INFRASTRUCTURE ASSET MANAGEMENT ANALYTICS STRATEGIES FOR SYSTEMIC RISK MITIGATION AND RESILIENCE ENHANCEMENT

INFRASTRUCTURE ASSET MANAGEMENT ANALYTICS STRATEGIES FOR SYSTEMIC RISK MITIGATION AND RESILIENCE ENHANCEMENT

By

Eric Goforth, B.Eng.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

Doctor of Philosophy (2022) (Civil Engineering)	McMaster University Hamilton, Ontario
TITLE:	Infrastructure asset management analytics strategies for systemic risk mitigation and resilience enhancement
AUTHOR:	Eric Goforth B.Eng. (McMaster University)
SUPERVISORS:	Dr. Wael El-Dakhakhni Dr. Lydell Wiebe
NUMBER OF PAGES:	xix, 191

Dedications

To my wife Carly

Lay Abstract

Effective infrastructure asset management systems are critical for organizations that own, manage, and operate infrastructure assets. Infrastructure asset management systems contain main components (e.g., engineering, project management, resourcing strategy) that are dependent on information and data. Inherent within this system is the potential for failures to cascade throughout the entire system instigated by such dependence. Within asset management, such cascading failures, known as systemic risks, are typically caused by stakeholders not using the same information for decision making or being overwhelmed by too much information. This thesis employs analytics strategies including: *i)* descriptive analytics to present only relevant and meaningful information necessary for respective stakeholders, *ii)* predictive analytics to forecast the resilience key performance indicator, rapidity, enabling all stakeholders to make future decisions using consistent projections, and *iii)* prescriptive analytics to optimize the asset management system by introducing additional information connections between main components. Such analytics strategies are shown to mitigate the systemic risks within the asset management system and enhance the resilience of infrastructure in response to an unplanned disruption.

Abstract

The effective implementation of infrastructure asset management systems within organizations that own, operate, and manage infrastructure assets is critical to address the main challenges facing the infrastructure industry (e.g., infrastructure ageing and deterioration, maintenance backlogs, strict regulatory operating conditions, limited financial resources, and losing valuable experience through retirements). Infrastructure asset management systems contain connectivity between major operational components and such connectivity can lead to systemic risks (i.e., dependence-induced failures). This thesis analyzes the asset management system as a network of connected components (i.e., nodes and links) to identify critical components exposed to systemic risks induced by information asymmetry and information overload. This thesis applies descriptive and prescriptive analytics strategies to address information asymmetry and information overload and predictive analytics is employed to enhance the resilience. Specifically, descriptive analytics was employed to visualize the key performance indicators of infrastructure assets ensuring that all asset management stakeholders make decisions using consistent information sources and that they are not overwhelmed by having access to the entire database. Predictive analytics is employed to classify the resilience key performance indicator pertaining to the forced outage rapidity of power infrastructure components enabling power infrastructure owners to estimate the rapidity of an outage soon after its occurrence, and thus allocating the appropriate resources to return the infrastructure to operation. Using predictive analytics allows decision-makers to use consistent and clear information to inform their decision to respond to forced outage occurrences. Finally, prescriptive analytics is applied to optimize the asset management system network by increasing the connectivity of the network and in turn decreasing the exposure of the asset management system to systemic risk from information asymmetry and information overload. By analyzing an asset management system as a network and applying descriptive-, predictive-, and prescriptive analytics strategies, this dissertation illustrates how systemic risk exposure, due to information asymmetry and information overload could be mitigated and how power infrastructure resilience could be enhanced in response to forced outage occurrences.

Acknowledgement

I would like to thank my supervisors Dr. Wael El-Dakhakhni and Dr. Lydell Wiebe for their guidance over these past years. I appreciated the discussion about complex ideas and brainstorming ways to solve them. I learn a lot from both of you that I will take into my career for the rest of my life. I would also like to thank my supervisory committee members Dr. Peijun Guo and Dr. Elkafi Hassini for their insights and constructive feedback throughout my work. In addition, I would like to thank Dr. Arnold Yuan for the great feedback and discussion as the external examiner of my Ph.D. defence.

