, V_{13}, and V_{15} in Table 5.1) [1] The results of the sensitivity analysis are shown in Figure 5.6.

Figure 5.6 shows that, if there had been a marking error at the time of the event, a pipe hit (i.e., "V_{23} = Yes") would have been most sensitive to increasing its probability due to the two following dig-in issues: "V_{11}: Dug prior to verifying marks by pot hole" and "V_{15}: Failed to maintain clearance after verifying marks." Potholing and clearance maintenance are expected to be performed consecutively in any excavation. As such, the following counterfactual query is studied: *Given evidence* $\{V_i = v_i\}$ *gathered from the event, how likely is that the consequent state of "*V_{23} *: Pipeline hit" been "No," instead of the observed "Yes," had the antecedent state of "*V_{11}*: Dug prior to verifying marks by pot hole" been "No," the antecedent state of "*V_{15}*: Failed to maintain clearance after verifying marks" been "No," and the antecedent state of "*V_{19} *: Locating & marking issue" been "Yes," in the event?* Mathematically, this query can be expressed as:

$$Q_D = Pr\left((V_{23})_{V_11=\text{No},V_{15}=\text{No},V_{15}=\text{Yes}} = \text{No} \,\middle|\, V_{23} = \text{Yes}, \{V_i = v_i\}\right) \tag{5.8}$$

The twin network shown in Figure C.5 is used to compute Equation 5.8, obtaining $Q_C = 39.25\%$. Therefore, if VC Tech had excavated using wrong marks, a pipe hit probability could have only been reduced from 68.03% (as

[1]The sensitivity derivative parameter S of a node state "$V_i = v$" given a target node state "$Y_i = y_i$" means that a change c on the value of $Pr(V_i = v_i)$ will cause a change of $S \cdot c$ on the value of $Pr(Y_i = y_i)$. For more information on the method, we refer the reader to [172].

shown in Scenario B) to 60.75% had they performed a pothole and subsequent maintenance of clearance at the excavation site (see Table 5.1).

5.4.1.3 Output: Lessons Learned

Both OSHA's and PHMSA's investigations identified a lack of a locate request notification as the root cause of the event. This claim is supported by the results obtained for scenarios A and C of this study. The "possible world" counterfactual analysis showed that a pipe hit probability could have been reduced from 25.75% to a 10.39% had a notification been provided to authorities (Table 5.1, Scenario A). In addition, this study showed that a pipe hit probability could have increased only to 11.85% had locating and marking works by USIC Locating service on the site had been incorrect (Table 5.1, Scenario C).

The results obtained for scenarios B and D, however, showed that if it was true that VC Tech used incorrect marks to perform the excavation activity, the probability of a pipe hit could have been 68.03% (Table 5.1, Scenario B); 2.25 times more likely that the expected prior of 25.75%. Furthermore, this value could have been reduced only to 60.75% had VC Tech performed the dig-in best practices of potholing, and subsequent site clearance, which were found to be the most influential practices on a pipe hit probability in this scenario (Table 5.1, Scenario D).

Therefore, the "possible world" counterfactual analysis performed on the Sun Prairie gas explosion showed that, in the particular case in which an excavator uses incorrect marks from a previous locating and marking work, only a locate request notification has the potential to significantly reduce a pipe hit probability. However, considering that a lack of notification is the major root cause of third-party damage in the U.S. [21], we conclude that it is necessary to find additional safety barriers beyond current dig-in best practices such as potholing and clearance maintenance to avoid similar events in the future.

5.5 Discussion

Counterfactual reasoning is prevalent in system safety in the form of "but-for" analyses due to the ability to incorporate the evidence from single events into an analyst's knowledge of a system in order to identify an event's causes and inform recommendations [28, 29, 194]. However, researchers have raised three important limitations on "but-for" counterfactuals for learning from events: *linearity*, *incompleteness*, and *uncomparable*. In this work, we created the probabilistic "possible worlds" approach to counterfactuals to tackle these limitations while keeping the benefits of counterfactual reasoning in system safety.

The first limitation on the current use of counterfactuals in system safety is *linearity*, meaning that event investigators are prone to turn complex events into linear cause-and-effect chains. Probabilistic "possible worlds" counterfactuals address this limitation by using a BN-based QRA model to assess the likelihood of counterfactual hypotheses. BNs are built to represent the current state of knowledge on the uncertainties about the phenomena, processes, and activities involved in a system [14]. As such, the use of a BN model enabled us to incorporate into the Sun Prairie gas explosion case study different relevant factors that are not currently considered in dig-in best practices procedures [155] such as notification awareness, excavator type, and a utility company's patrol frequency (see C.1). Furthermore, the use of the BaNTERA QRA model enabled us to better express the causal complexity of an excavation activity by incorporating commonly overlooked causal dependencies, such as the effect that a type of terrain has on the likelihood of locating and marking issues ($V_2 \rightarrow V_{19}$ in Figure 5.5) and the effect that a regulatory district has on an excavator's awareness on notification procedures and the quality of locating and marking works ($V_6 \leftarrow V_1 \rightarrow V_{19}$ in Figure 5.5).

The second limitation on the current use of counterfactuals is *incompleteness*, namely, event investigators tend to use counterfactuals to transform an event's evidence into proof of failure to perform according to a system's procedures and norms, therefore missing underlying reasons of an event. *Incompleteness* was clear in the official investigations following the Sun Prairie gas explosion. Both OSHA [193], and PHMSA [192] concluded, without further analysis, that the event's root cause was a lack of locate request notification. Although there was additional evidence of potential location and excavation errors, the event was reduced to additional evidence of the potential high consequences that not following notification procedures can have on pipeline safety. Drawing a parallel to a "but-for" approach to counterfactual reasoning, the statement *"but-for the lack of notification, the Sun Prairie gas explosion could have been avoided"* was found by authorities as a logical, valid, and sufficient explanation to the event. As such, the investigation was halted, and the only lesson learned was to keep promoting and enforcing compliance with notification procedures (see Section 5.4). However, as we showed in this work, much more could have been learned from this event.

As discussed above, using BN models in probabilistic "possible worlds" counterfactuals allows the inclusion of relevant factors beyond a system's procedures and standards in assessing counterfactual hypotheses. However, this capability does not ensure that these factors will be considered in the analysis. To overcome this issue, the proposed methodological framework for counterfactual reasoning (Figure 5.2) includes a sensitivity analysis step that identifies which variables of a system (different from the previously queried "possible world" antecedents) could have increased the probability of an undesired consequence in a counterfactual scenario. This step is suggested to be performed after assessing an analyst's initial counterfactual hypotheses, thereby expanding the bounds of what can be learned from a past event. This capability was demonstrated in the Sun Prairie gas explosion case study by showing that, even though dig-in best practices had been performed, a pipe hit still could have been highly probable. Considering that there has been a call for action beyond current damage prevention practices in the U.S. [155], a "possible world" counterfactual analysis of the Sun Prairie gas explosion provided relevant evidence into exploring further actions beyond dig-in best practices to prevent damages when all previous safety barriers have failed.

The last limitation of the current use of counterfactuals in system safety is that they are *uncomparable*; that is, there is no rigorous method to compare and prioritize counterfactual hypotheses on event investigations. To tackle this issue, probabilistic "possible worlds" counterfactuals use the QRA principles of uncertainty quantification and scenario analysis.

To incorporate uncertainty quantification in a counterfactual analysis of an event, we acknowledge a past event as one of many possible scenarios that could have unfolded. As such, we propose to assess the likelihood of a counterfactual hypothesis by transforming it into a probabilistic query to be answered through BN-based methods (see Figure 5.2). Quantifying the uncertainty on the realization of a past event through probabilities allows different counterfactual hypotheses to be directly compared. This capability of "possible worlds" counterfactuals was demonstrated, for instance, in the case study when assessing the likelihood of the counterfactual hypothesis "A locate request notification could have avoided a pipe hit." In Table 5.1, we showed that the effect that a locate request notification could have had in a pipe hit probability could have been similar independently if a locate and marking error had been present at the excavation site (10.39% compared to 11.85%).

Although counterfactual hypotheses can be compared through probabilities, their calculation is not straightforward. Traditionally, a risk assessment quantifies the probability of a specific scenario; that is, assessing how a cause can potentially lead to a specific consequence. Counterfactuals, however, are based on the evidence gathered from a past event, namely, a scenario that has already unfolded. As such, it was necessary to establish the mathematical background and methods needed to incorporate an event's evidence into analyzing a counterfactual scenario in a system safety context. As demonstrated in this work's case study, this was done through a BN-based twin network method, which allows a transparent representation of a counterfactual's elements (namely, its factual and possible worlds antecedent, consequent, and background conditions), and the integration of evidence through probabilistic inference methods (see Section 5.2).

Despite the abovementioned benefits of "possible worlds" counterfactuals in system safety, it is important to consider its limitations when informing "lessons learned" and recommendations. Insights that can be learned from a "possible worlds" counterfactual analysis are bounded by the BN-based QRA model used to describe the system and the event of interest. In particular, an accurate causal model of the event is necessary to provide useful information when learning from an event [195]. If a causal model is not accurate, it is possible to overlook potential causal mechanisms or background variables that can affect the correct assessment of the likelihood of a counterfactual hypothesis of interest [18]. Additionally, accurate causal modeling is also important due to the fact that counterfactual hypotheses are not possible to verify; an event already happened the way it did. Although natural experiments have been recommended to verify counterfactuals [29], we hold that the use of "possible world" counterfactuals should not be limited to the existence of these experiments. Rather, we recommend that the analyzed counterfactual hypotheses do not diverge from what is plausible to answer with the used QRA model [18, 186].

Finally, the proposed methodology in Section 5.3 assumes that an event's twin network counterfactual model is solely based on the variables included in the analyzed system's original causal BN model. Although this assumption holds for well-understood systems, for new technologies, highly-complex systems, and events with sparse/conflicting evidence, it may be necessary to add new variables to model a "possible world" event. An example concerning sparse and conflicting evidence can be seen in legal argumentation, which can be a direct ramification of a high-consequence accident such as the Sun Prairie explosion. As described by Neil et al. [196], a model representing two competing narratives of an event (one as a "factual" and the other as a "possible" world) will have commonalities. However, variables such as factual evidence and source credibility can differ in both narratives.

Similar challenges can be seen when performing counterfactual analyses on new technologies. An example of this can be seen when analyzing the transportation of hydrogen blends in natural gas pipelines. Consider the following counterfactual query: *"Given the evidence on the Sun Prairie explosion, could the consequences have been different if the transported gas was a hydrogen blend, instead of natural gas?"* A twin network model of this query will require additional nodes in the "possible world" to incorporate the effect of hydrogen on natural gas pipelines, such as hydrogen embrittlement on steel pipes and potentially higher operational pressures. Addressing these challenges presents an excellent opportunity to expand the insights and lessons learned from an event's investigation in the face of complex systems, events, and new technologies. The proposed "possible worlds" approach can serve as the backbone of these analyses in the future.

5.6 Concluding Remarks

This work established a "possible worlds" approach to counterfactual reasoning for system safety that improves a "*learning from events*" safety strategy. The proposed probabilistic "possible worlds" approach to counterfactuals enables the integration of an analyst's causal knowledge of a system (in the form of a BN-based risk assessment model) with the best available evidence on an event of interest. As a result, counterfactual hypotheses, which are commonly used in the practice of system safety, can now be rigorously assessed through causally-sound probabilistic methods. This approach overcomes current "but-for" counterfactuals' limitations, as was demonstrated in the case study on the 2018 Sun Prairie gas explosion.

Learning from single past events is crucial in the practice of system safety, and the new "possible worlds" approach provides a new foundation for enabling this. In summary, this work's major contributions are:

- The establishment of the "possible world" approach to counterfactual reasoning that enables, for the first time, a probabilistic assessment of counterfactuals in system safety.
- Demonstrating that "possible world" counterfactuals built on a QRA model and the best available evidence on an event provide an objective, defensible, and transparent basis for addressing counterfactual reasoning in system safety.
- Posit the mathematics of "possible worlds" counterfactuals to expand current capabilities of counterfactual methods in QRA and system safety for learning from past events (e.g., incident and accident investigations).
- The demonstration of the probabilistic "possible worlds" approach to counterfactual reasoning with the Sun Prairie explosion case study and the creation of extended safety recommendations compared to past investigations on the event.

The results presented in this chapter enabled us to achieve the second objective of this dissertation and, therefore, expanded modern QRA models based on BNs to provide causal insights on all three rungs of the ladder of causality. The completion of O1 and O2 can greatly impact the type of studies and insights that can be obtained from complex engineering systems and processes. The next chapter of this dissertation will demonstrate the benefits in risk analysis of exploiting the causal reasoning capabilities of modern QRA's with BN models.

Chapter 6: Application: A Causal Assessment of the Risk of Third-Party Excavation Damage to Natural Gas Pipelines Transporting Hydrogen

6.1 Introduction

The wide adoption of hydrogen technologies will require an extensive pipeline network for hydrogen delivery. As shown in Chapter 2, transporting hydrogen through existing natural gas pipelines can be challenging and pose unique risks that must be assessed before deploying this novel fuel. Regarding TPD, hydrogen affects several important properties that can increase the frequency and severity of TPD incidents in both carbon steel and plastic pipelines. Therefore, governments and industries will require in-depth studies and insights into the risks posed by hydrogen transportation through pipelines, which will serve as the baseline for future safety codes and standards that will be key for the wide and safe adoption of hydrogen. As shown throughout this research, the causal reasoning methods for QRA established in this dissertation are built toward generating a comprehensive set of risk insights on engineering systems. The use of these causal reasoning methods can be crucial in supporting the safe transportation of hydrogen through existing pipelines.

This chapter presents the results of RT3, which focuses on assessing the risk of TPD on hydrogen pipelines using the causal reasoning methods established in RT1 and RT2. Chapter 2 showed that past QRAs on hydrogen transportation through existing pipelines focus only on consequences (e.g., societal impact) and simplify the frequency of pipeline incidents to a single expected value equal to those observed in natural gas. However, there is evidence that hydrogen can affect both the frequency and severity of incidents, including those caused by TPD [127, 129, 140]. Reducing the risk of TPD to a single (and unchanged) frequency value can be useful in preliminary risk assessments on hydrogen pipelines. Still, analysts could be overlooking relevant risk-reduction insights on one of the major causes of pipeline incidents in the U.S. and the world [197]. Therefore, this chapter shows how risk assessment insights based on the three rungs of the ladder of causality can be highly valuable to the early deployment of hydrogen technologies, as they can inform recommendations for future actions and policies required to enable the safe transportation of hydrogen through existing natural gas pipelines. Furthermore, this work demonstrates how BNs models can be used as a unified approach to exploit causal reasoning to improve, in general, the QRA of complex engineering systems and processes.

The rest of this chapter is structured as follows: Section 6.2 presents the analysis scope and QRA methodology employed in this research task. Section 6.3 presents a description of the U.S. natural gas pipeline system, a third-party excavation process, and the effect of hydrogen on natural gas pipelines. Section 6.4 presents the methods, information sources, assumptions, and process to model the risk of TPD to natural gas pipelines transporting hydrogen. Section 6.5 shows the resulting risk model for TPD and the total risk evaluation, SHyTERP. Section 6.6 presents the results of four independent studies developed for providing causal insights on the risk of TPD to hydrogen pipelines. Last, Section 6.8 concludes with recommendations for contributing to the safe transportation of hydrogen through natural gas pipelines.

Additionally, the results of this work are anticipated to be submitted as a manuscript to the *International Journal of Hydrogen Energy*.

6.2 Analysis Scope and Risk Assessment Methodology

This work considers underground transmission and distribution pipelines, which can suffer from dents, gauges, and punctures in excavation activities. Punctures are a major cause of high-consequence incidents in the U.S. [198] and can result in jet fires if ignition occurs immediately or explosions if the ignition is delayed. In this work, only jet fires are considered[1]. Hydrogen releases can easily and rapidly ignite, increasing the likelihood of jet fires compared to natural gas releases. Jet fires can pose a major threat to personnel and equipment located near an excavation site and can be the primary step of domino-effect incidents [199]. As such, this work uses a QRA approach to (1) assess the risk of TPD and jet fires to hydrogen transportation through existing natural gas pipelines and (2) provide insights and recommendations

[1]Therefore, the term "ignition" will be equivalent to "immediate ignition" throughout this chapter

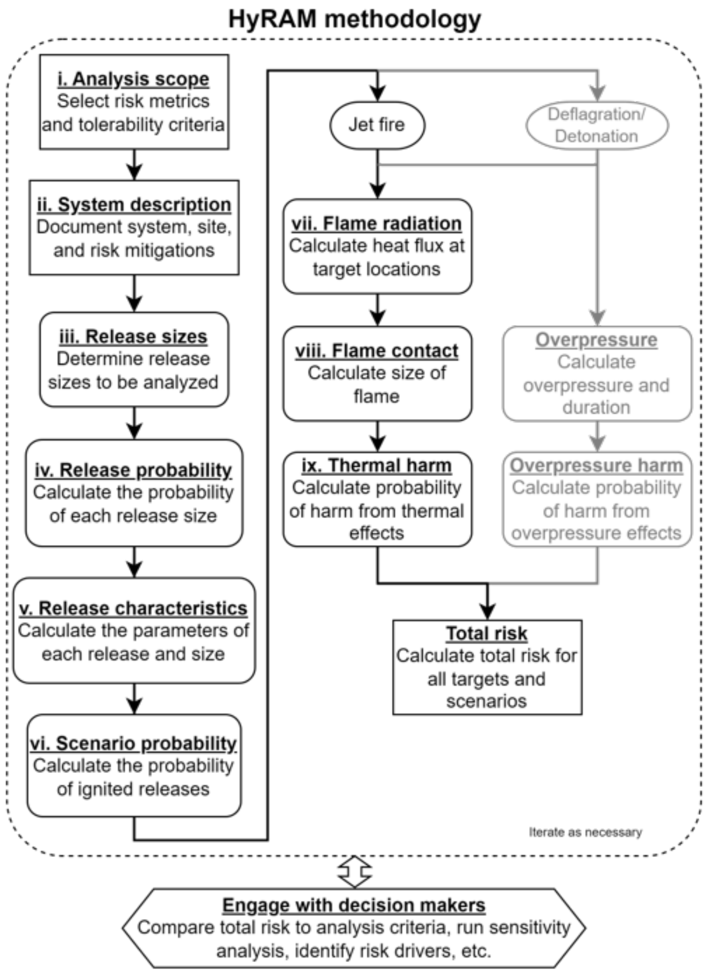


Figure 6.1: HyRAM methodology for the QRA of hydrogen systems. Adapted from [34]. This work only uses the steps involved in the analysis of jet fire scenarios. All other steps are shaded in grey.

for future actions/policies to handle risks.

The QRA performed in this work measures risk by estimating the probability per excavation activity of three undesired consequences caused by TPD to hydrogen pipelines: punctures (equivalent to a gas release), ignited releases, and thermal harm from jet fires. Probability is measured as the likelihood that these consequences occur per excavation activity in the U.S. and is analyzed separately for distribution mains and transmission pipelines.

Although the estimated probability values in this QRA can provide relevant insights as frequency estimations, their accuracy is difficult to assess due to the lack of hydrogen-specific data, the magnitude of the U.S. pipeline system, and the complexity of third-party excavation activities [20, 23]. Furthermore, third-party excavations and damages are a quasi-random process [167] that involves several interacting factors, including the environment in which they are performed and the actors and objects involved in the excavation. As such, this work develops a TPD risk model capable of simulating an excavation process and its potential consequences on both natural gas and hydrogen settings.

The total risk of TPD to hydrogen pipelines is assessed by calculating an excavation's probability of puncture, ignited release, and thermal harm for both gases. In terms of risk tolerability, hydrogen pipelines must ensure the same risk levels as the existing natural gas pipeline system. Lastly, we will conduct four case studies designed to engage with stakeholders and inform different aspects of the discussion around the use of hydrogen in natural gas pipelines.

- **Study 1**: Compare the probabilities of puncture, ignited release, and thermal harm caused by TPD in hydrogen and natural gas contexts.

- **Study 2**: Evaluate the effect of hydrogen on the risk of punctures, ignited releases, and thermal harms in a distribution system composed of different shares of steel and plastic pipes.

- **Study 3**: Evaluate the probability of alternative consequences to past incidents experienced by transmission pipelines had they transported hydrogen instead of natural gas.

- **Study 4**: Identify the most relevant risk-influencing factors affecting the risk of TPD to distribution and transmission pipelines transporting hydrogen.

Case Study 1 is designed to provide a baseline comparison between the TPD risks for both hydrogen and natural gas. Case Study 2 will provide insight into how the risk of TPD changes for distribution systems comprising different proportions of steel and plastic pipelines. Case Study 3 seeks to understand how past incidents could have changed if hydrogen were in the pipeline instead of natural gas. Case Study 4 seeks to identify the factors that are most responsible for differences in risk between hydrogen and natural gas.

The risk assessment methodology used in this work follows the Hydrogen Risk Assessment Model (HyRAM) methodology from Groth et al. [34], shown in Figure 6.1, which is tailored to assessing the risk of hydrogen systems. The analysis scope was covered in this section, while the remaining sections will show how each step in the HyRAM methodology is informed and performed to assess the risk of TPD to hydrogen pipelines.

6.3 System and Process Description

Third-party excavation damages to existing natural gas pipelines transporting hydrogen will be assumed to occur in the U.S. natural gas pipeline system and to follow this country's safety practices. The effect of hydrogen on natural gas pipelines will also be focused on U.S. infrastructure. Each of these aspects is described as follows:

- **U.S. Natural gas pipeline system.** Natural gas is delivered throughout the U.S. from high to low-pressure systems. Starting from gas wells and gathering systems, natural gas is first transported through an intra- and interstate transmission pipeline system. This system is composed of more than 500,000km of pipelines with high pressure (4-12MPa) and large diameter (50-1,120mm) and is mainly made from high-strength steels (mainly API 5L X42-X80) [126, 200]. Additionally, transmission lines are required to be buried between 0.9-1.2m depending on the area's land type, population, and use. Also, to ensure that gas remains pressurized, compression stations are placed at 65-160km intervals. Transmission pipelines feed into a lower-pressure distribution system through a "city gate," bringing natural gas to homes and businesses. This distribution system comprises more than 3 million km of mains and service pipelines, with pressures ranging from 0.1-0.6MPa and diameters of 25-560mm. Additionally, they are required to be buried between 0.3-0.5m. Distribution pipelines are mainly made of plastics (∼70%) and low-grade steels (∼28%) and are the most exposed to TPD.

  Pipeline operators use a supervisory control and data acquisition system to monitor for any flow and pressure problems and incorporate the ability to operate compressors and valves remotely. This allows operators in a control center to adjust and isolate the flow rate through a pipe in case of any incident, such as a gas release due to TPD.

- **Third-party excavation process.** A gas release is a major hazard, so a number of safety barriers are defined and used in an excavation process to prevent it. A general excavation process in the U.S. has been structured around the Common Ground Alliance's (CGA) best digging practices [155]. According to the CGA, a third-party excavator marks the anticipated digging area before excavating. They then call the local 811 One Call center or submit an online form providing site information at least two working days before their activities start. The responsible utilities are notified, and a utility service representative locates and marks any underground facilities. Then, the One Call center gives clearance to the excavator to start their activities. The excavator proceeds to dig safely by exposing underground utilities using hand tools to avoid accidental damage. A set of additional digging best practices are recommended by CGA to excavators in order to prevent damaging underground utilities [155]. These steps represent the main safety barriers used to prevent a gas pipeline from getting hit. A failure to appropriately execute any of these steps, in addition to a pipe hit with a strong enough force to overcome the pipe's inherent resistance, may result in a gas release. After identifying a gas release, excavators should immediately contact relevant authorities to enable utility managers to employ their emergency response plans (e.g., impede gas flow). If this recovery barrier fails, consequences can range from property damage to thermal harm and loss of life caused by jet fires or explosions. It is important to highlight that jet fire scenarios involve the immediate ignition of a gas release. Therefore, the impediment of the flow of gas and the execution of an emergency response plan will likely be performed after the consequences of a jet fire are observed. The safety and recovery barriers associated with a third-party excavation activity are summarized in the bow-tie diagram of Figure 6.2.

- **The effect of hydrogen on natural gas pipelines.** Transporting hydrogen through existing natural gas pipelines is expected to change the operational conditions of transmission and distribution pipeline systems. Hydrogen is expected to be unable to maintain similar energy flows to the current natural gas system, so hydrogen pipelines must operate at higher pressures and/or velocities [201]. This increase in pressure or velocity is due to the lower volumetric energy density of hydrogen compared to natural gas (12.7 and 40MJ/m^3, respectively [202]). Additionally, hydrogen flows will experience more severe pressure drops compared to natural gas [203] and will require additional compression stations along existing pipelines. Hydrogen is also expected to pose additional safety challenges that are not yet well understood, including hydrogen embrittlement to pipeline steels and hydrogen's unique gas release and ignition behavior. Hydrogen embrittlement is expected negatively impact the ductility and fracture toughness of steel pipelines [136], which can affect the frequency of failures caused by gouges, dents, and punctures in excavation activities. Hydrogen release and ignition behavior can pose unique challenges in terms of ignited releases and jet fires. Hydrogen can ignite at much wider concentrations in air than natural gas, even with low-energy sources [143], due to its low minimum ignition energy of 0.017mJ (compared to methane's 0.3mJ) and wide flammability range of 4%-75% volume in air (compared to methane's 5%-15%) [202]. Subsequent fires are expected to be much faster and hotter than their natural gas counterpart, posing a major hazard to people near the source of an ignited release (e.g., excavator) [129]. Nevertheless, hydrogen jet flames are expected to have a smaller size and lower heat radiation than natural gas jet flames, potentially reducing the risk of thermal harm and loss of populations further from an ignition source.