I would also like to acknowledge Dr. Ahmed Yosri and Dr. Mohamed Ezzeldin for their insight and expertise that helped me take ideas and synthesize them into publishable material.

I have enjoyed so many discussions about work and life with Moustafa Naiem Abdel-Mooty. Your friendship has been very special to me, and we kept each other sane throughout the pandemic. I am so thankful to all my friends that have supported me and drove me to be my best throughout this journey.

I am grateful to the Canadian Electricity Association and especially Dan Gent for granting me access to your data so that my research can have tangible impacts on the real world.

My parents, Dennis and Karen Goforth, have always encouraged me to do my best in everything that I do and that I can accomplish anything that I put my

mind to. They were there to listen when things were not going well, and they were there to celebrate great achievements. Thank you for your encouragement and prayers. Thank you for your sacrifices to raise me to strive for my goals. To my in-laws, Elden and Julie Westendorp, thank you for your encouragement and prayers. To my brothers, Kyle and Evan, and my sister-in-laws Erica, Shaelyn, and Justine, thanks for always being around for a laugh and to make life more fun!

I would not be here without the constant support and sacrifices made by my wife Carly Goforth. She has been present for all ups and downs. Her love and encouragement has kept me motivated to finish this Ph.D. and move forward to realize my dreams. We are a true team and together we can accomplish all that God has in store for our lives (Colossians 3:17-20).

To my Lord and Saviour Jesus Christ in whom all things are made that have been made. Thank you for giving me this opportunity to explore more about your creation. Thank you for answering my prayers when I was struggling and for rejoicing with me when I succeeded. I can accomplish nothing without your strength that fulfills and sustains me.

Table of contents

List of Figures

List of Tables

Declaration of Academic Achievement

This dissertation was prepared in accordance with the guidelines set by the school of graduate studies at McMaster University for sandwich theses that contain a compilation of research papers published or prepared for publishing as journal articles. This dissertation presents the work carried out solely by Eric Goforth, where technical advice and guidance were provided for the whole thesis by the academic supervisors Dr. Wael El-Dakhakhni and Dr. Lydell Wiebe. Further editorial comments and modeling discussion for the papers presented in Chapters 3 and 4 were provided by Dr. Ahmed Yosri. Information from outside sources, which has been used towards analysis or discussion, has been cited where appropriate. The original contributions of the author to each paper (Chapter) and the reasons for including them in this dissertation are outlined below:

Chapter 2: Eric Goforth, Wael El-Dakhakhni, Lydell Wiebe, "**Network Analytics for Infrastructure Asset Management Systemic Risk Assessment.**" *Journal of Infrastructure Systems*. Accepted for publication October 2021. DOI: 10.1061/(ASCE)IS.1943-555X.0000667.

The idea for the paper came from Eric Goforth through discussions with Wael El-Dakhakhni with the former preparing the five tool descriptions and demonstrating the implementation of Tools 1, 2, and 3. Eric Goforth prepared the manuscript, and the manuscript was reviewed and further edited by Wael El-Dakhakhni and Lydell Wiebe. This work is included in this dissertation as it is the first study to view the asset management system as a network of connected subject areas and it demonstrates how a network understanding can facilitate identifying systemic risks, caused by information asymmetry and information overload, that threaten the operation of an effective asset management system within an infrastructure owning, operating, or managing organization.

Chapter 3: Eric Goforth, Ahmed Yosri, Wael El-Dakhakhni, Lydell Wiebe, "**Rapidity Prediction For Power Infrastructure Forced Outages: A Data-driven Approach.**" *Journal of Energy Engineering*. Accepted for publication January 2022. DOI: 10.1061/(ASCE)EY.1943-7897.0000836.

The idea for the paper came from Eric Goforth through discussions with Wael El-Dakhakhni and Ahmed Yosri. The dataset employed in this paper came from the Canadian Electricity Association. Eric Goforth conducted the analysis and prepared the manuscript that was then further refined and edited by Ahmed Yosri, Wael El-Dakhakhni, and Lydell Wiebe. This work is included in this dissertation as it provides a framework that can be used to predict the resilience metric of rapidity for power infrastructure forced outages soon after the occurrence of an outage. The framework was applied successfully using transmission line forced outages. This work implements the concept presented as Tool 4 in Chapter 2 and applies it using real historical data that is collected by current power utilities.