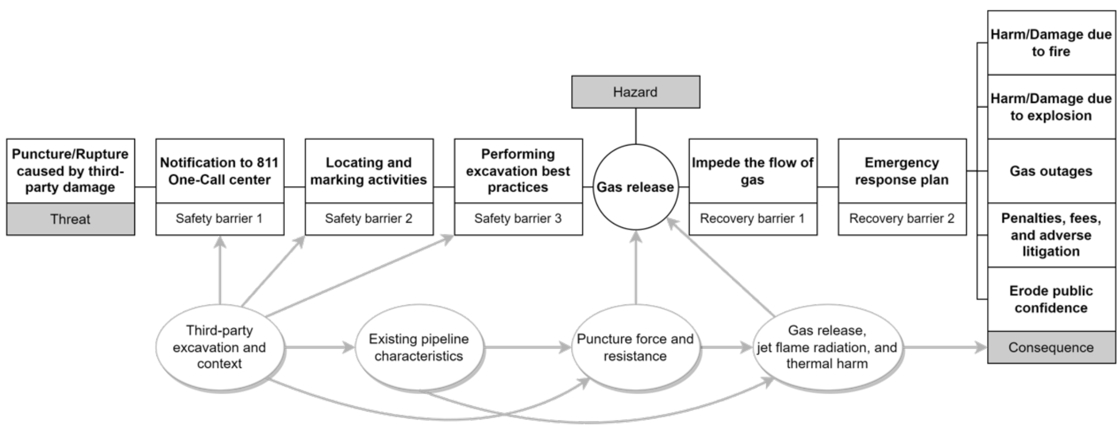


Figure 6.2: Bow-tie diagram for a TPD threat. The bow-tie's barriers and hazard will depend on a number of interacting variables, as shown in the grey circles in the figure. These dependencies and interactions are shown as gray arrows.

6.4 Modeling the Risk of Third-party Damage

The bow-tie diagram for a TPD threat in Figure 6.2 will be dependent on a number of interacting variables describing the excavation process, its context, and the potential consequences of a gas release. These variables can be grouped as follows:

1. *Third-party excavation and context*: This group contains all variables involved in the excavation process and the safety barriers present in CGA's best digging practices [155]. These variables account for the environment in which an excavation is performed and the actors involved in it. The excavation environment includes the excavation site characteristics (e.g., location class, soil type, and pipe depth) and overarching legal & regulatory context (e.g., mandatory or voluntary notifications to one-call centers). The excavation actors include the third-party excavators (e.g., professional or general public), utility managers (e.g., utility company and locating and marking contractors), and the one-call center (e.g., online submission or calling 811) involved in the excavation process. These variables will affect the sufficiency of safety measures taken on an excavation (i.e., notification, locating and marking activities, and excavation practices) and the probability of a pipe hit. These variables were thoroughly studied in the development of BaNTERA in Chapter 4.

2. *Existing pipeline characteristics*: This group contains the variables describing the design and operational characteristics of natural gas pipelines. These characteristics include materials (e.g., API 5L steels, cast iron, and polymers), pressure (e.g., maximum allowable and operating pressure), geometry (e.g., diameter and wall thickness), and their depth of cover. These variables affect a pipeline's resistance to puncture and subsequent gas release characteristics.

3. *Puncture force and resistance*: This group contains the variables interacting in the process of puncturing an underground pipeline. Puncture force is affected

by the equipment used in the excavation activity, its exerted force, and its tooth geometry. Puncture resistance is affected by the pipeline's ultimate strength, toughness, and mechanical degradation conditions (e.g., hydrogen embrittlement). These variables affect the probability of a puncture given a pipe hit in an excavation activity and the size of a gas release. These variables were thoroughly studied in the development of BaNTERA in Chapter 4.

4. *Gas release, jet flame radiation, and thermal harm*: This group contains the variables governing the physical characteristics of the gas released from a pipeline and the potential consequences it can produce. These variables include the release size, associated mass flow, ignition probability, potential radiative heat flux from a jet fire, and the probability of burns and loss of life due to thermal exposure. These variables affect the probability of ignition and thermal harm.

The dependencies among the variables described above are shown as grey arrows in Figure 6.2. It is important to highlight that all of these variables will depend on the specific pipeline system (distribution or transmission) and transported gas (natural gas or hydrogen) being studied.

6.4.1 Methods: Bayesian Networks and Causal Reasoning

The variables and dependencies shown in Figure 6.2 will be integrated into a single model in order to encode the many different scenario configurations of an excavation process. A BN modeling approach is used to integrate the many factors involved in an excavation process and its potential consequences, as it is tailored towards representing the joint probability distribution of multiple variables and their dependencies. Furthermore, BNs causal reasoning capabilities will be crucial in assessing and comparing the risks of natural gas pipelines to hydrogen pipelines.

BNs have already been demonstrated to be a suitable modeling approach for the QRA of TPD in natural gas pipelines, as demonstrated with the BaNTERA model in Chapter 4. The model developed in this work is heavily based on BaNTERA and includes additional variables and data on existing hydrogen transportation, pipeline characteristics, and potential consequences, allowing it to perform the QRA described in Section 6.2. Additionally, the studies to be performed in this QRA will require causal reasoning under uncertainty in the three rungs of the ladder of causality. Causal reasoning with BNs will be used for this task as follows:

- *Associative reasoning (rung 1)*: Informs Study 4 described in Section 6.2. This is done through a sensitivity analysis of the developed TPD risk model for hydrogen pipelines by modifying the marginal probability of the different variables of the model and propagating this information as evidence in the BN to inform any potential increase of the probability of a puncture, ignited release, and thermal harm. Associations will be based on the factorization formula in Equation 2.3 of Chapter 2.

- *Intervention reasoning (rung 2)*: Informs Study 2 described in Section 6.2 by simulating an excavation process in a generic distribution system. Interventions will be used to force the system to be composed of different proportions of plastic and steel pipelines and to transport hydrogen or natural gas. These interventions will be used to estimate the causal effect that pipeline materials could have on the probability of punctures, ignited releases, and thermal harm in a distribution system transporting hydrogen and natural gas. Interventions will be based on the methods established in Chapter 3.

- *Counterfactual reasoning (rung 3)*: Informed Study 3 described in Section 6.2 by simulating known past incidents caused by third parties in a utility company's natural gas transmission system. These past incidents will be modified through probabilistic counterfactual methods to evaluate what could have happened in terms of puncture and ignition probabilities had hydrogen been transported through the affected pipelines instead of natural gas. Counterfactuals will be based on the methods established in Chapter 5.

6.4.2 Information Sources

Multiple information sources, from public databases to codes and standards, are used in this work to inform the proposed TPD risk model. In order to reflect the present conditions of pipeline TPD risk, only data from 2016 to the present was used. Much of the information used in this work was also used to build the BaNTERA model in RT2.1. Therefore, the following sources will only describe the additional information used for expanding BaNTERA to hydrogen and to enable the assessment of consequences from TPD.

- *BaNTERA, A Bayesian Network for Third-Party Excavation Risk Assess-*

ment. Our previous work and model on TPD risk, BaNTERA [23], provides a comprehensive taxonomy of the actors, objects, and environments within a third-party excavation activity. Furthermore, BaNTERA provides a dependency structure for the variables describing an excavation process, subsequent puncture damage, and conditional probabilities describing the dependencies between these variables. A utility partner, CGA, and information from GTI Energy and PHMSA informed this dependency structure and probabilities. We refer the reader to Chapter 4 for a description of the information sources used to build BaNTERA and, therefore, part of the model proposed in this work.

- *GTI Energy*. GTI Energy provided guidance on informing the operational pressure ranges of existing pipelines in the U.S., as they are not commonly available to the public. Additionally, GTI Energy has conducted numerous research on the effect of hydrogen on the existing U.S. natural gas pipeline system [126, 204] and provided us with guidance on the modifications needed to account for the effect of hydrogen on TPDs. These modifications are thoroughly described in Section 6.4.4.

- *U.S. DoT PHMSA*. PHMSA's annual report on natural gas pipeline mileage and facilities from 2021 [200] was used to inform the characteristics of existing pipelines, such as materials and pipeline sizes in transmission and distribution systems.

- *Codes and standards*. This study used two standards and one code to inform variables on existing pipeline characteristics. The API 5L [205] and ASTM D2513 [206] specifications for steel and plastic pipe, respectively, were used to correlate a pipe size with its wall thickness. The U.S. code of federal regulations CFR49.192 [207] on natural gas pipeline safety was used to inform the dependencies between a pipe size, material, and wall thickness with its design pressure.

- *HyRAM+*. Hydrogen Plus Other Alternative Fuels Risk Assessment Models (HyRAM+) is a software toolkit developed by Sandia National Laboratories [208] that integrates publicly available data and models relevant to assessing the safety in the use, delivery, and storage infrastructure of hydrogen and other alternative fuels (e.g., methane). HyRAM+ provides a reduced-order physics module capable of simulating a gas release through an orifice, plume dispersion, and a jet flame's temperature, trajectory, and radiative heat flux. This information will inform variables associated with a gas release and potential jet flame radiation. It is important to highlight that HyRAM+ is also available as a Python API, enabling its integration with external models and its use on a probabilistic framework by allowing the simulation of a gas release under multiple scenarios.

6.4.3 Modeling Process

The next step in this work is to propose and develop a process that enables the development of a risk model on TPDs to underground pipelines in both natural gas and hydrogen settings. A flowchart of this process is presented in Figure 6.3 and will be used to generate the necessary information required in a TPD model capable of estimating the probability of punctures, ignited releases, and thermal harm associated with a third-party excavation process (required for Study 1 from Section 6.2), in addition to allowing causal reasoning under uncertainty (required for Studies 2, 3, and 4 from Section 6.2).

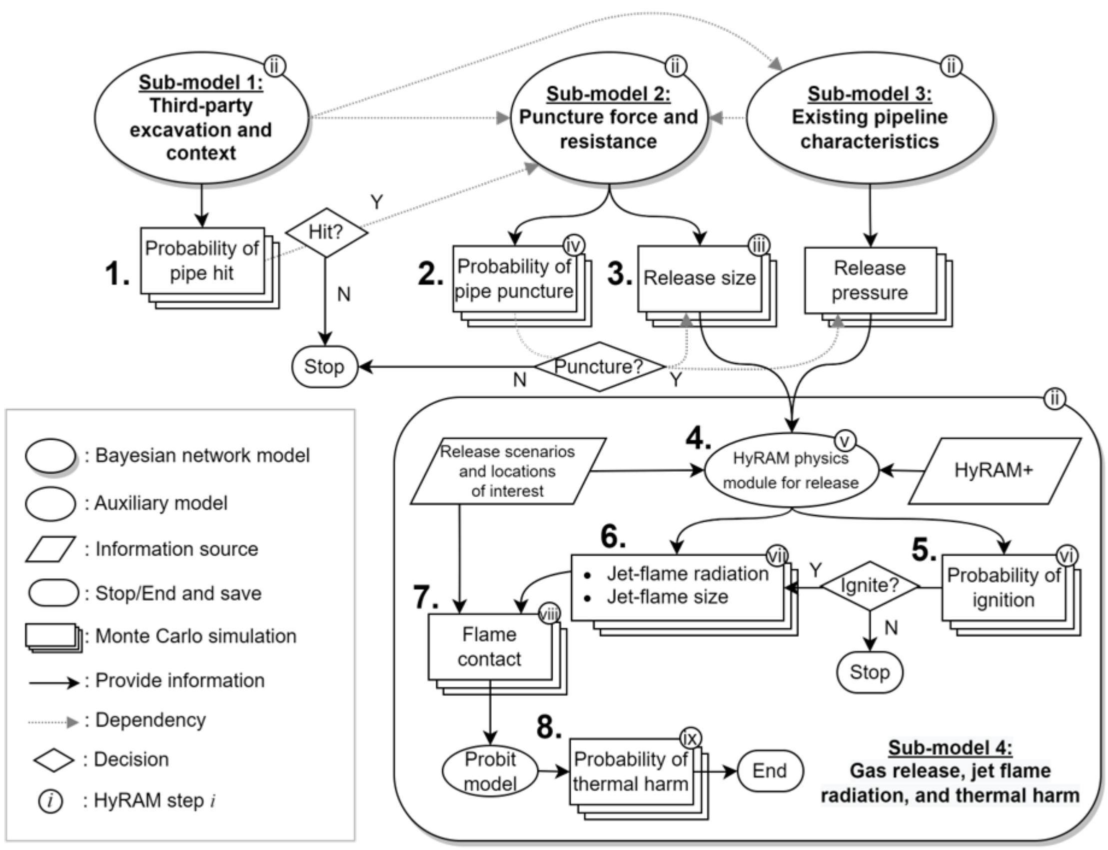


Figure 6.3: Process flowchart for modeling the risk of TPDs. Step i of the HyRAM methodology is omitted from this figure, as the analysis scope defined in Section 6.2 is what defined the structure of the flowchart

The flowchart in Figure 6.3 is informed by the sources described in Section 6.4.2. The variables that compose the groups “third-party excavation and context,” “puncture force and resistance,” and “existing pipeline characteristics” (described in

Section 6.4) are modeled using a BN approach and are called “sub-model 1,” “sub-model 2,” and “sub-model 3,” respectively. Information from utility partner, CGA, GTI Energy, and PHMSA, were already integrated into the BaNTERA model, and BaNTERA’s dependencies and conditional probabilities between variables are used to generate sub-model 1 and sub-model 2. GTI Energy, PHMSA, and pipeline codes and standards inform the variables and dependencies in Sub-model 3. Sub-models 1-3 encode in their structure a description of the threat, safety barriers, and hazard of the bow-tie diagram shown in Figure 6.2. As such, they represent Step ii of the HyRAM methodology (system description). The dependencies and conditional probabilities between variables from sub-model 1 affecting sub-models 2 and 3 are set through a BN approach, that is, by defining shared arcs and conditional probability tables or functions between the sub-models. These dependencies are shown as grey arrows in Figure 6.3. For example, the variable “Area development,” which describes the density of buildings and population near a pipeline’s right-of-way, is part of sub-model 1. However, area development will also affect the likelihood of observing a particular type of excavation equipment in sub-model 2 and the pipe’s depth of cover in sub-model 3. Dependencies between variables in sub-model 3 affecting sub-model 2 are also set, such as the dependency between a pipe’s puncture resistance and its material.

The process flowchart of Figure 6.3, used for developing a TPD risk model, is explained as follows:

1. Sub-model 1 simulates a single excavation process. If no pipe hit is associated

with the simulation, the process stops and records an excavation without a hit.

2. If a pipe hit instance is obtained from sub-model 1, sub-model 2 will simulate puncture instances conditioned on existing pipeline characteristics and the equipment used in the simulated excavation. The simulation of a pipe puncture instance informs Step iv of the HyRAM methodology (release probability). If there is no puncture associated with the simulation, the process stops and records an excavation with a pipe hit but without consequences.

3. If a pipe puncture instance is obtained from sub-model 2, a potential release size is simulated based on the tooth geometry and pipe size associated with the simulated excavation. The simulation of a release size informs Step iii of the HyRAM methodology (release size). Additionally, sub-model 3 will simulate a release pressure equivalent to the pipeline's operating pressure associated with the simulated excavation.

4. The simulated release size and pressure are fed into the HyRAM+ toolkit's physics module to simulate the gas release characteristics. This auxiliary model uses the HyRAM+ toolkit as an information source for the physics models describing the released gas equation of state, combustion, developing flow, and unignited release characteristics. The simulation results from the HyRAM+ toolkit inform Step v of the HyRAM methodology (release characteristics). The HyRAM+ toolkit's physics module will be used to estimate the steady-state mass flow rate of the released gas.

5. The steady-state mass flow rate of the released gas will determine its ignition probability (this correlation is further explained in Section 6.4.4). As such, an ignited release will be simulated and conditioned on the gas mass flow. The simulated ignition informs Step vi of the HyRAM methodology (scenario probability). If there is no ignition associated with the simulation, the process stops and records an excavation with a pipe hit and puncture but without an ignited release or harm.

6. If an ignited release instance is obtained in the simulation, the HyRAM+ toolkit's physics module will be used to simulate a jet-flame size and thermal radiation. This step uses a set of manually defined release scenarios of interest, such as the release angles along the diameter of a pipe. The simulated jet-flame results from the HyRAM+ toolkit inform Step vii of the HyRAM methodology (flame radiation).

7. The resulting jet flame from the previous step is used to determine flame contact on a set of manually defined locations in the area surrounding the simulated release. These locations represent the positions in which people can be exposed to thermal radiation and will depend on the simulated size of the jet fire. The defined flame contact locations inform Step viii of the HyRAM methodology (flame contact).

8. Last, the simulated thermal radiation at the locations of interest defined in the previous step informs a Probit model that estimates the probability of harm and fatality for a given thermal exposure. The simulated probability

of harm and fatality informs Step ix of the HyRAM methodology (thermal harm). At this point, the process ends and records an excavation with a pipe hit, puncture, ignited release, and potential thermal harm and fatality.

Steps 4-8 describe the variables affecting gas release, jet-flame radiation, and thermal harm. These variables will be modeled as a Bayesian network and called "sub-model 4." This sub-model encodes in its structure a description of the hazard and consequences of the bow-tie diagram shown in Figure 6.2, thereby representing Step ii of the HyRAM methodology.

The process described above is performed separately for distribution and transmission pipelines. Also, for each system, 10 million simulations are used to obtain a probability distribution over the simulated pipe hits, punctures, ignited releases, and thermal harm. The assumptions used in the modeling process, which will affect the results of these simulations, are described in the next section.

6.4.4 Modeling Assumptions

A number of modeling assumptions are used in this work to inform the different variables in sub-models 1-4, and some of these assumptions will be highly dependent on the gas transported. Modeling assumptions involving a third-party excavation activity are already defined in our previous work on the BaNTERA model (see Chapter 4). For the present study, we assume that an excavation process involving natural gas pipelines will not change in a hydrogen setting. That is, the safety barriers shown in the bow-tie diagram of Figure 6.2 will remain unchanged. However, this is not true for the hazard and consequences shown in the figure. For sub-models 2-4, the following assumptions are made:

6.4.4.1 Puncture Resistance and Hydrogen Embrittlement

Puncture is driven by a ductile failure that involves the plastic deformation of the pipe wall area in contact with the excavation equipment tooth, the generation of a failure on the inside surface of the pipe directly beneath the tooth, and the propagation of the failure through the pipe's wall until the penetration of the tooth. The Brooker's model for puncture failure [165] has been proven as a suitable model for generating failure predictions in steel pipelines under multiple tooth geometries, pipeline designs, and material grades [23,122]. As such, this model is used to describe puncture failure in steel pipelines and is mathematically defined as:

$$\begin{gathered} R_p = 7.0074 \times 10^{-7} t(\sigma_u + 410.4)(L + 22.41)W(3.142 + W) \\ \text{Puncture failure: } R_p < F \end{gathered} \tag{6.1}$$

Where R_p is a pipe's puncture resistance in [N], F the puncture load in [N] exerted by excavation equipment, t is the pipe wall thickness in [mm], σ_u is the ultimate stress of the pipeline material in [MPa], and L and W are a tooth's length and width, respectively, in [mm]. This model has been adapted in [122] to consider plastic pipelines by modifying the term $(\sigma_u + 410.4)$ to $(\sigma_u + 0.4)$ in Equation 6.1. This modification implies that punctures in plastic pipes are expected to be 2 times more likely than steel pipelines. Equation 6.1 is the center of sub-model 2 “puncture force and resistance” in the risk modeling process of Figure 6.3.

A major concern regarding the use of hydrogen on natural gas pipelines made of steel is hydrogen embrittlement. Hydrogen can dissociate on the pipeline's inner surface, dissolve into the metal lattice, and change its mechanical response to stress. A number of works have shown that steels in both transmission and distribution systems can be negatively affected by hydrogen. For example, San Marchi and Somerday [135] show that carbon pipe steels ranging from API 5L Grade B to X80 can present a significant degradation of reduction in area under tensile loads as well as a reduction of fracture toughness (50% reduction on average). Similar results can be found in the literature [136, 137], with all of them concluding that high-strength steels can be the most affected by hydrogen deterioration. For low-strength steel, research by Nykyforchyn et al. [138] and Tsyrul'nyk et al. [139] showed that steel pipes used for distribution mains are highly sensitive to hydrogen action, showing a significant decrease in impact toughness and brittle fracture resistance. To the best of the authors' knowledge, there is a research gap in the study of the effect of hydrogen on a pipeline's puncture resistance. However, according to material science experts from our industry partners, hydrogen's effect on punctures will likely be similar to its effect on fractures. Hydrogen embrittlement could potentially decrease the plastic deformation associated with a puncture failure, allowing a faster propagation of the crack formed at the inner surface of a pipe (in contact with hydrogen) at the location of the excavation equipment's tooth. Therefore, for this preliminary risk model, the following linear multiplicative factor is assumed to account for the effect of hydrogen embrittlement on a pipe's puncture resistance:

$$
\begin{gathered}
R_{p,\,H_2} = HE \times R_p \\
HE \sim Beta(5.3, 4.4)
\end{gathered} \tag{6.2}
$$

Where $R_{p,\,H_2}$ is the puncture resistance of a metal pipeline in contact with hydrogen, HE is the multiplicative puncture resistance decrease factor accounting for hydrogen embrittlement, and $Beta(\,)$ is the Beta probability distribution. As seen in Equation 6.2, HE is assumed to follow a Beta distribution constrained between 0 and 1. This distribution was defined based on an expert elicitation on the effect of hydrogen on puncture resistance, for which the following percentiles were obtained for the values that the HE factor can take: 1%: 0.1, 5%: 0.3, 50%: 0.55, 95%: 0.9. As such, it is expected that a natural gas pipeline's puncture resistance will decrease, on average, by a factor of 0.55 in a hydrogen setting. This decrease in puncture resistance can greatly impact the risk of TPD in transmission pipelines, which are mostly made of steel. However, distribution lines will also be impacted as a third of them in the U.S. are made from steel.

6.4.4.2 Operational Conditions

Natural gas pipelines in the U.S. are mainly underground and can be assumed to have a uniform gas temperature of 288K and no heat transfer between the gas and the surrounding soil (i.e., isothermal flow is assumed). As it was described in Section 6.3, operational pressures will be very different in distribution and transmission systems. In this work, the following probability distributions were elicited from a utility expert for an underground pipeline's operational pressure:

$$
\begin{gathered}
P_D \sim Uniform(4.2,\, 8.4) \\
P_T \sim Triangular(0.11,\, 0.17,\, 0.51)
\end{gathered}
\tag{6.3}
$$

Where P_D and P_T are the operational pressure in MPa of the U.S. distribution and transmission system, respectively. Operational pressures are commonly lower than a pipeline's design pressure, which will be determined by the pipe's material, diameter, and wall thickness [200]. To account for the dependency between pressure and a pipeline geometry, the design pressure dictated in the code CFR49.192.C [207] will also be included in the model. The operational pressure of a pipeline is assumed to be the minimum value between its design pressure dictated in CFR49.192.C and the elicited operational pressures in Equation 6.3. A pipeline's geometry and pressure are the center of sub-model 3 "existing pipeline characteristics" in the risk modeling process of Figure 6.3.

Additionally, if natural gas is replaced with hydrogen, the amount of hydrogen supplied to end users must satisfy the energy content required. In standard conditions, hydrogen is expected to provide less energy than natural gas in pipelines, where the transported volume of gas is essential. Natural gas contains 3.3 times more energy than hydrogen in terms of volume (hydrogen and natural gas net heating values are 10,246 kJ/m^3 and 33,906 kJ/m^3, respectively). Although part of this energy difference will be compensated by hydrogen's faster velocity through pipelines (3 times faster than natural gas [203]), a 30% increase in operational pressure will still be required to satisfy the energy content required by users of the existing natural gas system [201, 204]. As such, this work assumes that hydrogen pipelines will operate at 1.3 times the pressure of existing natural gas pipelines.

6.4.4.3 Gas Release and Ignition

In this work, we assume that a puncture will produce a circular release with a diameter equivalent to the tooth area that punctures the pipe. As shown in our previous work for BaNTERA [23], a tooth length is assumed to be $L \sim Uniform(10, 150)$mm, and the tooth width to be $W \sim Uniform(3, 20)$mm. As such, a puncture equivalent diameter will be assumed to range between 6.2 to 60mm. If a tooth's equivalent circular diameter is higher than a pipeline's diameter, a full rupture of the pipe is assumed. Additionally, a discharge coefficient of 1 is assumed for conservatism.

For a gas release scenario, this work assumes an isentropic release of the gas, constant pressure inside the pipeline, and constant outlet temperature [209]. Additionally, methane (CH_4) will be used to represent natural gas. The mass flow from a simulated release will be assumed to be in a steady state and is calculated using the HyRAM+ toolkit [208]. This calculation is based on a simulated pipeline's pressure, diameter, and release size obtained from sub-models 1-3 in Figure 6.3. The calculated mass flow from a release will be correlated to an immediate ignition probability (see sub-model 4) as defined by Tchouvlev et al., [210], and shown in Table 6.1. This work does not consider mass flow adjustments due to potential crater formation.

Although static electricity buildup and discharge is a known ignition hazard in releases from plastic pipelines [142], its effect on ignition probability will not be considered in this work.

(a)

Natural Gas (Methane)	
Mass flow, kg/s	Ignition probability
<1	0.007
1-50	0.047
>50	0.200

(b)

Hydrogen	
Mass flow, kg/s	Ignition probability
<0.125	0.008
0.125-6.25	0.053
>6.25	0.230

Table 6.1: Immediate ignition probabilities for (a) natural gas and (b) hydrogen. Based on [5].

6.4.4.4 Thermal Harm and Loss

Jet fires from ignited releases will be considered perpendicular to a pipeline's right-of-way. An individual's exposure to thermal radiation (measured as radiant heat flux) will be calculated at every meter in a radial line of 202m starting from the ignition source. This decision was made according to CFR49.192.5 [207] definition of location areas for risk assessment purposes. Additionally, thermal radiation will be measured at the height of 1.5m from the gas release source. In general, pipelines are required to be buried between 0.3-1.2m, so it can be assumed that the radiation from a jet fire at the height of 1.5m could contact an individual. Although pipelines can be buried deeper underground, 1.5m was assumed as a conservative height for flame contact. As such, thermal radiation will be evaluated at every potential location of an individual inside a radius of 202m from a jet fire and summarized as a probability distribution. These assumptions are depicted in Figure 6.4.

Thermal radiation is calculated using the HyRAM+ toolkit assuming the Yuceil/Otugen notional nozzle model [208], relative humidity of 0.89, and release pressure and size simulated from sub-models 2 and 3 in the risk modeling process of Figure 6.3.

Thermal harm and loss are calculated using the Tsao & Perry probit model for 2nd-degree burns and fatalities, as suggested in [5] for jet fires. Thermal harm depends on the thermal dose unit V, which combines the heat flux intensity, I, and exposure time, t, and is mathematically represented as follows:

$$V = I_{contact}^{4/3} \times t \tag{6.4}$$

Where $I_{contact}$ is the radiant heat flux $[W/m^2]$ from a jet fire, and t is the exposure duration in seconds. The use of 20s of thermal exposure was based on recommendations from the Dutch Decree on the External Safety of Pipelines (BEVB) and previous QRAs for hydrogen pipelines [128]. The thermal harm probit equations for fatalities and burns are:

$$\begin{gathered} \text{Fatality}: \; Y_{thermal\,harm} = -36.38 + 2.56 \times \ln(V) \\ \text{2nd Degree Burns}: \; Y_{thermal\,harm} = -43.14 + 3.0186 \times \ln(V) \end{gathered} \tag{6.5}$$

The result of Equation 6.5 can be used to calculate the probability of thermal harm ($P_{thermal\,harm}$) as follows:

$$Pr_{thermal\,harm} = F(Y_{thermal\,harm} \,|\, \mu = 5, \sigma = 1) \tag{6.6}$$

Where $F(\bullet)$ is the normal cumulative distribution function. The radiant heat flux from a jet fire and associated probability of thermal harm and loss will be included in sub-model 4, "gas release, jet-flame radiation, and thermal harm" in the risk modeling process of Figure 6.3.

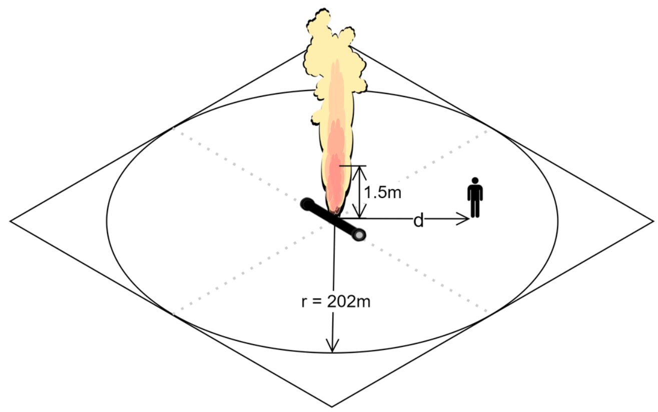


Figure 6.4: Thermal exposure to jet-fire. An individual will be considered to be located at a distance d $\in [0, 202m]$ from the ignition source.

6.5 Results

6.5.1 SHyTERP Model: Assessing Safe Hydrogen Transportation through Existing and Repurposed Pipelines

The resulting model of this research was named SHyTERP (Safe Hydrogen Transportation through Existing and Repurposed Pipelines). The SHyTERP model was specified for TPD risk assessment purposes and, as explained in Section 6.4, it comprises four sub-models, each of which is a BN encompassing distinct variables describing a third-party excavation process and its consequences. These BNs are presented in Figures D.1-D.4. As can be seen in these figures, four variables (highlighted in pink) are directly influenced by hydrogen, describing the gas' effect on a pipeline's pressure, resistance to puncture, gas release ignition probability, and jet flame characteristics. Table 6.2 shows network characteristics of sub-models 1-4. The parameterization of the variables not described in Sections 6.4.3 and 6.4.4 can be found in our previous work on BaNTERA [23]. Sub-model 1 is the most extensive of the four sub-models, including 67% of the variables in the proposed risk model and 68% of its probabilistic dependencies. Also, sub-model 3 "pipeline characteristics," is the most influential in the risk model and has the highest connectivity as sub-models 1, 2, and 4 depend on variables contained in sub-model 3. Overall, the proposed SHyTERP model for TPD risk is highly complex, presenting a total of 73 interacting variables and 110 conditional probabilities between them.

			# Cross-dependencies			
Sub-model (SM)	**# Variables**	**# Dependencies**	**SM-1**	**SM-2**	**SM-3**	**SM-4**
SM-1: Third-party excavation and context	49	75	-	3	1	0
SM-2: Puncture force and resistance	8	7	0	-	0	4
SM-3: Existing pipeline characteristics	10	11	1	3	-	2
SM-4: Gas release, jet flame radiation and thermal harm	6	5	0	0	0	-
Third-party damage risk model (overall)	**73**	**110**				

Table 6.2: Characteristics of the BNS composing sub-models 1-4 in SHyTERP. Variables and dependencies correspond to the number of nodes and arcs in each network. Cross-dependencies correspond to the number of arcs from a sub-model in a row to another in a column of the table.

6.5.2 Total Risk and Validation

10 million Monte Carlo simulations of SHyTERP were run for transmission and distribution systems transporting natural gas and hydrogen. For hydrogen pipelines, the average, minimum, and maximum values of the simulations are shown in Tables 6.3 and 6.4 (see 100% H_2 columns). Additionally, each simulation's cumulative distribution function (CDF) is presented in Figure 6.5. These results represent the total risk for each system, divided into the probability of pipeline puncture, imme-

diate ignition, 2nd-degree burns, and fatality per excavation activity. A frequentist interpretation of the probabilistic results of the model will be used below in order to facilitate the comparison of the total risk of TPD's risk in hydrogen and natural gas pipelines, which is often discussed in terms of frequencies.

Based on the marginal probabilities presented in Table 6.3, hydrogen distribution pipelines are expected to be punctured 2.84 times every 1,000 third-party excavations. 5.32×10^{-2} of those punctures are expected to ignite and produce a jet fire, and 2.67×10^{-4} of these fires are expected to cause a fatality. Based on Table 6.4, hydrogen transmission pipelines are expected to be punctured 1.55 times every 1 million third-party excavations. 0.13 of those punctures are expected to ignite and produce a jet fire (matching the fact that higher release rates are expected from punctures in transmission pipelines), and 2.24×10^{-2} of these fires are expected to cause a fatality. These results show that, overall, ignited releases in distribution pipelines are expected to be much more frequent than ignitions in transmission lines due to higher exposure to puncture damage. However, an immediate ignition is expected to be significantly more likely in a release from a transmission pipeline. The same trend was obtained for a fatality. In addition, it should be highlighted that uncertainty ranges for hydrogen ignitions and thermal harm are wide, and the 50th percentile of the CDFs presented in Figure 6.5 show that these events could be, in general, highly infrequent.

In addition to calculating the total risk, the proposed SHyTERP risk model for TPD was validated against U.S. performance measures and statistics collected by PHMSA between 2010-2020 [173]. Validation is a challenging task, as the proposed

	Distribution pipelines (average, min, max)		
	100%CH_4 (per excavation)	100%H_2 (per excavation)	μ_{H_2}/μ_{CH_4}
$Pr(Puncture)$	2.73×10^{-3} (2.24×10^{-3}, 3.25×10^{-3})	2.84×10^{-3} (2.34×10^{-3}, 3.40×10^{-3})	1.04
$Pr(Ignition)$	1.99×10^{-5} (0.00, 6.97×10^{-5})	1.51×10^{-4} (2.99×10^{-5}, 2.89×10^{-4})	7.60
$Pr(Burn)$	8.57×10^{-8} (0.00, 9.95×10^{-6})	1.48×10^{-7} (0.00, 1.99×10^{-5})	1.73
$Pr(Fatality)$	4.22×10^{-8} (0.00, 9.95×10^{-6})	4.03×10^{-8} (0.00, 9.95×10^{-6})	0.95

Table 6.3: Average, minimum, and maximum simulated values for distribution pipelines transporting natural gas (100% CH_4) and hydrogen (100% H_2).

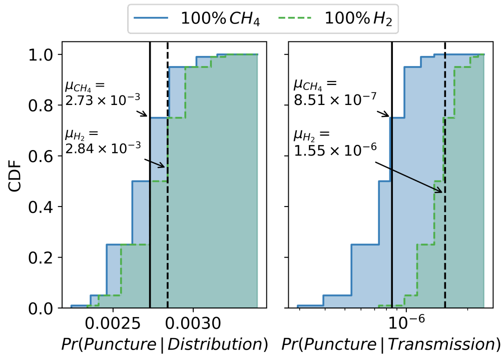
	Transmission pipelines (average, min, max)		
	100% CH_4 (per excavation)	100% H_2 (per excavation)	μ_{H_2}/μ_{CH_4}
$Pr(Puncture)$	8.51×10^{-7} (2.94×10^{-7}, 1.72×10^{-6})	1.55×10^{-6} (7.35×10^{-7}, 2.4×10^{-6})	1.82
$Pr(Ignition)$	3.92×10^{-8} (0.00, 2.45×10^{-7})	1.96×10^{-7} (0.00, 5.39×10^{-7})	5.00
$Pr(Burn)$	5.96×10^{-10} (0.00, 2.93×10^{-8})	4.87×10^{-9} (0.00, 9.31×10^{-8})	8.17
$Pr(Fatality)$	5.74×10^{-10} (0.00, 4.90×10^{-8})	4.40×10^{-9} (0.00, 9.80×10^{-8})	7.67

Table 6.4: Average, minimum, and maximum simulated values for transmission pipelines transporting natural gas (100% CH_4) and hydrogen (100% H_2).

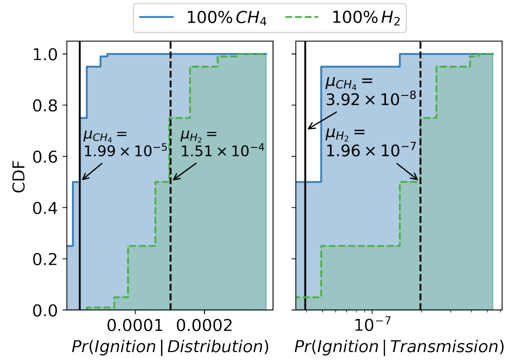
model is highly complex and nationwide data on pipeline's risk is limited to incidents involving considerable consequences, a small subset of the day-to-day events occurring in the U.S. pipeline system [167]. Additionally, there is no operational data on repurposed hydrogen pipelines, and excavation damage data is much more exhaustive and consolidated for distribution pipelines. Therefore, the validation of the proposed model will be done only on natural gas distribution pipelines. Assuming that there is an average of 20 million unique third-party excavations per year in the U.S. (based on one-call notification statistics [20]), the validation results for the proposed risk model are shown in Table 6.5. To obtain these results, a second model based on SHyTERP was developed for natural gas. This model has the same nodes presented in Figures D.1-D.4 but with modifications to represent natural gas transportation through the pipes [2].

Table 6.5 shows that SHyTERP can conservatively match PHMSA's perfor-

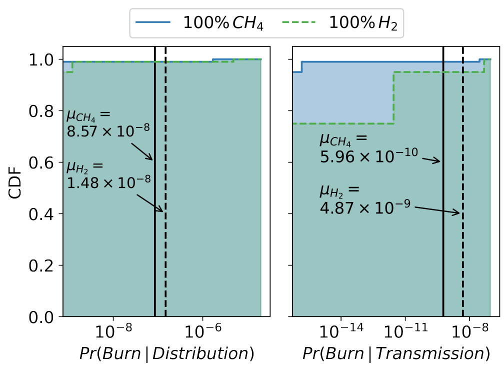
[2] Pink nodes in Figures D.1-D.4 were adapted for natural gas. Hydrogen embrittlement and pressure increase were eliminated from the model. Additionally, immediate ignitions and jet fire radiation were specified for natural gas.



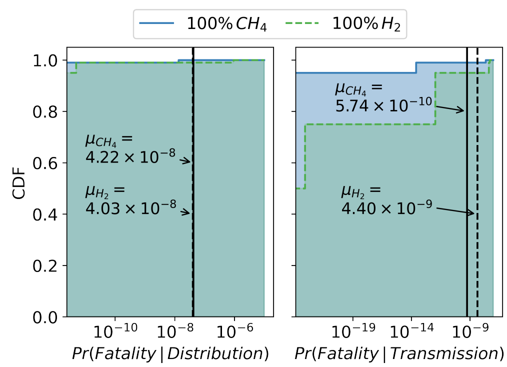
(a) CDF for the probability of puncture



(b) CDF for the probability of ignition



(c) CDF for probability of 2nd degree burn



(d) CDF for the probability of fatality

Figure 6.5: Total risk of TPD to natural gas pipelines transporting hydrogen.

mance measures and statistics for the U.S. natural gas distribution system. The proposed model estimates a lower expected number of damages per 1,000 tickets, which is 0.81 times the current U.S. statistic. Nevertheless, it should be considered that it only simulates punctures, one of the many failures a pipeline can suffer from in an excavation (e.g., dents and gouges). Regarding yearly rates for injuries and fatalities, the proposed model shows a very close result in terms of injuries, being 1.2 times higher than PHMSA's statistics. However, the model's results for yearly fatalities are conservative compared to PHMSA's statistics.

Performance measure/ statistic	PHMSA Average 2010-2020	Proposed Risk Model (Average)
Excavation damages per 1,000 tickets	3.10	2.52* (punctures)
Injuries per year	1.40	1.71
Fatalities per year	0.10	0.86

Table 6.5: Validation of the proposed SHyTERP risk model for third-party excavation damages to natural gas pipelines. Validation is only done for natural gas distribution systems. (*) The proposed risk model can only simulate punctures, not damages in general. PHMSA defines damage as any impact that degrades the integrity of a gas distribution facility, in which puncture is only a part of it.

6.6 Case studies

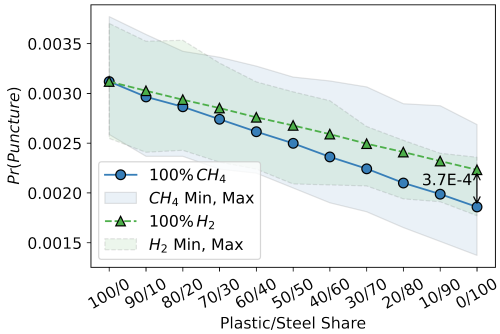
As explained in Section 6.2, four studies are performed to provide insights on the difference in TPD risk between natural gas and hydrogen distribution and transmission pipelines. The results of these studies are shown below.

6.6.1 Study 1: Comparison of Third-Party Damage Risk in Hydrogen and Natural Gas Settings

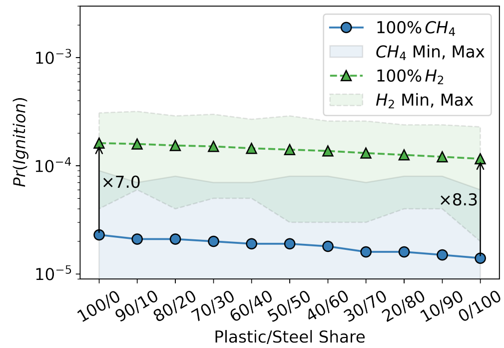
To provide a comparison of TPD risk in both hydrogen and natural gas contexts, SHyTERP's and its natural gas counterpart simulation results are shown in Tables 6.3 and 6.4 (see 100% CH_4 columns), and Figure 6.5.

Table 6.3 on distribution pipelines shows that the expected probabilities of punctures, burns, and fatalities are similar between the two gases. However, immediate ignitions are expected to be 7.6 times more frequent in hydrogen pipelines compared to natural gas pipelines. Despite the high expected frequency of ignited releases, jet fires from hydrogen distribution pipelines are expected to be less fatal than natural gas. This result can be attributed to the fact that hydrogen flames have less radiative emissive power than natural gas flames, especially when there are low concentrations of gas in the air, as it is expected to occur on a small release from a distribution line [130]. Figure 6.5 shows that the simulated CDFs for punctures, ignitions, and thermal harm in hydrogen pipelines are shifted to the right compared to natural gas CDF results, showing a higher total risk overall. Although fatalities are expected to be less frequent in TPDs to hydrogen distribution pipelines, the CDF results in Figure 6.5d show that a fatality is expected in 5% of the simulated excavations. For natural gas distribution pipelines, a fatality was expected only in 1% of the simulations, which shows that fatalities from hydrogen incidents could be more frequent than in natural gas incidents.

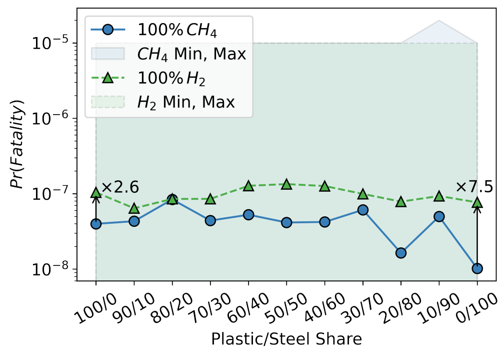
Table 6.4 on transmission pipelines shows that hydrogen's total risk is consistently higher compared to natural gas. An interesting result is that the frequency of thermal harm in hydrogen transmission pipelines seems to be mainly dependent on an increase in ignited releases rather than on the additional puncture failures driven by hydrogen embrittlement. This result is supported in Figures 6.5a and 6.5b. Figure 6.5a shows that a puncture is possible for all of the simulated excavations in both natural gas and hydrogen transmission settings and that their mean puncture probability differs by a factor of 1.82 (see Table 6.4). Figure 6.5b shows that ignitions were expected to be possible on only 50% of the simulated excavations in natural gas pipelines, increasing to 95% in a hydrogen setting. Furthermore, the ignition of hydrogen from releases in transmission pipelines is expected to be 5 times more frequent than from natural gas (see Table 6.4).



(a) Probability of puncture



(b) Probability of ignition



(c) Probability of fatality

Figure 6.6: Probability of puncture, ignition, and fatality in a distribution system for different shares of plastic and steel pipelines in both hydrogen and natural gas settings.

6.6.2 Study 2: The Effect of Hydrogen on the Risk of a Distribution System Composed of Different Shares of Steel and Plastic Pipes

The U.S. pipeline network has a heterogeneous share of plastic and steel distribution pipelines, and major work is being carried out in some states to transform the system to fully plastic in the near future (to the extent possible). However, plastic pipelines are more prone to getting punctured in excavation activities. As such, we sought to understand how injecting hydrogen into their distribution system could increase the risks posed by TPD.

To inform this inquiry, an intervention analysis (see Section 6.4.1) was performed by estimating the effect of hydrogen on a distribution system composed of 11 different percentual shares of plastics and steel. The intervention queries, Q, to be solved with the SHyTERP model proposed in this work can be expressed as follows:

$$
\begin{gathered}
Q = Pr\left(C = Yes \,|\, do\left(Material = \{Plastic = p,\, Steel = 100\% - p\}\right)\right) \\
C = \{Puncture,\, Ignition,\, Fatality\} \\
p = \{0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}\,\%
\end{gathered}
\tag{6.7}
$$

10 million simulations were performed for each percentual share of materials and for the transportation of both hydrogen and natural gas. The results obtained in the intervention analysis are shown in Figure 6.6 for the expected probabilities of puncture, ignition, and fatality, in addition to the minimum and maximum values obtained for them in the simulations (that is, the uncertainty range for the simulation's results).

Figure 6.6(a) on the probability of puncturing a pipeline shows that punctures are expected to be less frequent in distribution systems composed of steel, independently of the gas being transported through it. A major concern to our partners was that hydrogen embrittlement could significantly increase the likelihood of puncture on steel pipelines; however, our results show that the difference in punctures between natural gas and hydrogen is expected to be only 3.7 additional punctures every 10,000 excavations. Also, it is interesting to see that the uncertainty range for steel hydrogen pipelines is narrower than in natural gas. The reason for this is that hydrogen pipes were considered to operate with an additional 30% of internal pressure (see Section 6.4.4), which will limit the pipeline's design characteristics (e.g., wall thickness and diameter) compared to a natural gas pipeline. For hydrogen systems composed of 100% plastic pipes, it can be seen that the expected probability of puncture is the same regardless of the transported gas (which is also true for its uncertainty ranges). Further, the model's results suggest that a difference in the probability of punctures will not be significant in distribution networks with up to 70%/30% share of plastic and steel pipes, both in terms of expected values as well as uncertainty ranges for natural gas and hydrogen pipelines.

Although the risk of punctures in natural gas and hydrogen distribution pipelines were estimated to be similar, Figure 6.6(b) shows that ignited releases are expected to be consistently more frequent in hydrogen pipelines. This increase is expected to be more evident in systems made of steel, as an increase in punctures caused by hydrogen embrittlement will add up with the higher ignition probability of hydrogen. Analyzing the probability of fatalities caused by ignited releases in Figure 6.6(c), it can be seen that fatalities in hydrogen distribution pipelines are expected to be similar independently of the material shares and consistently higher by a factor of 2.87 (on average) compared to a natural gas system. Additionally, lower probabilities of fatality are observed for systems with material shares of over 80% steel pipes, which can be associated with a steel's higher resistance to puncture in addition to natural gas's lower probability of ignition. This reduction in fatalities is not present in systems transporting hydrogen. However, it should be noticed that uncertainty ranges for fatalities are significantly wide, ranging from 0 to 10^{-5} for both gases. As such, no certain conclusions can be provided in terms of fatalities beyond that they are expected to be in the same order of magnitude for both gases.

6.6.3 Study 3: The Effect of Hydrogen on the Outcomes of Past Incidents in Transmission Pipelines

Excavation damages to transmission pipelines are infrequent; however, they can lead to catastrophic consequences if the pipeline is punctured and the released gas ignited. Past incidents on the transmission pipelines have been observed to follow common patterns (e.g., damaged pipelines are usually 200-300mm in diameter and located in a Class 3 location). Future excavation incidents are expected also to follow these patterns, so a relevant insight towards assessing the risk of hydrogen transportation through existing transmission pipelines is to evaluate what could have happened in past incidents had hydrogen been transported through the pipelines instead of natural gas.

To inform this inquiry, a counterfactual analysis (see Section 6.4.1) was performed on a utility company's 2020-2021 recorded incidents to evaluate how transporting hydrogen could have impacted the probability of punctures and ignitions at the time of the events. Incidents in this time period included 18 pipe hits, one puncture, and no ignited releases. Then, to perform the counterfactual analysis, the following steps were performed on each incident instance:

1. Evaluate the prior probability of puncture and ignition conditioned on all

evidence recorded in the event on the variables included in the proposed TPD risk model (see Section 6.5.1). This prior probability will represent the moment just before the incident happens.

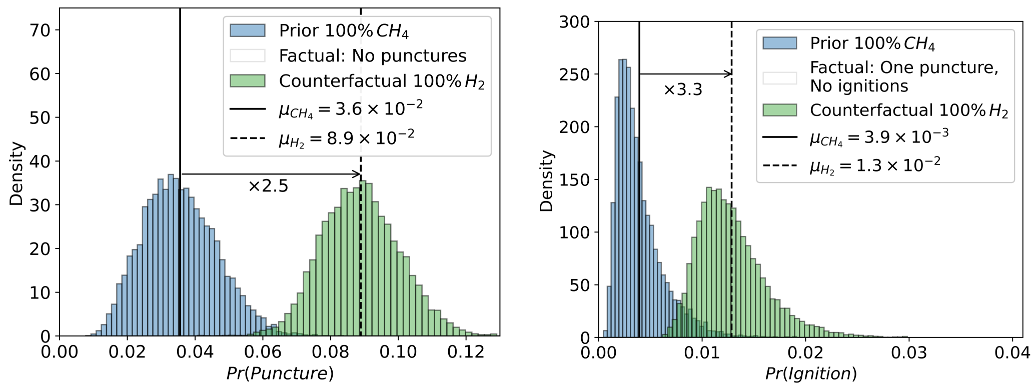
2. Evaluate with the model the following counterfactual queries Q:

$$Q = Pr(C_{Gas=H_2} = Yes \,|\, C = No,\, Gas = CH_4,\, E = e)$$
$$C = \{Puncture,\, Ignition\} \tag{6.8}$$

where $E = e$ is all evidence recorded in the event on the variables included in the proposed TPD risk model.

The prior and counterfactual probabilities of punctures and ignitions evaluated in steps 1 and 2, respectively, are used to compare hydrogen and natural gas transportation in transmission systems. In order to account for the uncertainty in the known population of past incidents observed in the utility's pipelines, ten thousand bootstrapping simulations [211] were performed on both the prior and counterfactual probabilities of punctures and ignitions. The results of these simulations are shown in Figure 6.7.

Figure 6.7(a) shows the average prior probability of puncturing a pipeline in the 19 incidents observed by the utility company. SHyTERP's natural gas counterpart estimates the prior puncture probability as 3.6×10^{-2} per excavation, which is equivalent to 0.69 expected punctures in the 19 incidents being studied. 1 puncture was observed between 2020-2021, which supports the model's accuracy. Looking now at the counterfactual probability of puncture had hydrogen been transported



(a) Puncture in past incidents

(b) Ignition in past incidents

Figure 6.7: Prior, factual, and counterfactual bootstrapped probabilities of puncture and ignition for past incidents in a utility's transmission system.

through the pipelines associated with these 19 incidents, the SHyTERP estimates that 2.5 times more punctures could have been expected (i.e., 8.9×10^{-2} punctures per excavation). As such, the counterfactual analysis shows that 2 punctures, rather than 1, could have been expected between 2020-2021 had hydrogen been transported through the pipelines rather than natural gas. Furthermore, Figure 6.7(a) shows that hydrogen could have caused the uncertainty range to shift towards higher probabilities of puncture in the 19 past incidents. Indeed, the lower bounds associated with the probability of punctures in a hypothetical hydrogen setting are expected to match the upper bounds of the factual natural gas setting.