Chapter 4: Eric Goforth, Ahmed Yosri, Wael El-Dakhakhni, Lydell Wiebe, "**An Optimal Configuration of Infrastructure Asset Management Systems Using a Genetic Algorithm and Complex Network Theory Approach.**" Submitted for publication to the *Journal of Infrastructure Systems* in January 2022.

The idea for the paper came from Eric Goforth through discussions with Wael El-Dakhakhni. Eric Goforth conducted the analysis and prepared the manuscript that was then further refined and edited by Ahmed Yosri, Wael El-Dakhakhni, and Lydell Wiebe. This work is included in this dissertation as it presents optimal configurations of an asset management system network that reduces the systemic risks caused by information asymmetry and information overload within a typical asset management system by adding connections. This work implements the concept presented as Tool 5 from Chapter 2 and links a typical asset management system network to current ISO 55001 standards highlighting the functionality of each asset management system component to the delivery of the overall asset management system objectives.

Chapter 1

INTRODUCTION

1.1. MOTIVATION

Infrastructure forms the backbone of human society, allowing for the distribution of resources and essential services to the general public. There are many different types of infrastructure that perform varying functions including power, water, sewage, transportation, telecommunications, facilities/housing, airports, marine ports, and agriculture. The quality and the operational effectiveness of such infrastructure systems impacts people's quality of life, the health of social systems, and the stability of economic and business activities (Uddin et al. 2013). For these reasons it is critical to have well maintained and well operating infrastructure systems. The Canadian Infrastructure Report Card (2019) graded the state of eight public infrastructure categories to be poor and the current state of infrastructure in the United States, as graded by the American Society of Civil Engineers (ASCE) in 2021, was a C- on average among 17 different infrastructure categories (American Society of Civil Engineers 2021). The ASCE 2021 report specifies that the total infrastructure investment gap to maintain a state of good repair among the 17 categories and earn a B grade is $2.59 trillion over 10 years (American Society of Civil Engineers 2021). That number increased from $2.1 trillion over 10 years as specified in the ASCE 2017 report. These condition assessments and investment gap for infrastructure across Canada and the United States highlight the need for effective management strategies considering whole lifecycle infrastructure spending to ensure the limited available resources are optimally allocated. Both the ASCE and the Canadian Infrastructure Report Card specify effective asset management (AM) practices as critical to address the main challenges faced by infrastructure asset owning organizations (e.g., infrastructure ageing and deterioration, maintenance backlogs, strict regulatory operating conditions, limited financial resources, and losing valuable experience through retirements) (Canadian Infrastructure Report Card 2019; American Society of Civil Engineers 2021).

1.2. Asset Management

AM is a strategic discipline that provides thoroughness and accountability to an organization in the decisions made throughout the whole lifecycle of infrastructure assets (Lloyd 2010). AM involves an integrated approach for forecasting into the future and inspecting the past to balance the needs of all stakeholders from the inception of an asset to its eventual disposal (Lloyd 2010). The Institute for Asset Management (IAM) presented a model of AM that connects major components of *Strategy & Planning*, *Asset Management Decision-Making*, *Lifecycle Delivery*, *Asset Information*, *Organization & People*, and *Risk & Review* as shown in Figure 1-1 (Institute for Asset Management 2015). The Federation of Canadian Municipalities and the National Research Council of Canada present AM as the combination of management, financial, economic, engineering, operational and other practices applied to physical assets with the objective of providing the required level of service in the most cost-effective manner (Federation of Canadian Municipalities and National Research Council 2005). These definitions of AM are applicable to all types of infrastructure asset owning, operating, and managing organizations and describe the system-nature of AM as each AM component is reliant on one or multiple component's information (International Organization for Standardization 2014; Canadian Network of Asset Managers 2018; United Nations 2021). To address the connectedness between AM components, organizations that own, manage, and operate infrastructure assets typically implement AM systems to achieve their organizational strategic plan and objectives (Hodkiewicz 2015).