Figure 6.7(b) shows the average prior probability of an ignited release in the 19 incidents observed by the utility company. SHyTERP's natural gas counterpart estimates the prior immediate ignition probability as 3.9×10^{-3} per excavation, which is equivalent to 7.4×10^{-2} expected punctures in the 19 incidents being studied. No ignitions were observed between 2020-2021, which supports the model's accuracy. Looking now at the counterfactual probability of ignited releases in a hypothetical hydrogen setting, the proposed SHyTERP model estimates that 3.3 times more ignited releases could have been expected on past incidents (i.e., 1.3×10^{-2} immediate ignitions per excavation). As such, the counterfactual analysis shows that for the 2 potential punctures expected on past incidents in a hydrogen setting, there could be a 25% chance of observing an ignited release. Furthermore, regarding the uncertainty of these estimations, we observe that hydrogen could have caused both a location and scale shift between the prior and counterfactual probability distributions for ignited releases. The 19 past incidents showed a prior probability of ignition that was positively skewed, presenting a mode of 2.2×10^{-3} ignited releases per excavation activity; an ignition was very unlikely to happen. However, the counterfactual probability distribution for an ignited release in a hypothetical hydrogen setting shows to be less skewed and with a mode of 0.011 ignited releases per excavation activity. Therefore, ignition would not have been unexpected.

6.6.4 Study 4: Most Relevant Risk-Influencing Factors

Last, a sensitivity analysis is performed on SHyTERP to identify the most relevant factors influencing the risk of TPD to natural gas pipelines transporting hydrogen. The target node for the sensitivity analysis is "Jet fire radiative heat flux" in sub-model 4 (see Figure D.4). The most relevant risk-influencing factors are considered as those variables that, by changing their marginal probabilities, will produce an increase in the observed heat radiation from a hydrogen release caused by TPD. This decision was based in that higher heat radiation from a hydrogen jet flame is correlated with a higher probability of fatalities, ignitions, and larger release sizes from punctures. The sensitivity analysis is based on Kjærulff and Van Der Gaag [172] sensitivity derivative parameter, S, for all the variables included in the models representing hydrogen transportation in both a distribution and a transmission system [3]. Table 6.6 shows the sensitivity analysis results.

(a)

100% H_2 **Distribution pipelines**

Rank	**Variable**	**Relevance**
1	Immediate ignition	1.00
2	Pipe pressure	0.57
3	Excav. preventive measures	0.53
4	Material	0.21
5	Wall thickness	0.12
6	Puncture force	0.11
7	Pipe diameter	0.10

(b)

100% H_2 **Transmission pipelines**

Rank	**Variable**	**Relevance**
1	Excav. preventive measures	1.00
2	H_2 embrittlement	0.72
3	Puncture force	0.68
4	Pipe pressure	0.53
4	Immediate ignition	0.53
5	Excavator tooth W	0.28
6	Wall thickness	0.25
7	Excav. equipment	0.24
7	Excavator tooth L	0.24

Table 6.6: Most relevant factors on the risk of TPD to hydrogen transmission and distribution pipelines. Relevance is calculated with respect to the model's variable with higher sensitivity derivative parameter S. The variables included in the table are all those whose S parameter was, at least, 10% of the higher-ranked variable.

Table 6.5(a) on hydrogen distribution pipelines shows that the three most relevant factors influencing the risk of TPDs are "Immediate ignition," "Pipe pressure," and "Excavation preventive measures." These factors match with the fact that hydrogen will present high ignition probabilities that, combined with potentially higher pressures to satisfy current energy needs by customers, can increase the frequency, variety (wider range of gas release characteristics), and consequences associated with ignited releases and jet fires. Therefore, improving TPD preventive measures in an excavation process will be crucial. Pipeline "Material" appears in

[3]The sensitivity derivative parameter S of a BN's node state "$V_i = v$" given a target node state "$Y_i = y_i$" means that a change c on the value of $Pr(V_i = v_i)$ will cause a change of $S \cdot c$ on the value of $Pr(Y_i = y_i)$.

the fourth place in terms of relevance. Distribution systems are mainly composed of plastic pipelines, which have a low resistance to punctures. As such, "Material" as a relevant factor highlights the need for materials with higher resistance to punctures but that maintain the benefits of current plastics pipelines. For instance, a material alternative could be fiber-reinforced composite pipelines for distribution applications, which have been shown to have better puncture resistance properties compared to polyethylene [212].

Table 6.5(b) on hydrogen transmission pipelines shows that the three most relevant factors influencing the risk of TPDs are "Excavation preventive measures," "H_2 embrittlement," and "Puncture force." This result shows that, as expected, embrittlement will be a major factor to consider in the integrity management and risk assessment of transmission pipelines, as it can have uncertain effects on a pipeline's resistance to damage. Moreover, there could be a need to change the current culture and practices involved in the excavation processes. For instance, horizontal directional drilling (HDD) is gaining wide popularity among excavators; however, it can exert two times more force on a pipeline compared to an excavator [122]. Combined with the uncertainties associated with H_2 embrittlement, HDD could become a high-risk practice not worth doing compared to other excavation alternatives. Additionally, "Pipe pressure" and "Immediate ignition" are tied for fourth place. These results highlight that hydrogen embrittlement, combined with higher operating pressures and ignition probabilities, can cause an increase in the frequency of ignited releases that can affect the public perception of hydrogen.

For distribution and transmission systems, relevant factors in places 5-7 show that the integrity management of hydrogen pipelines could benefit from a good understanding of the puncture forces that pipelines can be subject to by excavation equipment and the resistance they can have towards puncturing.

6.7 Discussion

A TPD risk model, SHyTERP, was proposed in this work to assess the risk of TPD to both natural gas and hydrogen pipelines. The proposed model incorporated 73 variables involved in a third-party excavation process, 110 dependencies, and the uncertainties relevant to the risk of punctures, ignited releases, and thermal harm from jet fires (see D.1). A probabilistic simulation approach based on causal BNs was used to parameterize SHyTERP and assess the total risk of TPD to natural gas and hydrogen pipelines. SHyTERP was validated against PHMSA historical data, and the average risk results for hydrogen pipelines align well with current literature on the topic (e.g., ignited releases were found to be more frequent on hydrogen pipelines, and hydrogen embrittlement was found to be challenging to the puncture resistance of steel transmission pipes. See Tables 6.3 and 6.4). However, employing a causal model and probabilistic simulations generated further insights beyond average total risk results. The use of probabilistic simulations showed that the lower percentiles obtained for the probability of puncture, ignition, and thermal harm, are consistently higher for hydrogen pipelines compared to natural gas (see in Figure 6.5). This result is highly relevant for low-frequency events such as injuries and fatalities, as a small increase in yearly rates can cause a negative public reaction toward the acceptance of hydrogen transportation through pipelines.

The risk of TPD to distribution pipelines transporting hydrogen was found to be dominated by the probability of ignited releases (7.6 times more frequent than natural gas. See Table 6.3). Study 1 showed that hydrogen distribution pipelines are expected to have a similar probability of getting punctured and causing a fatality compared to natural gas pipelines. The similarity in puncture probability between gases can be explained by the high proportion of plastic pipelines in distribution systems in the U.S., which are assumed to be unaffected by hydrogen degradation. For steel distribution pipelines, Study 2 showed that hydrogen embrittlement could cause a significant increase in the probability of punctures. Additionally, no major differences in puncture probability were found between hydrogen and natural gas distribution systems with a 0-30% share of steel pipes (see Figure 6.6(a)). However, as expected, systems with higher proportions of steel show additional protection from third-party punctures. Plastic, on the other hand, has become the standard material used in distribution systems. Although plastic pipelines are less resistant to punctures than steel pipes, utilities have defined that the benefits of plastic in terms of installation costs and overall safety (e.g., no corrosion) outweigh the risk of additional punctures in excavations. However, in a hydrogen setting, these additional punctures in excavation activities can cause a significant increase in the frequency (and variety) of ignited releases and jet fires (see Study 4 results in Section 6.6.4). Hydrogen is more likely to ignite than natural gas, and plastic pipelines can amplify the probability of immediate ignition due to static discharge.

The total risk of TPD to transmission pipelines was found to be consistently higher for hydrogen transportation compared to natural gas (see Table 6.4). Higher puncture and ignition probability in a hydrogen setting, as well as a higher operational pressure, can cause significant risks of burns and fatalities from jet fire events (see Table 6.5(b) from Study 4). As shown in Study 1 (see Section 6.6.1), this result can be explained by a potential increase in the frequency of punctures caused by hydrogen embrittlement. In a hydrogen setting, excavation equipment could puncture a wider variety of pipelines, resulting in a wider variety of gas releases in terms of release characteristics (e.g., release size and pressure). Combined with hydrogen's higher probability of ignition in high-pressure systems than natural gas (see Table 6.1), higher thermal harm frequencies can be expected from jet fires. By analyzing past natural gas transmission pipeline incidents in the U.S., Study 3 reveals that transporting hydrogen could have significantly increased their average and lower probability bounds of puncture and ignition probabilities. That is, the risk of puncture and ignition caused by TPD could have been consistently higher in a hydrogen setting. The results in this work show that hydrogen injection to existing transmission pipelines in the U.S. should be accompanied by case-specific QRAs demonstrating that hydrogen transportation complies with tolerable risk limits, as the risk of TPD in hydrogen transmission systems could increase significantly (in both frequency and consequences) compared to natural gas pipelines. In addition, Study 4 showed that excavation practices in areas with underground transmission pipelines might need to be re-evaluated. The puncture force exerted on potentially less resistant steel transmission pipelines could cause a significant increase in the risk of TPD. As such, popular excavation methods that exert high puncture forces, such as excavators and HDD, could become too risky in certain areas if hydrogen is transported in underground steel pipelines.

6.8 Concluding Remarks

TPD is one of the major threats to natural gas pipelines today, and it could become a major challenge to the safe transportation of hydrogen through existing natural gas infrastructure. This work documents a first-of-kind model and assessment of the risks posed by TPD to natural gas pipelines transporting hydrogen, specifically on the risks posed by punctures, immediate ignitions, and jet fires on both distribution and transmission systems. The objective of this work was to use a QRA approach based on the causal reasoning methods established in chapters 3 and 5 to assess the risk of TPD to hydrogen transportation through existing natural gas pipelines and provide insights and recommendations for future actions/policies to handle risks.

The results of this chapter show the applicability of the ladder of causation to generate a comprehensive set of causal insights necessary to tackle complex challenges such as hydrogen transportation through natural gas pipelines. Traditional QRA methodologies have been employed in recent years to address this challenge. However, insights from these past works have been limited to reducing the risk of TPD to a single expected value and comparing it to natural gas. The work in this chapter shows that several risk-reduction insights can be overlooked with traditional QRA methods and that an extended adoption of causal reasoning with a modern

BN modeling approach can extend insights to enable the safe deployment of this crucial new fuel. In general, causal reasoning enabled us to make the following contributions to the risk assessment of future hydrogen pipelines:

- Found, through associative reasoning, that hydrogen can consistently increase the risk of TPD and identified its drivers in both distribution and transmission pipelines.

- Found, through intervention reasoning, that hydrogen can pose a major challenge in plastic distribution pipes by significantly increasing the frequency of ignited releases. However, no major changes in TPD total risk are expected between different proportions of steel and plastic pipes in distribution systems.

- Found, through counterfactual reasoning, that past incidents in transmission pipelines could have been consistently more probable if hydrogen had been transported through the pipes instead of natural gas.

- Created a comprehensive set of recommendations based on causal insights to reduce the risk of TPD to future hydrogen pipelines in the U.S.

Based on the causal insights obtained in this work and discussed in Section 6.7, the following recommendations are made for controlling the risk of TPD to future hydrogen pipelines:

- Future risk assessments on hydrogen pipelines should consider improved causal models (such as BNs) and probabilistic simulation to expand the variety and depth of causal insights into a pipeline's risk in a hydrogen setting.

- Future actions/policies on pipelines should not only focus on decreasing the average total risk posed by hydrogen transportation but should also ensure a small difference between the most (and least) likely risk scenarios for both gases.

- Future installations of hydrogen pipelines should evaluate the trade-offs between installing new plastic pipes or repairing and installing new steel pipes in distribution systems due to a potential increase in the frequency of ignited releases. This evaluation should also consider how the public perception of hydrogen safety can be affected by a potential increase in ignited releases TPDs, as excavation incidents are highly visible to the public.

- Plastics with additional resistance to puncture should be considered for future distribution pipelines, such as fiber-reinforced composites.

- Hydrogen injection to existing transmission pipelines should be accompanied by case-specific QRAs demonstrating that the pipeline complies with tolerable risk limits, as the risk of TPD in hydrogen transmission systems could increase significantly.

- We recommend a wider adoption of non-destructive vacuum excavation methods and more strict damage prevention practices in excavation sites near transmission pipelines.

The results presented in this chapter enabled us to achieve this dissertation's third and last objective and, therefore, demonstrated the proposed intervention and counterfactual reasoning methods on a comprehensive QRA of a complex engineering system or process. By doing so, this work showed how BNs could serve as a unified modeling approach to exhaustively address causal reasoning in QRAs and generate insights on the three rungs of the ladder of causality.

Chapter 7: Conclusions

7.1 Summary of Key Results and Contributions

This research expands upon the current causal reasoning capabilities of modern QRAs leveraging improved causal models such as BNs. Specifically, this work exploits the causal reasoning capabilities of BNs to support intervention and counterfactual reasoning in the QRA of complex engineering systems and processes. The objectives were to (1) establish the mathematical background and methods for enabling intervention reasoning in QRA, (2) establish the mathematical background and methods for enabling counterfactual reasoning in QRA, and (3) demonstrate these methods in a comprehensive QRA study. We approached these objectives as 3 research tasks, which results and contributions are further described later in this chapter. Overall, the results of this dissertation led to the following three major technical contributions to the risk analysis and reliability engineering field:

1. **Major contribution 1 (MC1)**: Established a set of methods and frameworks that enable using BN models as a unified approach for obtaining causal insights in the QRA of complex engineering systems or processes

2. **Major contribution 2 (MC2)**: Developed two comprehensive risk models,

based on BNs, for analyzing the risk of TPD to natural gas and hydrogen pipelines

3. **Major contribution 3 (MC3)**: Demonstrated the applicability of causal reasoning with BNs for two real-world case studies on the risk assessment of third-party excavation damage to natural gas and hydrogen pipelines

Figure 7.1 summarizes the technical contributions of this dissertation and shows how each of the major contributions described above enabled us to achieve the three research objectives of this work. Each research task conducted in this work provided key results that contributed to achieving these objectives and are summarized in the following sections.

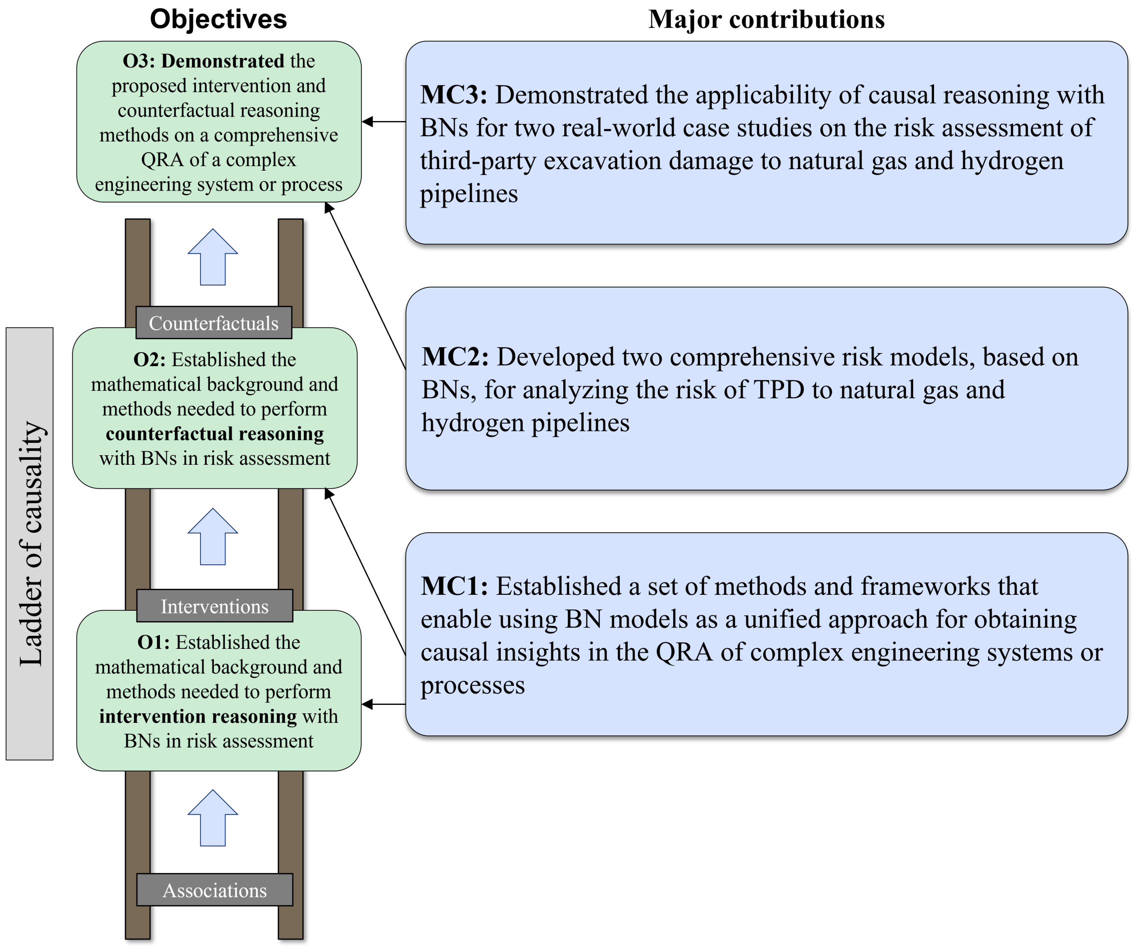


Figure 7.1: Three major technical contributions (MC) of this dissertation.

7.1.1 O1: Established the mathematical background and methods needed to perform intervention reasoning with BNs in risk assessment

Achieving the first objective of this dissertation, establishing the mathematical background and methods needed to perform intervention reasoning in modern QRA, required the results of RT1.

As shown in Chapter 3, RT1 resulted in a set of mathematical methods for performing intervention reasoning with BNs for QRA and, therefore, moved causal insights up to the second rung of the ladder of causality. These methods were based on the *truncated factorization formula* for fully-parameterized BNs (see Equation 3.2) and the *backdoor-formula* for non-fully parameterized models (see Equation 3.6). However, as described in Chapter 3, intervention reasoning cannot be performed on every BN model. The main requirement of a BN to permit the application of the equations shown before is that the model should include all confounding paths (back-door paths) between the intervened variable and the outcome of interest. To support analysts in complying with these requirements, a set of criteria (back-door criterion in Definition 4) and rules (do-calculus rules in Theorem 1) were provided to identify the variables in a BN that can bias the calculation of causal effects through intervention reasoning. To demonstrate the applicability of interventions in QRAs, the scope of this research activity was narrowed to using interventions in BNs to calculate the effect of decisions on a system's risk. A three-step framework and a set of three causal metrics were developed toward this goal (see Section 3.3). The proposed framework was built to comply with all requirements to perform intervention reasoning with BNs in a QRA. Additionally, the three causal metrics were created to inform decision-making by enabling the isolation (IE and CDE metrics in Equations 3.8 and 3.10, respectively) and comparison (CCE metric in Equation 3.9) of the causal effect that an intervention on a variable (or set) can have on the risk of a system.

Further, we developed two examples that demonstrate that current associative reasoning methods used in BN-based QRAs cannot support causal insights based on interventions. In the first example (see Section 3.4.4.1), we showed how the estimated causal effect of an intervention could be biased if no adjustment is performed to account for confounding paths between the intervened variable and the outcome of interest. Moreover, in this example, we were able to show that associations can potentially provide erroneous results due to the possibility of Simpson's reversal in the data used to parameterize a BN model. In the second example (see Section 3.4.4.2), we showed how insights provided by importance measures, a well-established and popular method in traditional QRAs, cannot be easily translated to a BN setting. The lack of awareness of intervention reasoning led previous research to apply associative reasoning methods for importance measure analysis with BNs. However, we found that this practice can lead to underestimated or even invalid values for an importance measure metric if a BN model for QRA does not comply with the requirements for intervention reasoning established in RT1.

The results of RT1 presented in Chapter 3 fill the gap of enabling analysts to reason about causality and interventions in modern BN-based QRAs for applications concerning changing conditions, such as the occurrence of an event or the implementation of an action or policy.

7.1.2 O2: Established the mathematical background and methods needed to perform counterfactual reasoning with BNs in risk assessment

Achieving the second objective of this dissertation, establishing the mathematical background and methods needed to perform counterfactual reasoning in modern QRA, required the results presented in RTs 1 and 2.

As shown in Chapter 5, the intervention reasoning methods established in RT1 were fundamental to the methods developed in RT2.2 by enabling the twin network representation of counterfactuals with BNs (see Section 5.2.3). The intervention techniques of Chapter 3 are required to model a "factual world" and a "possible world" in a BN model, as the latter has to be manipulated in order to represent an alternative scenario of a past event. The methods and requirements to enable counterfactual reasoning in BNs were established in Section 5.2 and show that a comprehensive causal BN model of a system or process is needed to enable counterfactual reasoning. This requirement is based on the fact that counterfactual modeling requires the identification and evidence updating of the common background conditions between an event and a counterfactual scenario. QRA models were found to be an excellent candidate to comply with this requirement, as they are built towards thoroughly representing the scenarios, probabilities, and consequences associated with the risk of a system or process. The BaNTERA model result of RT2.2 in Chapter 4 was used to demonstrate the established counterfactual methods. Although scarce, we found past QRAs attempting to perform counterfactual reasoning; however, these attempts did not assess the influence of the background conditions in their analysis. Ultimately, these past attempts were found to be answering an intervention query rather than a counterfactual query, which provides additional evidence for the need to understand the role of causality and the distinct types of reasoning in QRA (see Section 2.2.3).

The results of RT2 presented in Chapter 5 fill the gap of enabling analysts to reason about causality and counterfactuals in modern BN-based QRAs and provide causal insights on the third rung of the ladder of causality. To do this, we proposed the "possible world" approach to counterfactuals and developed an accompanying methodological framework to answer counterfactual queries in a QRA (see Section 5.3). This framework was focused on providing causal insights to explain past incidents and accidents and enable the assessment of multiple scenarios to provide deeper causal insights into past events.

7.1.3 O3: Demonstrated the proposed intervention and counterfactual reasoning methods on a comprehensive QRA of a complex engineering system or process

The last objective achieved in this dissertation was the demonstration of the established intervention and counterfactual methods on the QRA of a complex engineering system or process. The problem of TPD in natural gas pipelines was used as a case study throughout this research, which was expanded to hydrogen pipelines in RT3. An illustrative case study on the effectiveness of potholing in TPD prevention was performed in RT1 to demonstrate the application of the intervention reasoning methods and framework established in this dissertation (see Section 3.4). While the BN model used in RT1 was sufficient for illustrating intervention reasoning, it lacked the comprehensive causal structure required to perform counterfactual reasoning. RT2.1 addressed this problem by developing BaNTERA, a comprehensive Bayesian network for third-party excavation risk assessment. BaNTERA was the result of an extensive project aimed at providing an exhaustive causal structure and quantitative model to assess the risk of TPD to natural gas distribution and transmission pipelines (see Sections 4.5 and 4.6). The BaNTERA model was a successful endeavor and was used, first, to demonstrate the use of associative reasoning in QRAs by analyzing several case studies to inform the risk management of TPDs in the U.S. (see Section 4.8), and second, to develop a cloud-based GIS software to support a utility partner on the integrity management of their transmission pipelines (see Section 4.9). The BaNTERA model has a comprehensive causal structure suitable for counterfactual reasoning. As such, the counterfactual methods established in RT2.2 were applied with BaNTERA to provide causal insights on the 2018 Sun Prairie gas explosion (see Section 5.4). A previous study on the Sun Prairie explosion identified a lack of notification as the event's root cause. However, we demonstrated that "possible world" counterfactuals could extend this insight by showing that, in such a scenario, posterior preventive measures (e.g., performing test holes in the excavation site) would not have significantly reduced the likelihood of a TPD (see Section 4.10). This real-world case study showed the capability of BNs to reason about counterfactuals and provided additional insights from an accident's evidence compared to traditional methods such as "but-for" event analysis.

We fully demonstrated the applicability of causal reasoning with BNs for QRA by applying the methods established in RT1 and RT2 to RT3, the problem of TPD in natural gas pipelines transporting hydrogen. This was a complex challenge as hydrogen transportation through existing gas infrastructure is an open problem for achieving the U.S. decarbonization goals by 2050. RT3 resulted in the development of a novel algorithm and BN model, SHyTERP, to assess the risk of TPD to hydrogen pipelines. Different from recent QRAs on hydrogen pipelines, the SHyTERP model developed in this research activity is the first to account for the effect of hydrogen on both the frequency and consequence aspects of risk (see Section 6.4). Additionally, SHyTERP was based on a hybrid BN approach, which enabled us to obtain insights not only into expected risks values but also into simulating uncertainties surrounding the model's predictions (see Sections 6.5.2 and 6.6.1). RT3 showed how BNs could serve as a unified modeling approach to exhaustively address causal reasoning in QRAs and generate insights on the three rungs of the ladder of causality.

The three rungs of the ladder of causality were used in RT3 and led us to key findings on the risk of TPD to hydrogen pipelines. First, associative reasoning (rung 1 of the ladder) enabled us to conclude that hydrogen can consistently increase the risk of TPD, as hydrogen can increase the frequency of pipeline punctures and ignited releases in both distribution and transmission systems (see Section 6.6.1). Additionally, associative reasoning was used to perform a sensitivity analysis on ignited releases that can produce high radiative heat fluxes, which led us to identify critical variables that need to be controlled to ensure the safe transportation of hydrogen through existing pipelines. In distribution lines, these variables were immediate ignitions, increased pressure, TPD preventive measures, and pipe material. In transmission lines, the most important variables were TPD preventive measures, hydrogen embrittlement, and puncture force (see Section 6.6.4). Second, intervention reasoning (rung 2 of the ladder) was applied to SHyTERP to assess the causal effect of hydrogen on the risk of distribution systems composed of different proportions of steel and plastic pipes. This analysis found that no major changes are expected to be observed in punctures and fatalities when comparing hydrogen to natural gas pipelines, independently of the materials present in a pipeline system. However, ignited releases are expected to be significantly more frequent, especially in systems with high proportions of plastic pipes (see Section 6.6.2). Third, counterfactual reasoning (rung 3 of the ladder) was used to obtain insights into what could have happened in known past incidents in transmission systems if hydrogen had been transported through the pipelines instead of natural gas. This analysis found that the upper bound probability of puncture and ignition for past incidents in natural gas pipelines is close to the lower bound probability of these consequences if hydrogen is transported through the same pipelines involved in the studied incidents (see Section 6.6.3). Lastly, these causal insights were used to create recommendations to reduce the risk of TPD to future hydrogen pipelines in the U.S., including the use of composite materials for additional puncture protection in plastic distribution lines and evaluating the trade-offs of using high-force excavation methods near transmission pipelines given the potential increase in risk in a hydrogen setting (see

Section 6.8).

7.2 Work Products

7.2.1 Publications

7.2.1.1 Journal Papers

Included in this dissertation:

- Andres Ruiz-Tagle and Katrina M. Groth. "Comparing the risk of third-party excavation damage between natural gas and hydrogen pipelines" International Journal of Hydrogen Energy (anticipated).

- Andres Ruiz-Tagle, Enrique Lopez-Droguett, and Katrina M. Groth. "A novel probabilistic approach to counterfactual reasoning in system safety." Reliability Engineering & System Safety 228 (2022): 108785.

- Andres Ruiz-Tagle, Austin D. Lewis, Colin A. Schell, Ernest Lever, and Katrina M. Groth. "BaNTERA: A Bayesian Network for Third-Party Excavation Risk Assessment." Reliability Engineering & System Safety 223 (2022): 108507.

- Andres Ruiz-Tagle, Enrique Lopez Droguett, and Katrina M. Groth. "Exploiting the capabilities of Bayesian networks for engineering risk assessment: Causal reasoning through interventions." Risk Analysis 42, no. 6 (2022): 1306-1324.

Not included in this dissertation but influenced by the causal methods established in this research:

- Ahmad Al-Douri, Andres Ruiz-Tagle, and Katrina M. Groth "A Quantitative Risk Assessment of Hydrogen Fuel Cell Forklifts" International Journal of Hydrogen Energy (2022). Submitted: September 28, 2022. Minor comments: December 12, 2022.

- Mohammad Pishahang, Andres Ruiz-Tagle, Enrique Lopez Droguett, Marilia Ramos, and Ali Mosleh. "WiSE: A Bayesian Agent-Based Framework for Wildfire Safe Egress" In preparation.

- Ramin Moradi, Andres Ruiz-Tagle Palazuelos, Enrique Lopez Droguett, and Katrina M. Groth. "Toward a Framework for Risk Monitoring of Complex Engineering Systems with Online Operational Data: A Deep Learning-based Solution." Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability (2020): 1748006X221079964.

7.2.1.2 Conference Papers

- Colin A. Schell, Andres Ruiz-Tagle, Austin D. Lewis, and Katrina M. Groth. "Construction and Verification of a Bayesian Network for Third-Party Excavation Risk Assessment (BaNTERA)." Probabilistic Safety Assessment and Management PSAM 16, June 26-July 1, 2022, Honolulu, Hawaii.

- Mohammad Pishahang, Andres Ruiz-Tagle, Enrique López Droguett, Marilia

Ramos, and Ali Mosleh. "Integrating Human Behavior Modeling into a Probabilistic Wildfire Egress Planning Framework." Probabilistic Safety Assessment and Management PSAM 16, June 26-July 1, 2022, Honolulu, Hawaii.

- Marilia Ramos, Mohammad Pishahang, Andres Ruiz-Tagle, Enrique López Droguett, and Ali Mosleh. "WISE: A Probabilistic Wildfire Safe Egress Planning Framework." Probabilistic Safety Assessment and Management PSAM 16, June 26-July 1, 2022, Honolulu, Hawaii.

- Andres Ruiz-Tagle, Vincent P. Paglioni, Enrique Lopez Droguett, and Katrina M. Groth. "A Framework to Extrapolate and Evaluate Human Reliability Causal Models from Event Report Narratives." ANS International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA), November 7-12, 2021, Columbus, Ohio.

- Ramin Moradi, Andres Ruiz-Tagle Palazuelos, Katrina Groth, and Enrique Lopez Droguett. "Towards a Framework for Risk Monitoring of Complex Engineering Systems with Online Operation Data: A Deep Learning Based Solution." Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference (ESREL/PSAM15), November 1-5, 2020, Venice.

7.2.1.3 Technical Reports

- Gas Technology Institute "Final Report: Data Collection, Normalization, and Integration Methods to Enhance Risk Assessment Tools for Decision-

Making." Pipeline Hazardous Material and Safety Administration. (2022) 693JK31910002POTA. URL: https://primis.phmsa.dot.gov/matrix/PrjHome.rdm?prj=846

7.2.1.4 Presentations

- Andres Ruiz-Tagle and Katrina M. Groth "A Threat to Natural Gas Pipelines Transporting Hydrogen: Assessing the Risk of Third-party Excavation Damage." Technical Presentation to GTI Energy and Utility Partner, Chicago, July 27, 2022.

- Andres Ruiz-Tagle, Enrique Lopez Droguett, and Katrina M. Groth. "Addressing Causal Reasoning for Quantitative Risk Assessment of Engineering Systems: A Bayesian Network Approach." Society of Risk Analysis Annual Meeting, Virtual, December 5-10, 2021.

- Andres Ruiz-Tagle, Katrina M. Groth, Enrique Lopez-Droguett, Ernest Lever. "Round table: Transferring Modern Risk Assessment Models from Academia and Research Institutions to the Industry." Society of Risk Analysis Latin America, Virtual, December 2, 2021.

- Andres Ruiz-Tagle, Enrique Lopez Droguett, and Katrina M. Groth. "Graduate Research Lighting Round: Causal Reasoning with Bayesian Networks for the Quantitative Risk Assessment of Engineering Systems." ANS International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA), Columbus, November 11, 2021.

- Andres Ruiz-Tagle, Vincent P. Paglioni, Enrique Lopez Droguett, and Katrina M. Groth. "A Framework to Extrapolate and Evaluate Human Reliability Causal Models from Event Report Narratives." ANS International Topical Meeting on Probabilistic Safety Assessment and Analysis (PSA), Columbus, November 7-12, 2021.

- Andres Ruiz-Tagle, Austin Lewis, and Katrina Groth. "BaNTERA Model Demonstration." Technology Transfer Meeting, Virtual, March 9, 2021.

- Ramin Moradi, Andres Ruiz-Tagle Palazuelos, Katrina Groth, and Enrique Lopez Droguett. "Towards a Framework for Risk Monitoring of Complex Engineering Systems with Online Operation Data: A Deep Learning Based Solution." Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference (ESREL/PSAM15), Venice, November 1-5, 2020.

7.2.2 Models and Codes

This dissertation resulted in 11 BN models and 12 programming codes toward achieving the three objectives of this research. The majority of the models and codes were developed using BayesFusion GeNIe software [1], SMILE engine [213], and the Python programming language [214]. All of these work products are available to use for research in the Systems Risk and Reliability Analysis (SyRRA) lab. This material is available upon request to Andres Ruiz-Tagle at aruiztag@umd.edu. The models and codes for each research activity are listed as follows:

Research Task 1

BN models

- Interventions_BN_CaseStudy-2020.xdsl: GeNIe BN model used in the case study

Codes

- Interventions_CPT_CaseStudy-2020.py: Python script for parameterizing the BN model used in the case study
- Interventions_doCalcIdentification_CaseStudy-2020.R: R script [215] for identifying adjustable backdoor paths in the BN model used in the case study

Research Task 2.1

BN models

- BaNTERA-2021.xdsl: GeNIe BN model for BaNTERA
- model.json: AgenaRisk [175] model for the TPD risk management GIS software

Codes

- BaNTERA_DataProcessing_2021.py: Python script for processing input data to BaNTERA
- BaNTERA_Parameterization_2021.py: Python script for parameterizing BaNTERA
- BaNTERA_Results_2021.py: Python script for simulating target nodes of BaNTERA and generating results
- GIS application API and scripts: TIMP_DataProcessing.py; TIMP_ResultsProcessing.py; index.js

Research Task 2.2

BN models

- prior_BaNTERA-2021.xdsl: GeNIe BN prior model based on BaNTERA for the case study
- sA_TWIN_BaNTERA-2021.xdsl: GeNIe BN twin network mode for case study scenario A
- sB_TWIN_BaNTERA-2021.xdsl: GeNIe BN twin network model for case study scenario B
- sC_TWIN_BaNTERA-2021.xdsl: GeNIe BN twin network model for case study scenario C
- sD_TWIN_BaNTERA-2021.xdsl: GeNIe BN twin network model for case study scenario D

Research Task 3

BN models

- HyTPD_H2Distr-2022.xdsl: GeNIe BN model for the extended version of BaNTERA to hydrogen transportation through distribution pipelines
- HyTPD_H2Trans-2022.xdsl: GeNIe BN model for the extended version of BaNTERA to hydrogen transportation through transmission pipelines
- HyTPD_NGDistr-2022.xdsl: GeNIe BN model for the extended version of BaNTERA to natural gas transportation through distribution pipelines
- HyTPD_NGTrans-2022.xdsl: GeNIe BN model for the extended version of BaNTERA to natural gas transportation through transmission pipelines

Codes

- HeatFlux.py: Python script for the HeatFlux class used for simulating gas release, ignition, heat radiation from jet flames, and thermal harm
- HyTPD_Param-2022.py: Python script for parameterizing the extended version of BaNTERA to hydrogen pipelines
- HyTPD_Results-2022.py: Python script for simulating target nodes of the extended version of BaNTERA to hydrogen pipelines

7.3 Opportunities for Future Work

This dissertation is a first step toward introducing causal reasoning capabilities to modern QRA based on improved causal models based on BNs. The results of this dissertation can be divided into two major areas: causal reasoning methods and QRA models for assessing the risk of TPD to underground pipelines. There are several areas to expand and improve the work in these two areas.

7.3.1 Causal Reasoning Methods

The following points present future directions for expanding intervention and counterfactual reasoning in QRA.

Intervention Reasoning

- Intervention reasoning was studied on static BN models of a system or process. Further research should explore the use of interventions to obtain causal insights into dynamic systems. To do this, dynamic BNs [216] should be used as they accept the intervention reasoning methods established in this work (dynamic BNs are unwrapped static BNs with temporal dependencies). By doing so, QRA models will be able to reason about the temporal effect of interventions in a system or process behavior.

- Causal reasoning in BN-based QRA models is prone to multiple biases that can affect the accuracy of causal insights based on interventions in a system. This dissertation focused on using Pearl's *do*-operator to control for confounding bias. However, other biases, such as selection and transportability bias [45], can present a challenge. Bareinboim et al. [45] and Hund et al. [54] have demonstrated the use of the *do*-operator to control for multiple types of biases with causal graphical models, which should be explored in future work to improve the accuracy of causal insights with modern QRA models.

- The intervention reasoning methods established in this work should be further investigated for expanding the identification of causal explanations (e.g., identifying risk triggers [217]) on the risk of a system. Causal graphical models have been used in addition to the *do*-operator to search for decision-making opportunities [218], and parallels can be drawn for finding the major contributors to scenarios of interest in a risk assessment.

- Intervention reasoning in modern QRA can be affected by the quality of data used to perform it. However, methods such as the "backdoor" and "frontdoor" [18] criterion can provide with means of focusing data collection activities on ensuring accurate causal insights on a system. These activities, for example, can be guided by collecting data on an optimized set of variables identified through the backdoor criterion to control for confounding bias when providing causal insights in a QRA. These methods should be further investigated to support causal reasoning activities in modern QRA.

Counterfactual Reasoning

- Future research in QRA should focus on developing algorithms for automatically identifying relevant features (and their combination) that explain past incidents and events through counterfactual queries to a model. Counterfactual reasoning methods are becoming prevalent in other fields to offer causal explanations in both big and sparse-data settings [57]. This process is called "feature attribution," and QRA can benefit from cross-pollination. Attribution can be studied by evaluating the probability of necessity and sufficiency of a variable on the outcome of an event, and methods have been created to evaluate these probabilities [18]. Particular attention should be given to downward counterfactual reasoning as answering queries of the type "how the event could have been worse" has the potential of searching for variables that are critical to controlling for the risk management of a system.

- Evidence on the occurrence of past events is not always clear and succinct and can be explained by several competing narratives. Future work should be done on developing methodologies that incorporate competing information in a counterfactual model of a system in order to evaluate the risk of an event and provide causal insights. For instance, an event investigation and counterfactual model might represent a single narrative for the event development. However, conflicting hypotheses might require additional variables in a "possible world" model, and there should be a means of comparing potential explanations of an event [196]. The "possible world" methodology to counterfactual reasoning in QRA can serve as a baseline framework to model, evaluate, and compare competing narratives by evaluating their likelihood given an event's evidence.

- The proposed "possible world" approach to counterfactual reasoning in QRA should be expanded to include the evaluation of uncertainty on counterfactual insights. Many scenarios could develop given the conditions of a past event; however, analysts only have evidence of an event's "factual" outcomes. Stochastic simulation should be further explored to simulate a comprehensive range of potential developments of an event and their uncertainty, as this will provide further insight into comparing competing counterfactual hypotheses. Uncertainty in counterfactual insights was explored in Chapter 6. However, further work should be done on standardizing the analysis of counterfactual insights generated through simulation.

Besides expanding intervention and counterfactual reasoning methods, future work should also consider the applicability of causal reasoning with BNs in a big data setting. The CPTs and CPFs in BNs suffer from exponential growth when additional parent nodes are included in a model. Therefore, in big data applications, the resulting combinatorial explosion can make the data requirements for parameterizing BNs, as well as the computing of probabilistic inference algorithms, considerably expensive. Variational inference [219] applied to probabilistic graphical models is a promising solution to these challenges. Variational inference has the potential to allow fully-continuous BNs and to relax the discretization requirements that lead to the exponential growth of CPTs. Furthermore, variational inference methods are based on approximating posterior probability distributions as an optimization problem without requiring sampling, which has the potential to significantly decrease the cost of inference algorithms. As such, given that the methods established in this work hold for all causal BNs independently of the inference algorithm used for computation, variational inference could allow causal reasoning on highly complex systems under a big data setting.

7.3.2 QRA Models for TPD of Underground Gas Pipelines

The following points present future directions for expanding the models developed in this work for assessing the risk of TPD to natural gas and hydrogen pipelines.

BaNTERA Model

- More data should be collected on pipeline characteristics to improve BaNTERA's parameterization and predictive accuracy, particularly on excavations without pipeline incidents. Currently, organizations like PHMSA and CGA only collect incident data. Pipeline statistics for risk modeling are mainly based on as-designed characteristics. However, it is well known that pipelines operate at conditions that are well outside initial design parameters [220]. Future work on models like BaNTERA should make use of novel data streams to integrate statistics on on-site operational pipelines. We recommend, as a starting point, using an opportunistic sampling approach from TPD incidents as described in Sherwin, Lever, and Brandt [167]. This work shows that TPDs are a quasi-random process that could provide as-operated pipeline data on TPD locations, such as pipe pressures, materials, age, and health state, among others. This data can complement as-designed pipe characteristics to improve the accuracy of future versions of BaNTERA.

- More information should be collected on human factors affecting TPD's root causes of notification and excavation issues. As described above, TPD data mainly comes from post-incident databases. As such, data from TPD risk-influencing factors is restricted to excavation instances that led to pipe damage. Millions of near-pipe excavations are performed every year in the U.S. without consequences. However, information on these excavations can only be obtained from notification centers such as 811. Considering that more than 30% of TPDs occur without notification, it is highly important for BaNTERA to include information on non-notified and non-incident data. We recommend the use of survey and expert-elicited data to complement statistics obtained from notified excavations. A first effort was made in BaNTERA's development by including survey data on excavators' awareness of notification practices [23]. However, additional efforts should be made to collect survey data on notification and excavation practices among excavators.

- BaNTERA only included pipe hits and puncture damage as target variables for assessing the probability of TPD. Punctures are the least likely outcome of an excavation activity (but with the highest potential consequences [198]). As such, assessing punctures only provide a lower bound for the probability of TPD. Further efforts should be made to develop and integrate parametric models describing dents and gouges, which are significantly more frequent pipeline failure modes in excavation activities. Although the immediate consequences of dents and gouges can be negligible, they can accelerate time-based failure mechanisms such as fatigue and corrosion. Differently from punctures, dents are gouges are usually not reported to authorities, creating latent risks that could be catastrophic [178]. Future versions of BaNTERA should integrate dents and gouges in its formulation to characterize better the risk of TPD.

Hydrogen TPD Risk Model

- The accuracy of jet fire scenario modeling should be improved in future work. The Hydrogen TPD risk model of Chapter 6 did not consider wind and crater formation, which could be highly relevant in releases from transmission and large distribution pipelines. Craters can change the release size conditions of punctured pipelines, as well as enable potential domino effects due to impingement on uncovered surrounding pipelines. Winds can significantly change a population's exposure to risk depending on their position with respect to the direction of the wind. Commercial software such as DNV's Phast has the capability of modeling the effect of wind and craters in gas releases from underground pipelines [128], and their recently released Python API would allow us to integrate it into a simulation routine for assessing excavation scenarios. Future work should consider extending our hydrogen TPD model to incorporate simulations with the Phast software and simulator.

- The model of Chapter 6 was a preliminary risk model to assess the risk of TPD in natural gas and hydrogen pipelines. As such, the analysis of this work focused on generic insights into transmission and distribution pipelines. Future work should focus on more granular insights into the individual and societal risk of ignited releases in specific pipeline systems. Operational conditions of pipelines are highly heterogeneous throughout the U.S., so a set of representative systems should be further analyzed in both distribution and transmission systems. Special considerations should be made on location classes, depth of pipeline cover, and the density of pipeline infrastructure. Risk profiles and transects on these selected pipelines and scenario configurations could provide a more precise assessment of the risk of TPD to future hydrogen pipelines.

- Assessing the risk of delayed ignitions in both confined and unconfined settings is highly relevant for hydrogen pipelines. Hydrogen releases' dispersion and flammable cloud formation behave very differently from natural gas releases. Therefore, future work should expand the TPD model proposed in Chapter 6 to incorporate delayed ignited releases, deflagration/detonation, and overpressure harm. Similarly to the previous future work recommendation, delayed ignition scenarios are recommended to be studied on a selected set of pipeline configurations. The reason for this is that overpressure harm is highly dependent on enclosure characteristics, so risks can change significantly between high-density and open areas.

7.4 Research Impact

A seminal paper by Aven and Zio [31] identified the question *How to understand causality in a risk context?* as a foundational issue in modern risk assessment and management. Causality is a key concept in risk assessments [19] and is crucial for obtaining insights into the causes or explanations for accidents and harm. Recent research for modernizing QRA has tackled multiple foundational issues identified in [31], including definitions of terminology and fundamental principles [35], the development of quality tests for risk analyses [221], and the development of methods for analyzing the risk of complex systems [9, 52]. However, research on causality for modern QRAs has received little attention [11, 30, 146, 180].

In this research, we provide a baseline for improving QRA by expanding the use of causality in risk analysis and propose the widely popular BN modeling approach as a unified model for causal reasoning in modern QRA. The introduction of Pearl's ladder of causality to the QRA community serves as a fundamental pillar for the research of causality, identifying the basic concepts and requirements to reason about causality in QRA. Causal insights on the risk of a system are commonly rooted in a specific rung of the ladder of causality. However, this research showed that risk analysts had been prone to errors in the formulation of causal insights due to an unclear understanding of the distinctions between associations, interventions, and counterfactuals (see Chapters 3 and 5). This research develops and provides the methods that risk and reliability analysts need to reason about causality in QRAs and provide new and transparent causal insights to inform risk management and decision-making.

In addition to establishing a baseline for the study of causality in modern QRA, this research demonstrates that causality is readily applicable within QRA's widely popular BN modeling approach. The methods and frameworks established in this work enable risk analysts to reason about causality and advance the research of complex challenges in modern QRA that require improved versions of causal analysis. Our research results are already being used to address novel problems in QRA, such as investigating triggering factors of marine geohazard events [217] and evaluating the effectiveness of repairs and planned maintenance activities on assets over their life cycle [222]. The results of this dissertation prove that the proposed intervention and counterfactual methods are readily available for addressing complex challenges regarding causality in the QRA of complex engineering systems and processes.

Appendix A: Appendix for Chapter 3

A.1 Conditional Probability Tables for the BN model of Chapter's 3 illustrative case study

The tables below present each node CPTs from the BN model used in the case study.

Property Owner [%]	19.4
Contractor [%]	74.8
Government Entity [%]	5.8

Table A.1: Z_1: Third-Party Excavator CPT

	Z_1: Third-Party Excavator		
	Property Owner	Contractor	Government Entity
Yes [%]	30.9	21.1	5.3
No [%]	69.1	78.9	94.7

Table A.2: Z_2: Potholing Performed? CPT

	Z_1: Third-Party Excavator		
	Property Owner	Contractor	Government Entity
Yes [%]	94.7	66.4	61.1
No [%]	5.3	33.6	38.9

Table A.3: Z_3: Other Excavation Best Practices Performed? CPT

	Z2: Potholing Performed?			
	Yes		No	
	Z3: Other Excavation Best Practices Performed?			
	Yes	No	Yes	No
Yes [%]	100.0	79.9	60.9	0.0
No [%]	0.0	20.1	39.1	100.0

Table A.4: Z_4: Sufficient Excavation Practices? CPT

Yes [%]	87.7
No [%]	12.3

Table A.5: Z_5: Sufficient L&M Practices? CPT

	Z_1: Third-Party Excavator		
	Property Owner	Contractor	Government Entity
Yes [%]	17.9	61.9	82.2
No [%]	82.1	38.1	17.8

Table A.6: Z_6: Sufficient Notification Practices? CPT

	Z_4: Sufficient Excavation Practices?							
	Yes				No			
	Z_5: Sufficient L&M Practices?							
	Yes		No		Yes		No	
	Z_6: Sufficient Notification Practices?							
	Yes	No	Yes	No	Yes	No	Yes	No
Yes [%]	100.0	55.2	94.2	42.3	95.9	59.5	90.0	0.0
No [%]	0.0	44.8	5.8	57.7	4.1	40.5	10.0	100.0

Table A.7: Z_7: Sufficient Preventive Measures? CPT

Appendix B: Appendix for Chapter 4

B.1 Hierarchical Taxonomy at the Third-Party Damage Site

1. Actors	**2. Objects**	**3. Environment**
1.1. Third-Party	2.1. Facility	3.1 Physical
1.1.1. On-Site	2.1.1. Type of Facility	3.1.1. Geography
1.1.1.1. Excavator Operator	2.1.1.1. Transmission/Distribution	3.1.2. Hazards
1.1.1.1.1. Training/Experience	2.1.2. State of Facility	3.2. Legal
1.1.1.1.2. Attitude Towards Pipeline	2.1.2.1. Previous Damage	3.2.1. Location
1.1.2. Organizational	2.1.2.2. Deteriorated Facility	3.2.2. Oversight
1.1.2.1. Excavation Policy	2.1.2.3. Abandoned Facility	3.2.2.1. Enforcement Policy
1.1.2.2. Training Practices	2.1.3. Pipe Characteristics	3.2.3. Public
1.2. Facility Owner/Management	2.1.3.1. Pipe Wall Thickness	3.2.3.1. Awareness of Pipeline
1.2.1. On-Site	2.1.3.2. Pipe Material	3.2.3.2. Attitude Towards Pipeline
1.2.1.1. Locator Training/Experience	2.1.3.2.1. Ultimate Strength of Pipe Material	3.3. Excavation
1.2.1.2. Marker Training/Experience	2.1.4. Depth of Cover	3.3.1. Depth of Excavation
1.2.2. Organizational	2.1.5. Additional Facility Protection	3.3.2. Excavation Dynamic
1.2.2.1. Maintenance Policy	2.2. Excavation/Construction	3.3.2.1. Interactions
1.2.2.2. Training Practices	2.2.1. Excavator	3.3.3. Previous Excavation History at Site
1.2.2.3. Popularity	2.2.1.1. Excavator Tooth Length	
1.3. Notification Center	2.2.1.2. Excavator Tooth Width	
1.3.1. Quality of Responses	2.2.2. Notification Ticket	
1.3.2. Quality of Records	2.2.2.1. Dig Time	
	2.2.2.2. Dig Location	
	2.2.3. Locator Equipment	
	2.2.4. Facility Maps/Records	
	2.2.5. Excavation Monitoring Tools	

Table B.1: Comprehensive taxonomy of actors, objects and environments within a third-party excavation context

B.2 BaNTERA Nodes

Node ID	Node Name	Node Definition	Node Type	Discrete/ Continuous	Parameterization Method Used	Data Sources Used*
E1	Regulatory District	Regulatory District	Environment	Discrete	Elicitation - Relative count of events	C
E2	Area of Incident	Area of Incident	Environment	Discrete	Elicitation - Relative count of events	B
E3	Area Development	Area Development	Environment	Discrete	Elicitation - Relative count of events	B
E4	Excav. Depth (ft)	Depth of Excavation (ft)	Excavation Context	Continuous	Elicitation - Histogram	D
E5	Patrol Freq.	Patrol Frequency	Environment	Discrete	Elicitation - Relative count of events	D
EAC1	Excavator	Excavator Type	Third-Party Context	Discrete	Elicitation - Relative count of events	A
EAC2	Awareness	Notification Awareness	Third-Party Context	Discrete	Elicitation - Relative count of events	C
EAC3	Qualif. Locator	Qualified Locator	Third-Party Context	Discrete	Elicitation - Relative count of events	D
EAC4	Loc. Tech.	Locating Technology	Third-Party Context	Discrete	Elicitation - Relative count of events	D
EAC5	Tracer Wire	Tracer Wire Functionality	Third-Party Context	Discrete	Elicitation - Relative count of events	D
EAC6	Records Map	Facility Records/Map	Third-Party Context	Discrete	Bayesian update	Prior - D; Evidence - A
EC1	Excav. Equip.	Excavation Equipment	Excavation Context	Discrete	Elicitation - Relative count of events	B
EC2	Tooth L (mm)	Excavator tooth length (mm)	Excavation Context	Continuous	Elicitation - Distribution fitting	D
EC3	Cover Depth (ft)	Depth of Cover (ft)	Excavation Context	Continuous	Elicitation - Distribution fitting	D
EC4	Tooth W (mm)	Excavator tooth width (mm)	Excavation Context	Continuous	Elicitation - Distribution fitting	D
EC5	Adjacent to Pipe	Excavation adjacent to pipeline	Excavation Context	Discrete	Elicitation - Relative count of events	D
EC6	Excav. >Cover	Excavation Deeper than Cover	Excavation Context	Discrete	Elicitation - Conditional Assignment	N/A
EC7	Puncture Force (kN)	Puncture force (kN)	Excavation Context	Continuous	Elicitation - Histogram	D
FC1	Abandoned	Abandoned Facility	Facility Context	Discrete	Bayesian update	Prior - D; Evidence - A
FC2	Pipe System	Pipeline System	Environment	Discrete	Elicitation - Relative count of events	A
FC3	Material	Material	Facility Context	Discrete	Elicitation - Relative count of events	D
FC4	Wall thick (mm)	Pipe wall thickness (mm)	Facility Context	Continuous	Elicitation - Histogram	D
FC5	Ult. Strength (MPa)	Ultimate strength of material (MPa)	Facility Context	Continuous	Elicitation - Distribution fitting	D
FC6	Puncture Resistance (kN)	Puncture resistance (kN)	Facility Context	Continuous	Elicitation - Functional equation	N/A
NR1	L&M	Action of locating and marking	Non-DIRT Root Cause	Discrete	Elicitation - Conditional assignment	N/A
NR2	Right-Way Detected	Right way of excavation detected	Non-DIRT Root Cause	Discrete	Elicitation - Relative count of events	D
NR3	Sabotage	Sabotage	Non-DIRT Root Cause	Discrete	Elicitation - Relative count of events	B
R1	Notification	Notification made to One Call Center/ 811	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R2	Dug outside	Excavator dug outside area described on ticket	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R3	Incorrect info.	Excavator provided incorrect notification information	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R4	Out of date	Correct dig time relative to valid ticket start date/time	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R5	Locator-MI	Locator error (marked inaccurately)	DIRT Root Cause	Discrete	Bayesian update	Prior - D; Evidence - A
R6	Tracer-MI	Tracer wire issue (marked inaccurately)	DIRT Root Cause	Discrete	Bayesian update	Prior - D; Evidence - A
R7	Abandoned-MI	Abandoned facility (marked inaccurately)	DIRT Root Cause	Discrete	Bayesian update	Prior - D; Evidence - A
R8	Records Maps-MI	Incorrect facility records / maps (marked inaccurately)	DIRT Root Cause	Discrete	Bayesian update	Prior - D; Evidence - A
R9	Locator-NM	Locator error (not marked)	DIRT Root Cause	Discrete	Elicitation - Relative count of events	D
R10	Incomplete Marks-NM	Incomplete marks at damage location (not marked)	DIRT Root Cause	Discrete	Bayesian update	Prior - D; Evidence - A
R11	Abandoned-NM	Abandoned facility (not marked)	DIRT Root Cause	Discrete	Elicitation - Relative count of events	D
R12	Tracer-NM	Tracer wire issue (not marked)	DIRT Root Cause	Discrete	Bayesian update	Prior - D; Evidence - A
R13	Unlocatable Fac.-NM	Unlocatable facility (not marked)	DIRT Root Cause	Discrete	Elicitation - Relative count of events	D
R14	Records Map-NM	Incorrect facility records / maps (not marked)	DIRT Root Cause	Discrete	Bayesian update	Prior - D; Evidence - A
R15	No Response-NM	No response from operator / contract locator (not marked)	DIRT Root Cause	Discrete	Elicitation - Relative count of events	D
R16	Backfill Error	Improper backfilling practices	DIRT Root Cause	Discrete	Elicitation - Relative count of events	C
R17	Other Excav. Issue	Improper excavation practice not listed elsewhere	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R18	Marks Fade Lost	Marks faded or not maintained	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R19	Fail Shore Support	Excavator failed to protect / shore / support facilities	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R20	Fail Test-hole	Excavator dug prior to verifying marks by test-hole (pothole)	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R21	Failed Clearence	Excavator failed to maintain clearance after verifying marks	DIRT Root Cause	Discrete	Elicitation - Relative count of events	C
R22	OC center Error	One Call Center error	DIRT Root Cause	Discrete	Elicitation - Relative count of events	C
R23	Other Issue	Root cause not listed elsewhere	DIRT Root Cause	Discrete	Elicitation - Relative count of events	A
R24	Previous Dmg. Pipe	Previous damage	DIRT Root Cause	Discrete	Elicitation - Relative count of events	C
R25	Deteriorated Fac.	Deteriorated facility	DIRT Root Cause	Discrete	Elicitation - Relative count of events	C
RG1	Notification Issue	Notification Issue	DIRT Group	Discrete	Elicitation - Conditional assignment	N/A
RG2	Locating issue	Locating issue	DIRT Group	Discrete	Elicitation - Conditional assignment	N/A
RG3	Excavation Issue	Excavation Issue	DIRT Group	Discrete	Elicitation - Conditional assignment	N/A
RG4	Misc. Issue	Non-actionable (miscellaneous) factors	DIRT Group	Discrete	Elicitation - Conditional assignment	N/A
RG5	Misc. Pipe Issue	Miscellaneous - pipe-based	DIRT Group	Discrete	Elicitation - Conditional assignment	N/A
SG1	Other Notif. Issue	Other Notification Issue	Sub-DIRT Group	Discrete	Elicitation - Conditional assignment	N/A
SG2	Mark Inacc.(MI)	Facility marked inaccurately	Sub-DIRT Group	Discrete	Elicitation - Conditional assignment	N/A
SG3	Not Marked (NM)	Facility not marked	Sub-DIRT Group	Discrete	Elicitation - Conditional assignment	N/A
T1	Preventive Measures	Sufficient Actionable Preventive Measures	Summary Target Nodes	Discrete	Bayesian update	Prior - D; Evidence - A
T2	Pipe Hit	Pipe Hit	Summary Target Nodes	Discrete	Elicitation - Conditional assignment	N/A
T3	Force>Resistance	Puncture Force>Puncture Resistance	Summary Target Nodes	Discrete	Elicitation - Conditional assignment	N/A
T4	Damage (Puncture)	Damage (Puncture Failure)	Summary Target Nodes	Discrete	Elicitation - Conditional assignment	N/A

*A - Partner utility company damage reports; B - PHMSA gas distribution and transmission incident data; C - CGA's 2018 and 2019 DIRT Database; D - Previous GTI Energy models

Table B.2: List of the nodes in BaNTERA, node characteristics, and the techniques and data sources used to parameterize each node

B.3 Parameterized TPD Bayesian Network Model

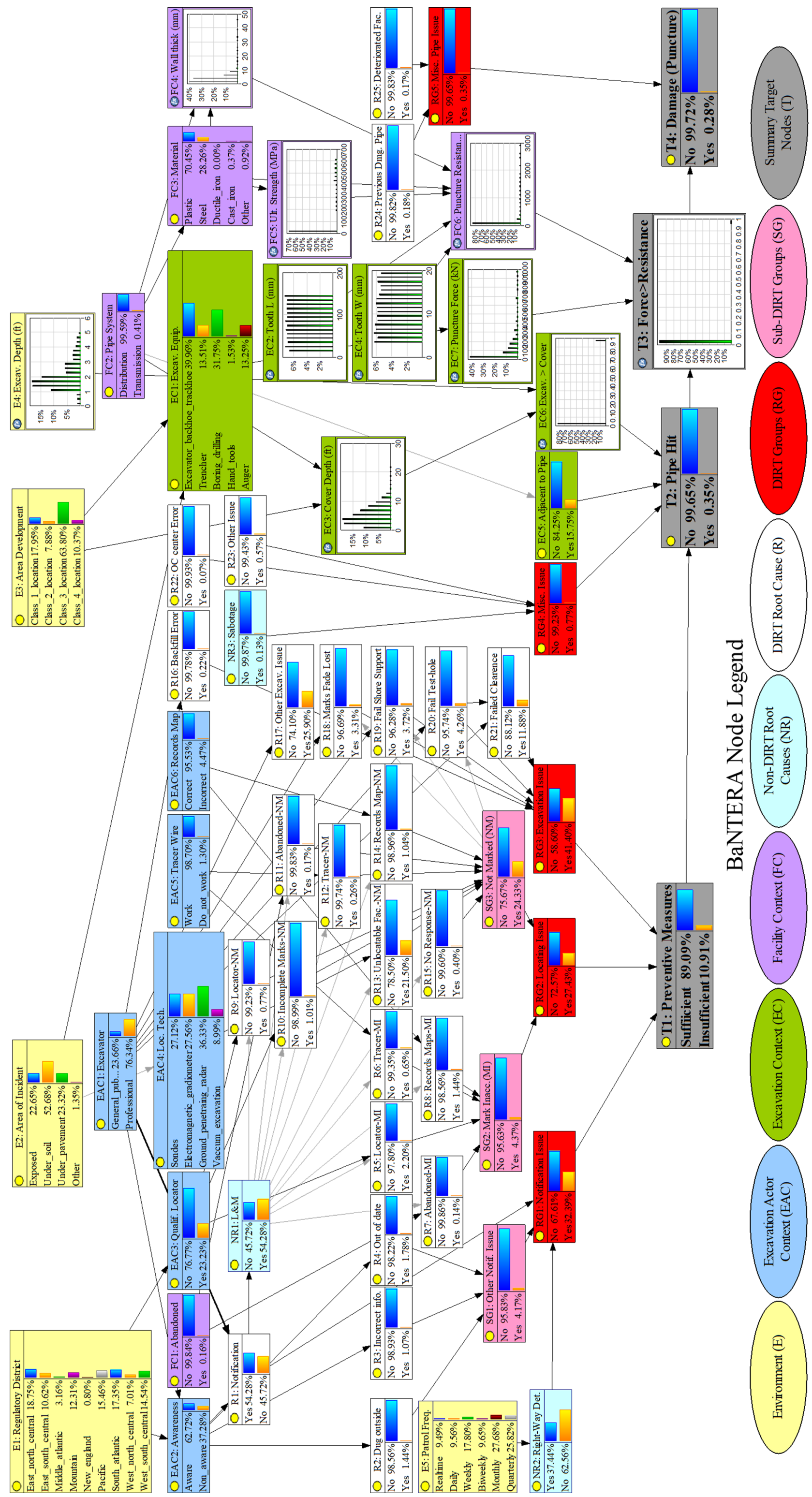


Figure B.1: Sample parameterization of BaNTERA's nodes. These values represent possible prior probabilities for a generic third-party excavation site. Parameterization and visualization made with GeNIe [1]

Appendix C: Appendix for Chapter 5

C.1 Nodes, States, and Marginal Prior Probabilities of the Bayesian Network Model used in the Sun Prairie Gas Explosion Case Study

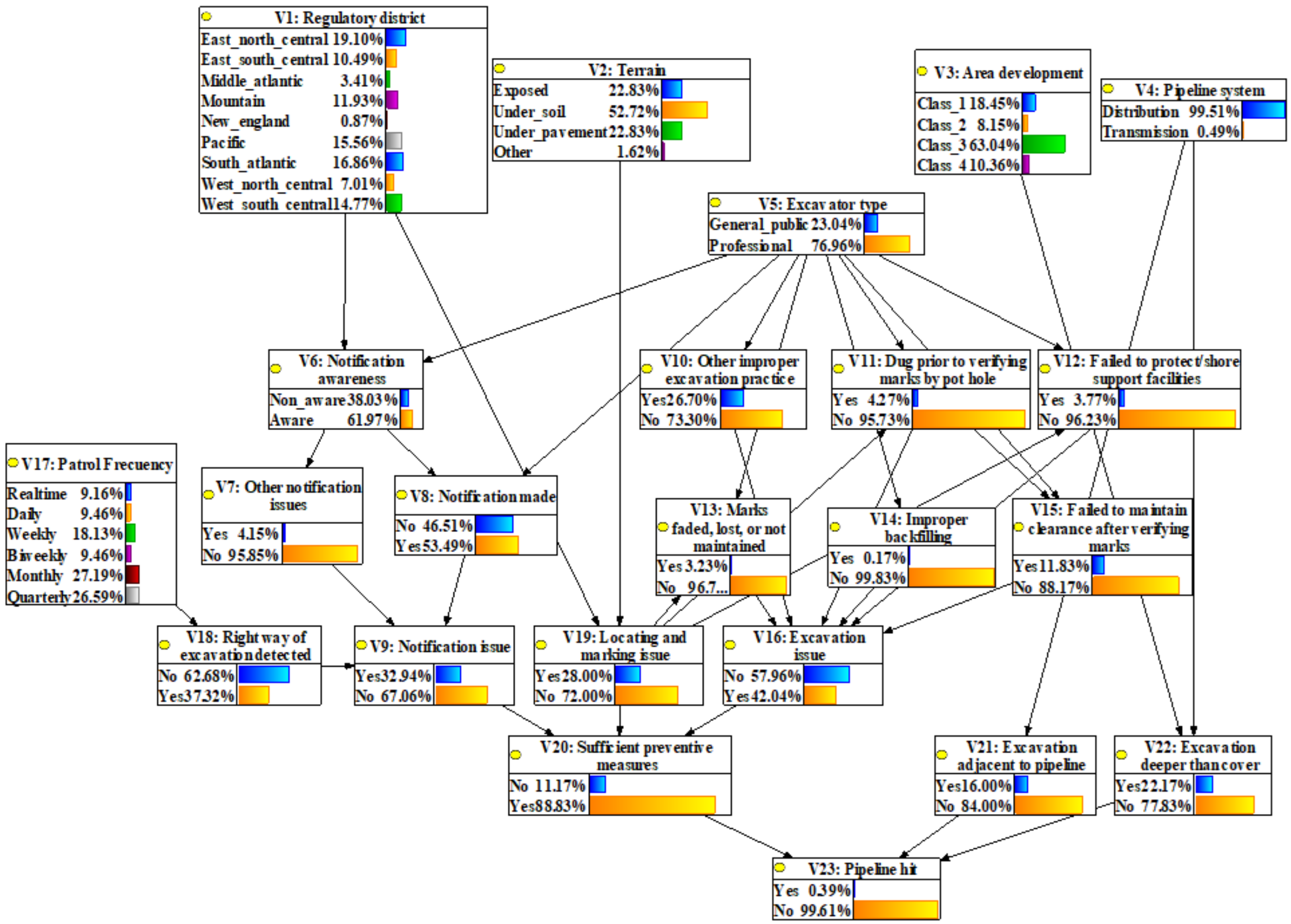


Figure C.1: Nodes, states, and marginal prior probabilities of the Bayesian network model used in the Sun Prairie gas explosion case study.

C.2 Twin Network Model used in the Counterfactual Scenarios

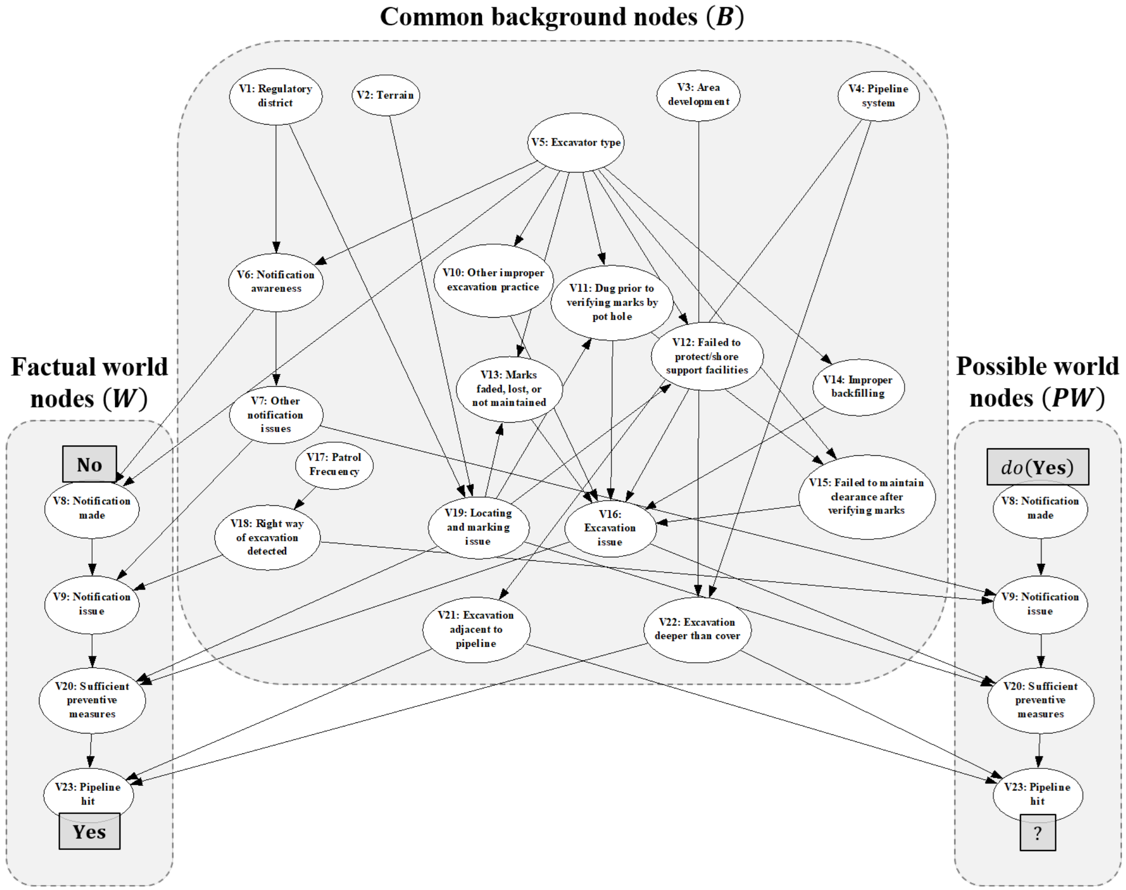


Figure C.2: Twin network model used in the counterfactual scenario A of the Sun Prairie gas explosion case study. Although it is not shown in the twin network figure for clarity, some common background nodes have instantiated evidence, as shown in Table 5.1.

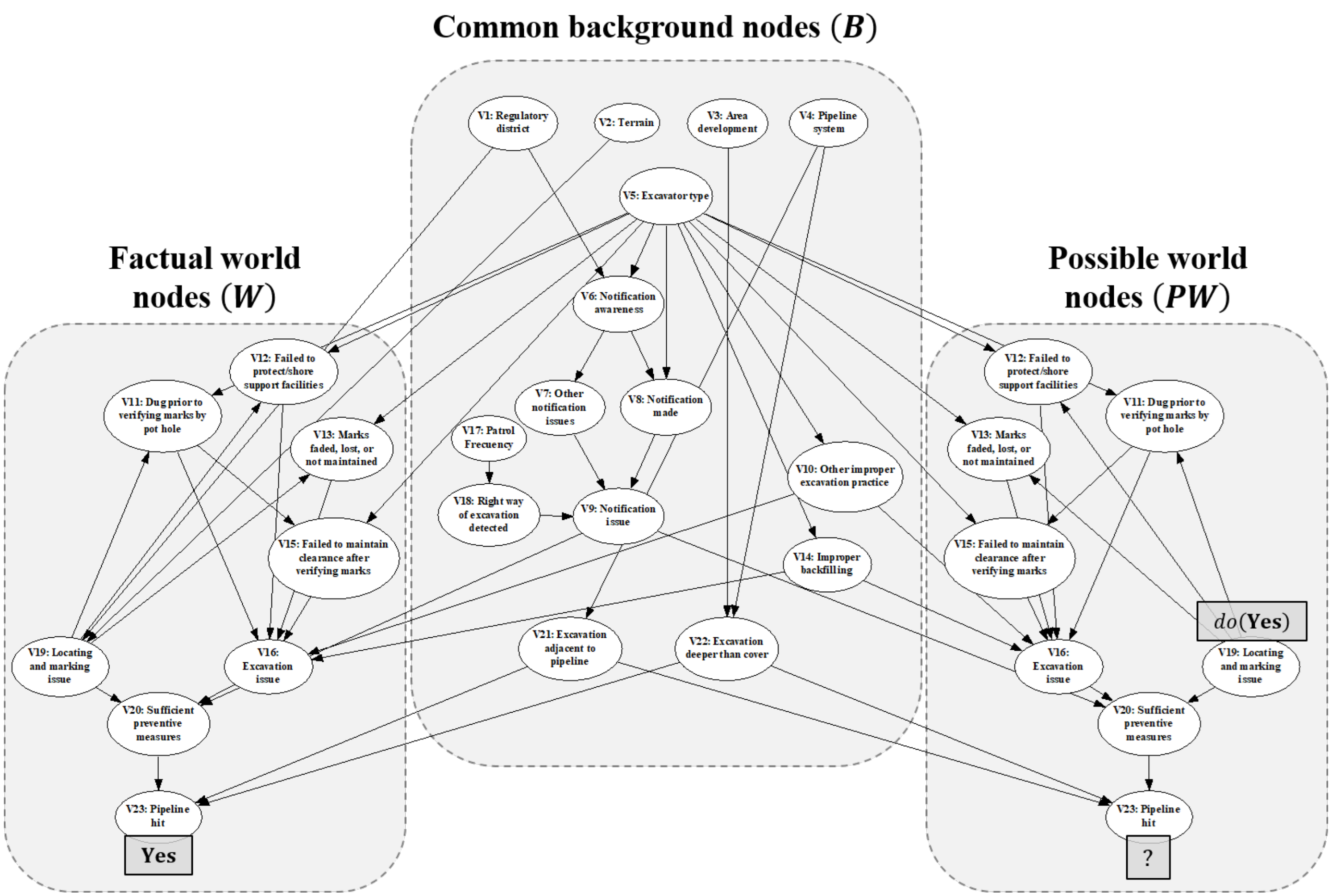


Figure C.3: Twin network model used in the counterfactual scenario B of the Sun Prairie gas explosion case study. Although it is not shown in the twin network figure for clarity, some common background nodes have instantiated evidence, as shown in Table 5.1.

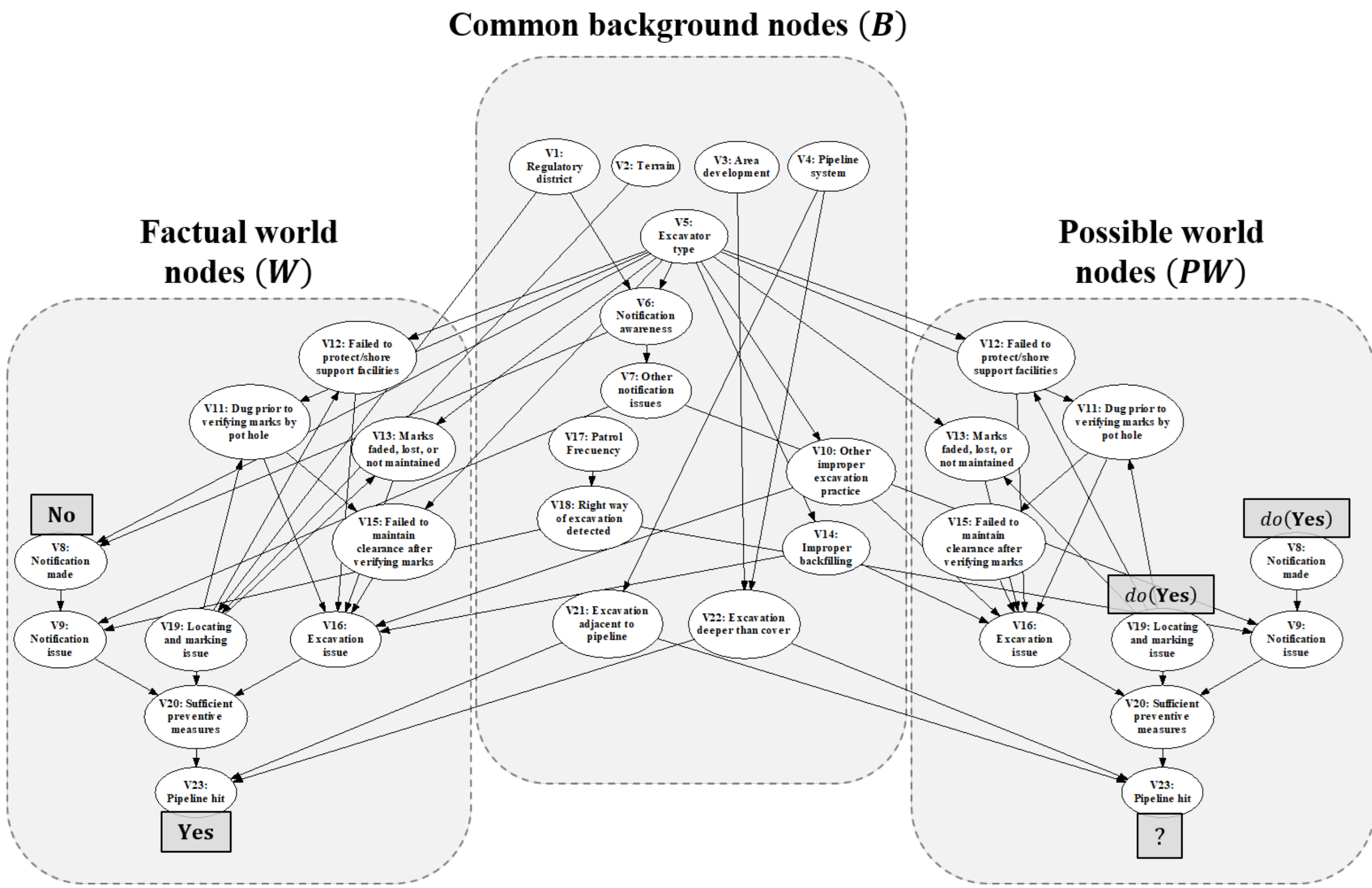


Figure C.4: Twin network model used in the counterfactual scenario C of the Sun Prairie gas explosion case study. Although it is not shown in the twin network figure for clarity, some common background nodes have instantiated evidence, as shown in Table 5.1.

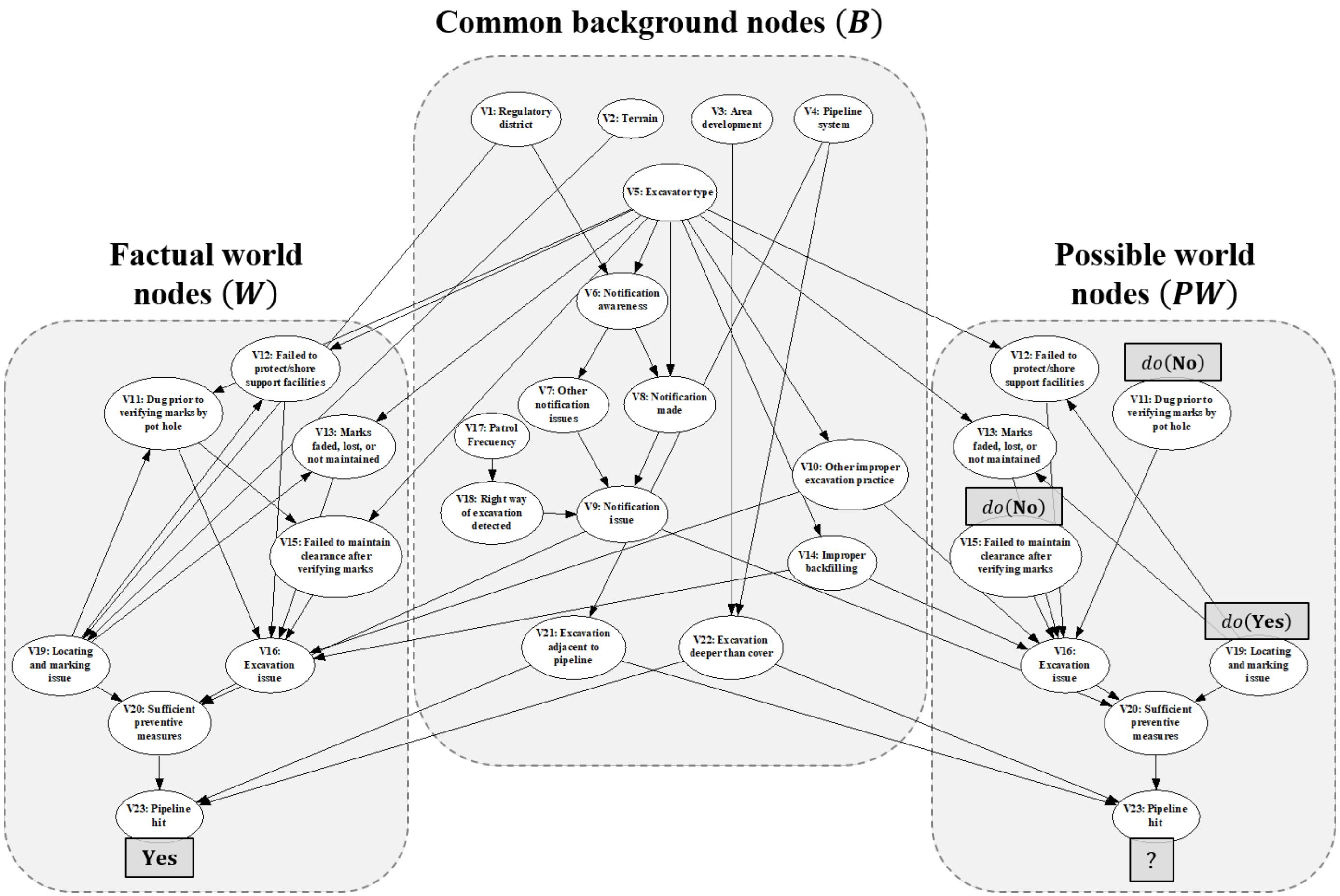


Figure C.5: Twin network model used in the counterfactual scenario D of the Sun Prairie gas explosion case study. Although it is not shown in the twin network figure for clarity, some common background nodes have instantiated evidence, as shown in Table 5.1.

Appendix D: Appendix for Chapter 6

D.1 SHyTERP Model: Assessing Safe Hydrogen Transportation through Existing and Repurposed Pipelines

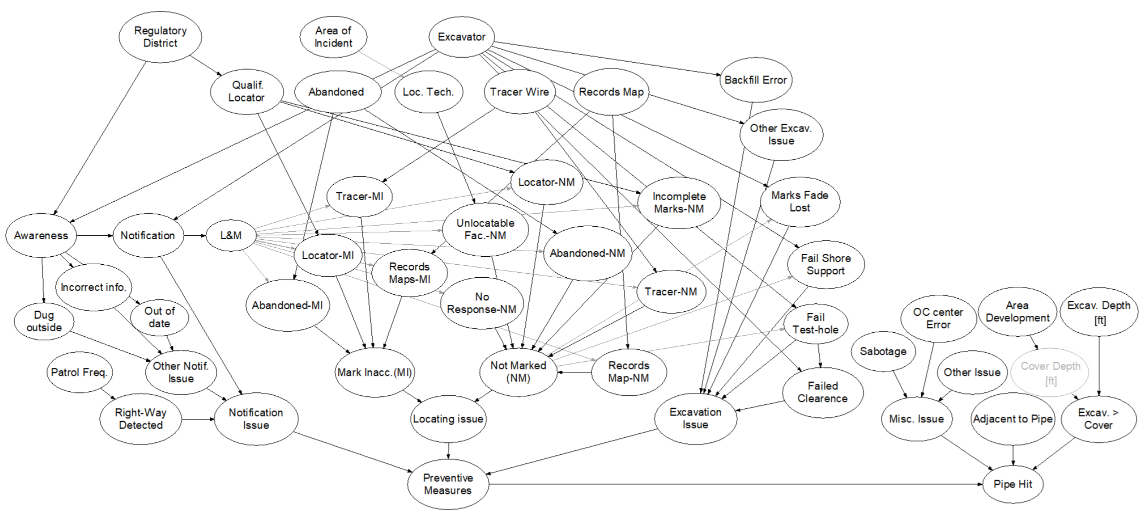


Figure D.1: Sub-model 1 - Third-party excavation and context.

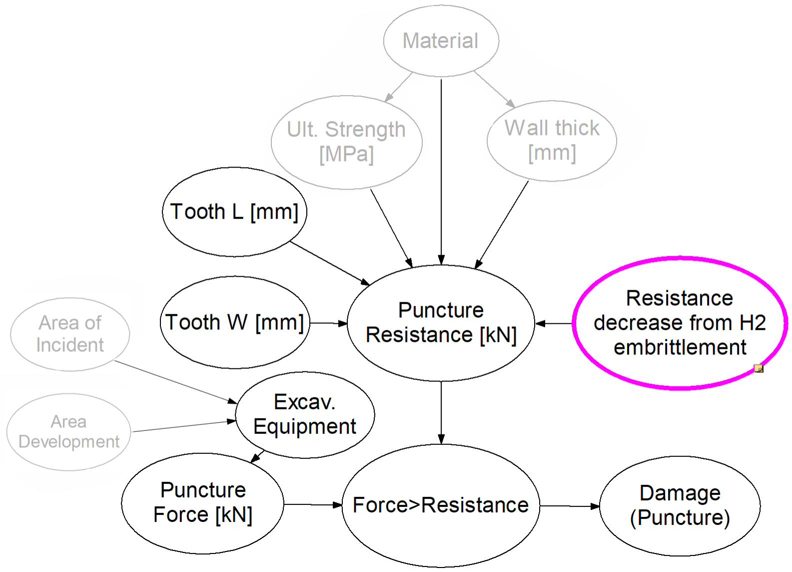


Figure D.2: Sub-model 2 - Puncture force and resistance. Variables directly affected by hydrogen are in pink.

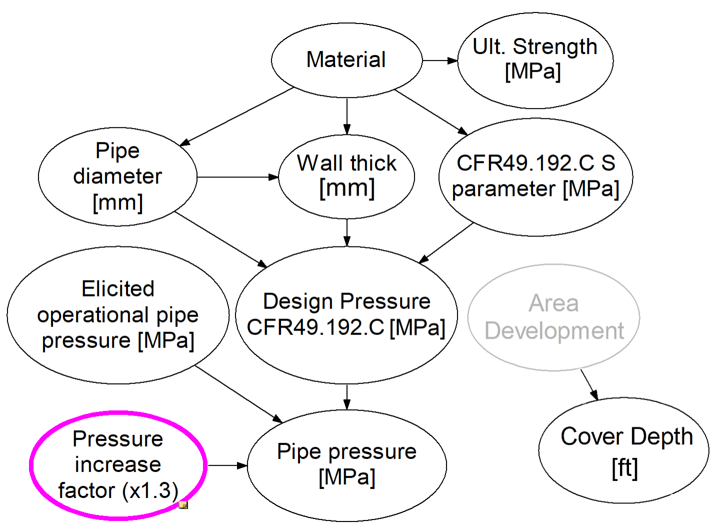


Figure D.3: Sub-model 3 - Existing pipeline characteristics. Variables directly affected by hydrogen are in pink.

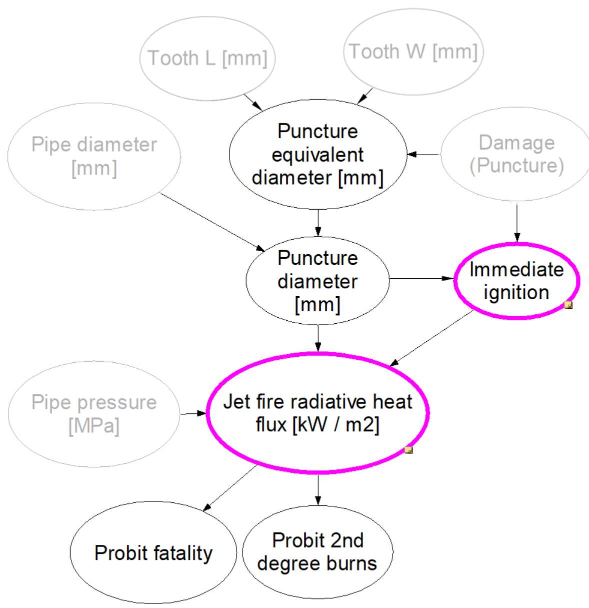


Figure D.4: Sub-model 4 - Gas release, jet flame radiation, and thermal harm. Variables directly affected by hydrogen are in pink.

Bibliography

[1] BayesFusion. GeNIe modeler software package, version 3.0 academic, 2022.

[2] Mohammad Modarres. *Risk analysis in engineering: techniques, tools, and trends*. CRC press, 2006.

[3] Nuclear Regulatory Commission. NRC: risk and performance concepts in the NRC's approach to regulation.

[4] Andrew Hopkins. *Disastrous decisions: the human and organisational causes of the Gulf of Mexico blowout*. CCH Australia Sydney, 2012.

[5] Jeffrey LaChance, Andrei Tchouvelev, and Angunn Engebo. Development of uniform harm criteria for use in quantitative risk analysis of the hydrogen infrastructure. *International Journal of Hydrogen Energy*, 36(3):2381–2388, 2011.

[6] Hy4Heat. Safety assessment: Conclusions report (incorporating quantitative risk assessment). Technical report, UK Department for Business, Energy & Industrial Strategy, 2021.

[7] Katrina M Groth, Reuel Smith, and Ramin Moradi. A hybrid algorithm for developing third generation hra methods using simulator data, causal models, and cognitive science. *Reliability Engineering & System Safety*, 191:106507, 2019.

[8] Louis Anthony Cox Jr. Improving causal inferences in risk analysis. *Risk Analysis*, 33(10):1762–1771, 2013.

[9] Sohag Kabir and Yiannis Papadopoulos. Applications of Bayesian networks and Petri nets in safety, reliability, and risk assessments: A review. *Safety Science*, 115:154–175, 2019.

[10] Katrina M Groth, Matthew R Denman, Michael C Darling, Thomas B Jones, and George F Luger. Building and using dynamic risk-informed diagnosis procedures for complex system accidents. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 234(1):193–207, 2020.

[11] Louis Anthony Cox Jr, Douglas A Popken, and Richard X Sun. *Causal analytics for applied risk analysis*. Springer, 2018.

[12] Ali Mosleh. PRA: a perspective on strengths, current limitations, and possible improvements. *Nuclear Engineering and Technology*, 46(1):1–10, 2014.

[13] Andrew Rae, Rob Alexander, and John McDermid. Fixing the cracks in the crystal ball: A maturity model for quantitative risk assessment. *Reliability Engineering & System Safety*, 125:67–81, 2014.

[14] Terje Aven. *Quantitative risk assessment: the scientific platform*. Cambridge University Press, 2011.

[15] George E Apostolakis. How useful is quantitative risk assessment? *Risk Analysis: An International Journal*, 24(3):515–520, 2004.

[16] Norman Fenton and Martin Neil. *Risk assessment and decision analysis with Bayesian networks*. CRC Press, 2018.

[17] Kathryn Roscoe, Anca Hanea, Ruben Jongejan, and Ton Vrouwenvelder. Levee system reliability modeling: The length effect and Bayesian updating. *Safety*, 6(1):7, 2020.

[18] Judea Pearl. *Causality*. Cambridge University Press, 2009.

[19] Terje Aven. *The science of risk analysis: Foundation and practice*. Routledge, 2019.

[20] Common Ground Alliance. DIRT annual report for 2019. Technical report, Common Ground Alliance (CGA), 908 King Street Suite 330, Alexandria, VA 22314, 2020.

[21] Pipeline and Hazardous Materials Safety Administration. Pipeline incident 20 year trends. Technical report, Pipeline and Hazardous Materials Safety Administration (PHMSA), 2022.

[22] Common Ground Alliance. DIRT annual report for 2021. Technical report, Common Ground Alliance (CGA), 908 King Street Suite 330, Alexandria, VA 22314, 2022.

[23] Andres Ruiz-Tagle, Austin D Lewis, Colin A Schell, Ernest Lever, and Katrina M Groth. Bantera: A bayesian network for third-party excavation risk assessment. *Reliability Engineering & System Safety*, 223:108507, 2022.

[24] Shanshan Fu, Yuerong Yu, Jihong Chen, Yongtao Xi, and Mingyang Zhang. A framework for quantitative analysis of the causation of grounding accidents in arctic shipping. *Reliability Engineering & System Safety*, 226:108706, 2022.

[25] Mario Van der Borst and Herman Schoonakker. An overview of PSA importance measures. *Reliability Engineering & System Safety*, 72(3):241–245, 2001.

[26] Nima Khakzad, Faisal Khan, and Paul Amyotte. Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. *Reliability Engineering & System Safety*, 96(8):925–932, 2011.

[27] Nancy Leveson. A new accident model for engineering safer systems. *Safety Science*, 42(4):237–270, 2004.

[28] Madeleine Rowland and Andrew Hopkins. *Learning from High Reliability Organisations*, chapter The importance of counterfactual thinking: the case of the 2003 Canberra Bushfires, page 237. CCH Australia Ltd., 2009.

[29] Andrew Hopkins. Issues in safety science. *Safety Science*, 67:6–14, 2014.

[30] Terje Aven. A risk science perspective on the discussion concerning Safety I, Safety II and Safety III. *Reliability Engineering & System Safety*, 217:108077, 2022.

[31] Terje Aven and Enrico Zio. Foundational issues in risk assessment and risk management. *Risk analysis*, 34(7):1164–1172, 2014.

[32] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic Books, 2018.

[33] The Daily Reporter Construction News. Tag Archives: Sun Prairie. Pages 1-2. *The Daily Reporter*, 2019/2020.

[34] Katrina M Groth and Ethan S Hecht. Hyram: A methodology and toolkit for quantitative risk assessment of hydrogen systems. *International Journal of Hydrogen Energy*, 42(11):7485–7493, 2017.

[35] Terje Aven, Yakov Ben-Haim, H Boje Andersen, Tony Cox, Enrique López Droguett, Michael Greenberg, Seth Guikema, Wolfgang Kröger, Ortwin Renn, Kimberly M Thompson, and Enrico Zio. *Society for risk analysis glossary*. Society for Risk Analysis, 2018.

[36] Stanley Kaplan and B John Garrick. On the quantitative definition of risk. *Risk Analysis*, 1(1):11–27, 1981.

[37] Floris Goerlandt, Nima Khakzad, and Genserik Reniers. Validity and validation of safety-related quantitative risk analysis: A review. *Safety Science*, 99:127–139, 2017.

[38] Hans J Pasman, William J Rogers, and M Sam Mannan. Risk assessment: What is it worth? shall we just do away with it, or can it do a better job? *Safety Science*, 99:140–155, 2017.

[39] Jan Erik Vinnem and Terje Aven. *Risk management with applications from the offshore petroleum industry*. Springer, 2007.

[40] Joel Tickner and David Kriebel. *Implementing the precautionary principle: perspectives and prospects*, chapter The role of science and precaution in environmental and public health policy, page 42. Edward Elgar Publishing, 2006.

[41] International Electrotechnical Commission. Iec/fdis 31010:2018 risk management — risk assessment techniques. Technical Report IEC/FDIS 31010:2009(E), International Organization for Standardization (ISO), 2018.

[42] Frank C Keil. Explanation and understanding. *Annual Review of Psychology*, 57:227, 2006.

[43] Peter Menzies. The causal structure of mechanisms. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(4):796–805, 2012.

[44] Elias Bareinboim, Juan D Correa, Duligur Ibeling, and Thomas Icard. On Pearl's hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: the works of Judea Pearl*, pages 507–556. Association for Computing Machinery, 2022.

[45] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

[46] Mohammad Modarres, Mark P Kaminskiy, and Vasiliy Krivtsov. *Reliability engineering and risk analysis: a practical guide*. CRC press, 2009.

[47] Mohammad Modarres, Mehdi Amiri, and Christopher Jackson. *Probabilistic physics of failure approach to reliability: modeling, accelerated testing, prognosis and reliability assessment*. John Wiley & Sons, 2017.

[48] Huairui Guo and Haitao Liao. Practical approaches for reliability evaluation using degradation data. In *Annual Reliability and Maintainability Symposium*, volume 7, pages 1–10, 2015.

[49] Roshanak Nateghi and Terje Aven. Risk analysis in the age of big data: The promises and pitfalls. *Risk Analysis*, 41(10):1751–1758, 2021.

[50] Andres Ruiz-Tagle Palazuelos and Enrique López Droguett. System-level prognostics and health management: A graph convolutional network–based framework. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 235(1):120–135, 2021.

[51] Nicola Paltrinieri, Louise Comfort, and Genserik Reniers. Learning about risk: Machine learning for risk assessment. *Safety Science*, 118:475–486, 2019.

[52] Ramin Moradi and Katrina M Groth. Modernizing risk assessment: A systematic integration of PRA and PHM techniques. *Reliability Engineering & System Safety*, 204:107194, 2020.

[53] Ramin Moradi, Andrés Ruiz-Tagle Palazuelos, Enrique Lopez Droguett, and Katrina M Groth. Toward a framework for risk monitoring of complex engineering systems with online operational data: A deep learning-based solution. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, page 1748006X221079964, 2020.

[54] Lauren Hund and Benjamin Schroeder. A causal perspective on reliability assessment. *Reliability Engineering & System Safety*, 195:106678, 2020.

[55] David A Broniatowski and Conrad Tucker. Assessing causal claims about complex engineered systems with quantitative data: internal, external, and construct validity. *Systems Engineering*, 20(6):483–496, 2017.

[56] Kangwon Seo and Rong Pan. Planning accelerated life tests with multiple sources of random effects. *Journal of Quality Technology*, 54(2):162–183, 2022.

[57] Ramaravind Kommiya Mothilal, Divyat Mahajan, Chenhao Tan, and Amit Sharma. Towards unifying feature attribution and counterfactual explanations: Different means to the same end. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 652–663, 2021.

[58] Olivier Salvi and Bruno Debray. A global view on ARAMIS, a risk assessment methodology for industries in the framework of the SEVESO II directive. *Journal of Hazardous Materials*, 130(3):187–199, 2006.

[59] Homayoon Dezfuli, Allan Benjamin, Christopher Everett, Gaspare Maggio, Michael Stamatelatos, Robert Youngblood, Sergio Guarro, Peter Rutledge, James Sherrard, Curtis Smith, et al. NASA risk management handbook. Technical report, National Aeronautics and Space Administration (NASA), 2011.

[60] Robert M Weinstock, Carol S Smidts, Ali Mosleh, Yung-Hsien Chang, Sankaran Swaminathan, Francisco J Groen, and Zhibin Tan. Quantitative risk assessment system (QRAS). Technical report, National Aeronautics and Space Administration (NASA), 2001.

[61] KD Russell, CL Hoffman, KJ Kvarfordt, CL Smith, ST Wood, and E Lois. Systems analysis programs for hands-on integrated reliability evaluations (SAPHIRE) version 6.0 system overview manual. Technical report, NUREG0CR-6532, U.S. Nuclear Regulatory Commission (NRC), 1999.

[62] Nancy G Leveson. Applying systems thinking to analyze and learn from events. *Safety Science*, 49(1):55–64, 2011.

[63] Shu Yang and B Wayne Bequette. Observational process data analytics using causal inference. *AIChE Journal*, page e17986, 2022.

[64] Per Hokstad and Marvin Rausand. Common cause failure modeling: status and trends. *Handbook of Performability Engineering*, pages 621–640, 2008.

[65] Eric Stemn and Theophilus Joe-Asare. The influence of accident manuals on the effectiveness of accident investigations–An analysis of accident management documents of Ghanaian mines. *Safety Science*, 135:105129, 2021.

[66] Linda Drupsteen and Peter Hasle. Why do organizations not learn from incidents? Bottlenecks, causes and conditions for a failure to effectively learn. *Accident Analysis & Prevention*, 72:351–358, 2014.

[67] Jean Christophe Le Coze. What have we learned about learning from accidents? Post-disasters reflections. *Safety Science*, 51(1):441–453, 2013.

[68] Chris Johnson and C.M. Holloway. A survey of logic formalisms to support mishap analysis. *Reliability Engineering & System Safety*, 80(3):271–291, 2003.

[69] National Transportation Safety Board. Aircraft accident report: Descent below visual glidepath and impact with seawall Asiana Airlines Flight 214. Technical report, U.S. National Transportation Safety Board, 2014.

[70] National Transportation Safety Board. Aircraft accident report: Flight into terrain during missed approach U.S. Air Flight 1016. Technical report, U.S. National Transportation Safety Board, 1995.

[71] Kate Branford, Andrew Hopkins, and Neelam Naikar. Guidelines for AcciMap analysis. In *Learning from High Reliability Organisations*, pages 1–12. CCH Australia Ltd, 2009.

[72] Jan Sanders and Peter Ladkin. Introduction to Why-Because analysis. Technical report, Computer Networks and Distributed Systems Group, Bielefeld University, 2012.

[73] Chris Johnson. *A handbook for the reporting of incidents and accidents*. Springer, UK, London, 2003.

[74] Nektarios Karanikas, Dimitrios Chionis, and Anastasios Plioutsias. “old” and “new” safety thinking: Perspectives of aviation safety investigators. *Safety Science*, 125:104632, 2020.

[75] Sidney W.A. Dekker. Reconstructing human contributions to accidents: the new view on error and performance. *Journal of Safety Research*, 33(3):371–385, 2002.

[76] Hollnagel Erik. *FRAM the functional resonance analysis method: modelling complex socio-technical systems*. CRC Press, 2017.

[77] G Woo, T Maynard, and J Seria. *Reimagining history: counterfactual risk analysis*. Lloyd's emerging risk report, 1 Lime St, London, United Kingdom, 2017.

[78] Julien Etienne. Knowledge transfer in organisational reliability analysis: From post-accident studies to normal operations studies. *Safety Science*, 46(10):1420–1434, 2008.

[79] Adam Hulme, Neville A Stanton, Guy H Walker, Patrick Waterson, and Paul M Salmon. What do applications of systems thinking accident analysis methods tell us about accident causation? A systematic review of applications between 1990 and 2018. *Safety Science*, 117:164–183, 2019.

[80] C.Y. Lam and A.M. Cruz. Risk analysis for consumer-level utility gas and liquefied petroleum gas incidents using probabilistic network modeling: a case study of gas incidents in Japan. *Reliability Engineering & System Safety*, 185:198–212, 2019.

[81] William Hughes, Wei Zhang, Diego Cerrai, Amvrossios Bagtzoglou, David Wanik, and Emmanouil Anagnostou. A hybrid physics-based and data-driven model for power distribution system infrastructure hardening and outage simulation. *Reliability Engineering & System Safety*, 225:108628, 2022.

[82] Edward J Oughton, Daniel Ralph, Raghav Pant, Eireann Leverett, Jennifer Copic, Scott Thacker, Rabia Dada, Simon Ruffle, Michelle Tuveson, and Jim W Hall. Stochastic counterfactual risk analysis for the vulnerability assessment of cyber-physical attacks on electricity distribution infrastructure networks. *Risk Analysis*, 39(9):2012–2031, 2019.

[83] Judea Pearl. Structural counterfactuals: A brief introduction. *Cognitive Science*, 37(6):977–985, 2013.

[84] Helge Langseth and Luigi Portinale. Bayesian networks in reliability. *Reliability Engineering & System Safety*, 92(1):92–108, 2007.

[85] Nima Khakzad. Which fire to extinguish first? a risk-informed approach to emergency response in oil terminals. *Risk Analysis*, 38(7):1444–1454, 2018.

[86] Austin Lewis and Katrina Groth. A review of methods for discretizing continuous-time accident sequences. In *Proceedings of the 29th European Safety and Reliability Conference*, pages 754–761, 2019.

[87] Judea Pearl. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th Conference of the Cognitive Science Society*, pages 329–334, 1985.

[88] Judea Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier, 2014.

[89] Jeevith Hegde, Ingrid Bouwer Utne, Ingrid Schjølberg, and Brede Thorkildsen. A Bayesian approach to risk modeling of autonomous subsea intervention operations. *Reliability Engineering & System Safety*, 175:142–159, 2018.

[90] Philippe Weber, Gabriela Medina-Oliva, Christophe Simon, and Benoît Iung. Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, 25(4):671–682, 2012.

[91] Lusine Mkrtchyan, Luca Podofillini, and Vinh N Dang. Bayesian belief networks for human reliability analysis: A review of applications and gaps. *Reliability Engineering & System Safety*, 139:1–16, 2015.

[92] Mario Brito and Gwyn Griffiths. A bayesian approach for predicting risk of autonomous underwater vehicle loss during their missions. *Reliability Engineering & System Safety*, 146:55–67, 2016.

[93] Limao Zhang, Xianguo Wu, Yawei Qin, Miroslaw J Skibniewski, and Wenli Liu. Towards a fuzzy Bayesian network based approach for safety risk analysis of tunnel-induced pipeline damage. *Risk Analysis*, 36(2):278–301, 2016.

[94] Jyoti Bhandari, Rouzbeh Abbassi, Vikram Garaniya, and Faisal Khan. Risk analysis of deepwater drilling operations using bayesian network. *Journal of Loss Prevention in the Process Industries*, 38:11–23, 2015.

[95] Judea Pearl. Causal diagrams for empirical research. *Biometrika*, 82(4):669–688, 1995.

[96] Xingguang Wu, Huirong Huang, Weichao Yu, Yuming Lin, Yanhui Xue, Qingwen Cai, and Jili Xu. Accident analysis and risk prediction of tank farm based on Bayesian network method. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, page 1748006X221139906, 2022.

[97] Yan Fu Wang, Min Xie, Kwai-Sang Chin, and Xiu Ju Fu. Accident analysis model based on bayesian network and evidential reasoning approach. *Journal of Loss Prevention in the Process Industries*, 26(1):10–21, 2013.

[98] Xiaoge Zhang and Sankaran Mahadevan. Bayesian network modeling of accident investigation reports for aviation safety assessment. *Reliability Engineering & System Safety*, 209:107371, 2021.

[99] Chio Lam and Wenxing Zhou. Statistical analyses of incidents on onshore gas transmission pipelines based on PHMSA database. *International Journal of Pressure Vessels and Piping*, 145:29–40, 2016.

[100] Pipeline and Hazardous Materials Safety Administration. Fact sheet: Excavation damage, 2023. U.S. Department of Transportation. Accessed March, 2023.

[101] Y. Zhang and W.G. Weng. Bayesian network model for buried gas pipeline failure analysis caused by corrosion and external interference. *Reliability Engineering & System Safety*, 203:107089, 2020.

[102] Yanbao Guo, Xiaoli Meng, Deguo Wang, Tao Meng, Shuhai Liu, and Renyang He. Comprehensive risk evaluation of long-distance oil and gas transportation pipelines using a fuzzy Petri net model. *Journal of Natural Gas Science and Engineering*, 33:18–29, 2016.

[103] Jun Li, Hong Zhang, Yinshan Han, and Baodong Wang. Study on failure of third-party damage for urban gas pipeline based on fuzzy comprehensive evaluation. *PLOS ONE*, 11(11):1–15, 11 2016.

[104] Golam Kabir, Rehan Sadiq, and Solomon Tesfamariam. A fuzzy Bayesian belief network for safety assessment of oil and gas pipelines. *Structure and Infrastructure Engineering*, 12(8):874–889, 2016.

[105] Kai Wen, J. Gong, Fanglin Chen, and Yuqing Liu. A framework for calculating the failure probability of natural gas pipeline. *Journal of Computer Science Technology Updates*, 1:1–8, 2014.

[106] Aihua Liu, Ke Chen, Xiaofei Huang, Didi Li, and Xiaochun Zhang. Dynamic risk assessment model of buried gas pipelines based on system dynamics. *Reliability Engineering & System Safety*, 208:107326, 2021.

[107] Mohammed Taleb-Berrouane, Faisal Khan, and Kelly Hawboldt. Corrosion risk assessment using adaptive bow-tie (ABT) analysis. *Reliability Engineering & System Safety*, 214:107731, 2021.

[108] Changxi Wang and Elsayed A. Elsayed. Stochastic modeling of corrosion growth. *Reliability Engineering & System Safety*, 204:107120, 2020.

[109] Babak Omidvar and Hamid Karimi Kivi. Multi-hazard failure probability analysis of gas pipelines for earthquake shaking, ground failure and fire following earthquake. *Natural Hazards: Journal of the International Society for the Prevention and Mitigation of Natural Hazards*, 82(1):703–720, May 2016.

[110] Ussama Kaewpradap, Marc Singer, Srdjan Nesic, and Suchada Punpruk. Comparison of model predictions and field data–the case of top of the line corrosion. In *CORROSION 2015*, pages 1007–1016. OnePetro, 03 2015.

[111] Qishi Chen and MA Nessim. Reliability–based prevention of mechanical damage to pipelines. Technical Report PR-244-9729, C-FER Technologies, 1999.

[112] W. Xiang and W. Zhou. Bayesian network model for predicting probability of third-party damage to underground pipelines and learning model parameters from incomplete datasets. *Reliability Engineering & System Safety*, 205:107262, 2021.

[113] Linlin Lu, Wei Liang, Laibin Zhang, Hong Zhang, Zhong Lu, and Jinzhi Shan. A comprehensive risk evaluation method for natural gas pipelines by combining a risk matrix with a bow-tie model. *Journal of Natural Gas Science and Engineering*, 25:124–133, 2015.

[114] Wei Liang, Jinqiu Hu, Laibin Zhang, Cunjie Guo, and Weipeng Lin. Assessing and classifying risk of pipeline third-party interference based on fault tree and SOM. *Engineering Applications of Artificial Intelligence*, 25(3):594–608, 2012.

[115] Wenhe Wang, Kuiling Shen, Beibei Wang, Chuanfu Dong, Faisal Khan, and Qingsheng Wang. Failure probability analysis of the urban buried gas pipelines using Bayesian networks. *Process Safety and Environmental Protection*, 111:678–686, 2017.

[116] Xinhong Li, Guoming Chen, Shengyu Jiang, Rui He, Changhang Xu, and Hongwei Zhu. Developing a dynamic model for risk analysis under uncertainty: Case of third-party damage on subsea pipelines. *Journal of Loss Prevention in the Process Industries*, 54:289–302, 2018.

[117] Michael Smith, Lewis Barton, Konstantinos Pesinis, and Ian Laing. Intelligent corrosion prediction using Bayesian networks. In *CORROSION 2019*. OnePetro, 2019.

[118] Michael Smith, Matthew Capewell, Konstantinos Pesinis, and Erika Santana. Bayesian network for data-driven inspection planning of oil and gas pipelines. In *CORROSION 2020*. OnePetro, 2020.

[119] Christopher Jackson and Ali Mosleh. Oil and gas pipeline third party damage (TPD) - a new way to model external hazard failure. In *Proceedings of the 14th Probabilistic Safety Assessment and Management, PSAM*, volume 14, pages 1–12, Los Angeles, CA, USA, 16-21 September, 2018.

[120] Xiaoyan Guo, Laibin Zhang, Wei Liang, and Stein Haugen. Risk identification of third-party damage on oil and gas pipelines through the Bayesian network. *Journal of Loss Prevention in the Process Industries*, 54:163–178, 2018.

[121] Yan Cui, Noor Quddus, and Chad V Mashuga. Bayesian network and game theory risk assessment model for third-party damage to oil and gas pipelines. *Process Safety and Environmental Protection*, 134:178–188, 2020.

[122] Tishun Peng. Modernizing tools for 3rd party damage. Technical Report 22412, GTI Energy, 2019.

[123] Babajide Kolade. Towards the development of a risk model for locates. Technical Report 22397, GTI Energy, 2019.

[124] U.S. Department of Energy. National clean hydrogen strategy and roadmap. Technical report, U.S. Department of Energy, 2022.

[125] P. W. Parfomak. Pipeline transportation of hydrogen: Regulation, research, and policy. Technical Report R46700, Congressional Research Service, 2021.

[126] M. Penev M.W. Melaina, O. Antonia. Blending hydrogen into natural gas pipeline networks: A review of key issues. Technical Report TP-5600-51995, National Renewable Energy Laboratory (NREL), 2013.

[127] Zine labidine Messaoudani, Fotis Rigas, Mahar Diana Binti Hamid, and Che Rosmani Che Hassan. Hazards, safety and knowledge gaps on hydrogen transmission via natural gas grid: A critical review. *International Journal of Hydrogen Energy*, 41(39):17511–17525, 2016.

[128] HAJ Froeling, Marc T Dröge, Gabriela F Nane, and Ad JM Van Wijk. Quantitative risk analysis of a hazardous jet fire event for hydrogen transport in natural gas transmission pipelines. *International Journal of Hydrogen Energy*, 46(17):10411–10422, 2021.

[129] Andrzej Witkowski, Andrzej Rusin, Mirosław Majkut, and Katarzyna Stolecka. Analysis of compression and transport of the methane/hydrogen mixture in existing natural gas pipelines. *International Journal of Pressure Vessels and Piping*, 166:24–34, 2018.

[130] Tim Harwood, Mark Danter, Russ Oxley, Stella Matthews, Catherine Spriggs, Dan Allason, Mike Acton, and Fiona Fylan. H21 Phase 1 technical sumary report. Technical report, HSE Science Division and DNV GL, 2021.

[131] Hans J Pasman and William J Rogers. Risk assessment by means of Bayesian networks: A comparative study of compressed and liquefied h2 transportation and tank station risks. *International Journal of Hydrogen Energy*, 37(22):17415–17425, 2012.

[132] Andrzej Rusin and Katarzyna Stolecka. Reducing the risk level for pipelines transporting carbon dioxide and hydrogen by means of optimal safety valves spacing. *Journal of Loss Prevention in the Process Industries*, 33:77–87, 2015.

[133] Francisco Filipe Cunha Lima Viana, Marcelo Hazin Alencar, Rodrigo José Pires Ferreira, and Adiel Teixeira De Almeida. Multidimensional risk assessment and categorization of hydrogen pipelines. *International Journal of Hydrogen Energy*, 47(42):18424–18440, 2022.

[134] Paola Russo, Alessandra De Marco, and Fulvio Parisi. Assessment of the damage from hydrogen pipeline explosions on people and buildings. *Energies*, 13(19):5051, 2020.

[135] Christopher W San Marchi and Brian P Somerday. Technical reference for hydrogen compatibility of materials. Technical Report SAND2012-7321, Sandia National Laboratories, 2012.

[136] Laurent Briottet, I Moro, and P Lemoine. Quantifying the hydrogen embrittlement of pipeline steels for safety considerations. *International Journal of Hydrogen Energy*, 37(22):17616–17623, 2012.

[137] K Xu. Hydrogen embrittlement of carbon steels and their welds. In *Gaseous Hydrogen Embrittlement of Materials in Energy Technologies*, pages 526–561. Elsevier, 2012.

[138] Hryhoriy Nykyforchyn, Olha Zvirko, Myroslava Hredil, Halyna Krechkovska, Oleksandr Tsyrulnyk, Oleksandra Student, and Leonid Unigovskyi. Methodology of hydrogen embrittlement study of long-term operated natural gas distribution pipeline steels caused by hydrogen transport. *Frattura e Integrita Strutturale*, 16(59), 2022.

[139] OT Tsyrul'nyk, HM Nykyforchyn, D Yu Petryna, MI Hredil, and IM Dz'Oba. Hydrogen degradation of steels in gas mains after long periods of operation. *Materials Science*, 43(5):708–717, 2007.

[140] Lin Zhang and RA Adey. Prediction of third party damage failure frequency for pipelines transporting mixtures of natural gas and hydrogen. In *Proceedings of the 3rd International Conference on Hydrogen Safety, Ajaccio-Corsica, France*, volume 1618, 2009.

[141] Institution of Gas Engineers and Managers. IGEM/H/1 - reference standard for low pressure hydrogen utilisation. Technical Report Communication 1860, Institution of Gas Engineers & Managers (IGEM), 2022.

[142] Gas Research Institute. Electrostatic discharger system for polyethylene pipes in gas distribution. Technical Report GRI-92/0460, Southwest Research Institute, 1992.

[143] G Hankinson, H Mathurkar, and BJ Lowesmith. Ignition energy and ignition probability of methane-hydrogen-air mixtures. In *Proceedings of the 3rd International Conference on Hydrogen Safety, Ajaccio-Corsica, France*, volume 1618, 2009.

[144] Paul Hünermund and Elias Bareinboim. Causal inference and data-fusion in econometrics. *arXiv:1912.09104*, 2019.

[145] Michael Joffe, Manoj Gambhir, Marc Chadeau-Hyam, and Paolo Vineis. Causal diagrams in systems epidemiology. *Emerging Themes in Epidemiology*, 9(1):1, 2012.

[146] Andres Ruiz-Tagle, Enrique Lopez Droguett, and Katrina M Groth. Exploiting the capabilities of Bayesian networks for engineering risk assessment: Causal reasoning through interventions. *Risk Analysis*, 42(6):1306–1324, 2022.

[147] Judea Pearl. The do-calculus revisited. *arXiv:1210.4852*, 2012.

[148] Tyler J VanderWeele and Ilya Shpitser. On the definition of a confounder. *Annals of Statistics*, 41(1):196, 2013.

[149] Ilya Shpitser and Judea Pearl. Identification of conditional interventional distributions. *arXiv:1206.6876*, 2012.

[150] Santtu Tikka and Juha Karvanen. Identifying causal effects with the R package `causaleffect`. *Journal of Statistical Software*, 76(12):1–30, 2017.

[151] Pipeline and Hazardous Materials Safety Administration. Annual report mileage summary statistics. Technical report, U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (PHMSA), 2020.

[152] Joseph S. Santarelli. Risk analysis of natural gas distribution pipelines with respect to third party damage. Master's thesis, Western University, Ontario, Canada, 4 2019.

[153] Common Ground Alliance. DIRT annual report for 2018. Technical report, Common Ground Alliance (CGA), 908 King Street Suite 330, Alexandria, VA 22314, 2019.

[154] Common Ground Alliance. White paper: Data-informed insights and recommendations for more effective excavator outreach. Technical report, Common Ground Alliance (CGA), 908 King Street Suite 330, Alexandria, VA 22314, 2019.

[155] Common Ground Alliance. The definitive guide for underground safety & damage prevention. Technical report, Common Ground Alliance (CGA), 908 King Street Suite 330, Alexandria, VA 22314, 2020.

[156] Colin R Blyth. On Simpson's paradox and the sure-thing principle. *Journal of the American Statistical Association*, 67(338):364–366, 1972.

[157] Judea Pearl. Comment: understanding Simpson's paradox. *The American Statistician*, 68(1):8–13, 2014.

[158] Ali Noroozian, Reza Baradaran Kazemzadeh, Seyed Taghi Akhavan Niaki, and Enrico Zio. System risk importance analysis using Bayesian networks. *International Journal of Reliability, Quality and Safety Engineering*, 25(01):1850004, 2018.

[159] Saran Srikanth Bodda, Abhinav Gupta, and Nam Dinh. Enhancement of risk informed validation framework for external hazard scenario. *Reliability Engineering & System Safety*, 204:107140, 2020.

[160] Esmaeil Zarei, Ali Azadeh, Nima Khakzad, Mostafa Mirzaei Aliabadi, and Iraj Mohammadfam. Dynamic safety assessment of natural gas stations using Bayesian network. *Journal of Hazardous Materials*, 321:830–840, 2017.

[161] Mohammad Yazdi and Sohag Kabir. A fuzzy Bayesian network approach for risk analysis in process industries. *Process Safety and Environmental Protection*, 111:507–519, 2017.

[162] Chengdong Wang. *Hybrid causal logic methodology for risk assessment.* University of Maryland, College Park, 2007.

[163] Pipeline and Hazardous Materials Safety Administration (PHMSA). Fact sheet: Excavation damage, 2020. Online; accessed July-2020.

[164] Daniel C Brooker. Experimental puncture loads for external interference of pipelines by excavator equipment. *International Journal of Pressure Vessels and Piping*, 82(11):825–832, 2005.

[165] Daniel C Brooker. Numerical modelling of pipeline puncture under excavator loading. Part II: parametric study. *International Journal of Pressure Vessels and Piping*, 80(10):727–735, 2003.

[166] Daniel C Brooker. Numerical modelling of pipeline puncture under excavator loading. Part I. development and validation of a finite element material failure model for puncture simulation. *International Journal of Pressure Vessels and Piping*, 80(10):715–725, 2003.

[167] Ernest Lever, Evan D. Sherwin, Katrina M. Groth, and Max Henrion. Data collection, normalization, and integration methods to enhance risk assessment tools for decision-making. Technical Report DOT693JK31910002POTA, Gas Technology Institute (GTI), 2022.

[168] Pipeline and Hazardous Materials Safety Administration (PHMSA). PHMSA distribution, transmission & gathering, LNG, and liquid accident and incident data, 2020. Online; accessed December-2020.

[169] Seth D Guikema and Jeremy P Coffelt. Modeling count data in risk analysis and reliability engineering. In *Handbook of Performability Engineering*, pages 579–594. Springer, 2008.

[170] Dana Kelly and Curtis Smith. *Bayesian inference for probabilistic risk assessment: A practitioner's guidebook.* Springer Science & Business Media, 2011.

[171] Gareth W. Parry and Mary T. Drouin. Risk-informed regulatory decision-making at the US NRC: dealing with model uncertainty. In *Model Uncertainty: Conceptual and Practical Issues in the Context of Risk-Informed Decision Making*, pages 9–20, College Park, MD, USA, 10 2009.

[172] Uffe Kjærulff and Linda C Van Der Gaag. Making sensitivity analysis computationally efficient. *arXiv:1301.3868*, 2013.

[173] Pipeline and Hazardous Materials Safety Administration (PHMSA). PHMSA national pipeline performance measures, 2021. Online; accessed July-2021.

[174] Pipeline and Hazardous Materials Safety Administration. A study on the impact of excavation damage on pipeline safety. Technical report, U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (PHMSA), 2014.

[175] Agenarisk developer, version 10, 2022.

[176] Franz Josef Lohmar. World geodetic system 1984—geodetic reference system of GPS orbits. In *GPS-Techniques Applied to Geodesy and Surveying*, pages 476–486. Springer, 1988.

[177] QGIS Development Team. *QGIS Geographic Information System*. QGIS Association, 2022.

[178] Daniel A. Ersoy, Ernest Lever, and Khalid Farrag. Latent third party damages statistical approach. Technical Report 21735, GTI Energy, 2015.

[179] Michael W Morris and Paul C Moore. The lessons we (don't) learn: Counterfactual thinking and organizational accountability after a close call. *Administrative Science Quarterly*, 45(4):737–765, 2000.

[180] Andres Ruiz-Tagle, Enrique Lopez-Droguett, and Katrina M Groth. A novel probabilistic approach to counterfactual reasoning in system safety. *Reliability Engineering & System Safety*, 228:108785, 2022.

[181] David Lewis. Counterfactuals and comparative possibility. In *Ifs: Conditionals, Belief, Decision, Chance and Time*, pages 57–85. Springer, 1973.

[182] Peter Menzies and Helen Beebee. Counterfactual Theories of Causation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2020 edition, 2020.

[183] Xiaowei Li, Tiezhong Liu, and Yongkui Liu. Cause analysis of unsafe behaviors in hazardous chemical accidents: Combined with HFACs and Bayesian network. *International Journal of Environmental Research and Public Health*, 17(1):11, 2020.

[184] Nini Xia, Patrick XW Zou, Xing Liu, Xueqing Wang, and Runhe Zhu. A hybrid BN-HFACS model for predicting safety performance in construction projects. *Safety Science*, 101:332–343, 2018.

[185] Alexander Balke and Judea Pearl. Counterfactual probabilities: Computational methods, bounds and applications. In *10th Uncertainty in Artificial Intelligence Conference Proceedings*, pages 46–54. Elsevier, 1994.

[186] Evangelia Kyrimi. *Bayesian Networks for Clinical Decision Making: Support, Assurance, Trust*. PhD thesis, Queen Mary University of London, 2019.

[187] Anthony Costa Constantinou, Barbaros Yet, Norman Fenton, Martin Neil, and William Marsh. Value of information analysis for interventional and counterfactual Bayesian networks in forensic medical sciences. *Artificial Intelligence in Medicine*, 66:41–52, 2016.

[188] Ankur Ankan and Abinash Panda. `pgmpy`: Probabilistic graphical models using python. In *Proceedings of the 14th Python in Science Conference (SCIPY 2015)*, pages 6–11, 2015.

[189] Gordon Woo. Downward counterfactual search for extreme events. *Frontiers in Earth Science*, 7:340, 2019.

[190] Roland Wenzlhuemer. Counterfactual thinking as a scientific method. *Historical Social Research/Historische Sozialforschung*, pages 27–54, 2009.

[191] Paul M Salmon, Miranda Cornelissen, and Margaret J Trotter. Systems-based accident analysis methods: A comparison of Accimap, HFACS, and STAMP. *Safety Science*, 50(4):1158–1170, 2012.

[192] Pipeline and Hazardous Materials Safety Administration. Gas transmission & gathering incident data - January 2010 to present. Technical report, U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration (PHMSA), 2021.

[193] OSHA. U.S. Department of Labor cites utility contractors following fatal explosion in Wisconsin. *OSHA News Release - Region 5*, 2019.

[194] Jean-Christophe Le Coze. Storytelling or theory building? Hopkins' sociology of safety. *Safety Science*, 120:735–744, 2019.

[195] Rachel Smallman and Amy Summerville. Counterfactual thought in reasoning and performance. *Social and Personality Psychology Compass*, 12(4):e12376, 2018.

[196] Martin Neil, Norman Fenton, David Lagnado, and Richard David Gill. Modelling competing legal arguments using Bayesian model comparison and averaging. *Artificial Intelligence and Law*, 27(4):403–430, 2019.

[197] Sérgio Barros da Cunha. A review of quantitative risk assessment of onshore pipelines. *Journal of Loss Prevention in the Process Industries*, 44:282–298, 2016.

[198] Michael Baker. Mechanical damage: Final report. Technical Report DTRS56-02-D-70036, U.S. Department of Transportation Pipeline and Hazardous Materials Safety Administration Office of Pipeline Safety, 2009.

[199] Shu-jiao Tong, Siu-ming Lo, Pei-hong Zhang, and Bao-zhi Chen. Jet fire consequence evaluation on the natural gas transported by pipelines. *Procedia Engineering*, 52:349–354, 2013.

[200] PHMSA. Pipeline mileage and facilities, 2020. Online; accessed November-2022.

[201] Irfan Ahmad Gondal and Mukhtar Hussain Sahir. Prospects of natural gas pipeline infrastructure in hydrogen transportation. *International Journal of Energy Research*, 36(15):1338–1345, 2012.

[202] Katrina M Groth and Ahmad Al-Douri. Hydrogen safety, risk, and reliability analysis. In *Hydrogen Economy Processes, Supply Chain, Life Cycle Analysis and Energy Transition for Sustainability*, pages 487–510. Academic Press, 2023.

[203] Abubakar Jibrin Abbas, Hossein Hassani, Martin Burby, and Idoko Job John. An investigation into the volumetric flow rate requirement of hydrogen transportation in existing natural gas pipelines and its safety implications. *Gases*, 1(4):156–179, 2021.

[204] Arun S. K. Raju, Alfredo Martinez-Morales, and Oren Lever. Hydrogen blending impacts study. Technical Report R1302008, The California Public Utilities Commission, 2022.

[205] API. Specification 5L, 46th edition. Technical report, American Petroleum Institute, 2022.

[206] ASTM. DS513-20 Standard specification for thermoplastic gas pressure pipe, tubing, and fittings. Technical report, American Society for Testing and Materials, 2020.

[207] PHMSA. Code of Federal Regulation 49 CFR Part 192. Technical report, Pipeline and Hazardous Materials Safety Administration, 2022.

[208] Brian D. Ehrhart, Ethan Hecht, and Katrina M. Groth. Hydrogen Risk Assessment Models (HyRAM) (v.3.1) (technical reference manual). Technical Report SAND-2021-5812 696344, Sandia National Lab, 2021.

[209] Daniel A Crowl and Joseph F Louvar. *Chemical process safety: fundamentals with applications*. Pearson Education, 2001.

[210] A Tchouvelev, R Hay, and P Benard. Comparative risk estimation of compressed hydrogen and cng refuelling options. Technical report, Canadian Hydrogen Association workshop on building Canadian strength with hydrogen systems, 2006.

[211] Donald B Rubin. The Bayesian bootstrap. *The Annals of Statistics*, pages 130–134, 1981.

[212] Dmitriy Sergeyevich Gustov and Evgeniy Anatolevich Lyubin. Investigating the possibility of using fiberglass reinforced polyethylene pipe for the transportation of natural gas. *Indian Journal of Science and Technology*, 9(14):1–11, 2016.

[213] BayesFusion. SMILE engine, version 2.0.9 academic, 2022.

[214] Python. Python, version 3.7, 2019.

[215] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.

[216] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[217] Ibsen Chivata Cardenas, Roger Flage, and Terje Aven. Marine geohazards exposed: Uncertainties involved. *Marine Georesources & Geotechnology*, pages 1–31, 2022.

[218] Tsai-Ching Lu and Marek J Druzdzel. Causal models, value of intervention, and search for opportunities. In *Advances in Bayesian Networks*, pages 121–135. Springer, 2004.

[219] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

[220] Michael Adamo and Kristine Wiley. Quantifying methane emissions from natural gas residential customer meters in california. Technical Report 22049, GTI Energy, 2020.

[221] John Lathrop, Willy Roed, Steven Ackerlund, Robert Waller, Terje Aven, Roger Flage, Ghaidaa Hetou, Richard Belzer, Amina Wilkins, Tim Trevan, Tes Yellman, Ted Simon, Dikmen Irem, Samuel Denard, Emma Soane, Tim Smith, Debra Decker, Patricia Larkin, and Ronald Dyer. *SRA Risk Analysis Quality Test*. Society for Risk Analysis, 2021.

[222] Haoyuan Zhang and D William R Marsh. Managing infrastructure asset: Bayesian networks for inspection and maintenance decisions reasoning and planning. *Reliability Engineering & System Safety*, 207:107328, 2021.