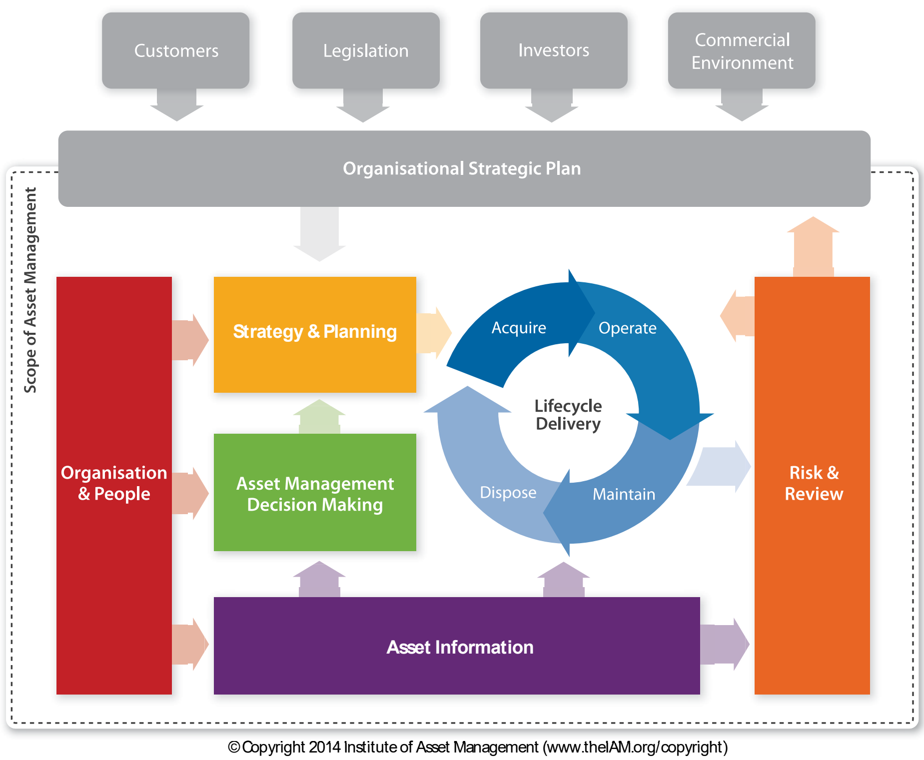


Figure 1-1. IAM's *six-box model* for infrastructure AM (Institute for Asset Management 2015).

An AM system is a set of interrelated and interacting elements within an organization, whose function is to establish the AM policy and AM objectives, and the processes needed to achieve those objectives (International Organization for Standardization 2014). An AM system is composed of components contributing to the functionality to the AM system (e.g., engineering, project management, operations, maintenance, strategic planning) (First State Investments 2010; Institute of Asset Management 2014). In order for an AM system to be efficient and effective, each component must fully deliver its intended function and value (Pell et al. 2015). When a component becomes dysfunctional and does not provide its intended value, an AM system ceases to operate as originally intended and does not provide the intended functionality for an organization (de la Pena et al. 2016).

1.3. SYSTEMIC RISK

Within the financial system a systemic risk is when one or multiple components of a system have the potential to fail, leading to potentially larger cascading failures throughout the system (López-Espinosa et al. 2015; Miller 2017). Any complex system that includes connected components is exposed to systemic risks. For example systemic risks were explored in bridge interconnection systems (Alzoor et al. 2021) and the power grid interconnection systems (Ezzeldin and El-Dakhakhni 2019). Such applications enable decision-makers to identify critically vulnerable components and develop strategies to mitigate systemic risks. Organizations that own, operate, and manage infrastructure assets have challenges that induce such systemic risks within their AM system, including information asymmetry between AM stakeholders and among the major components of an AM system, and information overload among AM stakeholders. Information asymmetry occurs when one party in a relationship has more or better quality information than anther (Bergh et al. 2019). Within an AM system, information asymmetry includes stakeholders that use different information to support their decision-making, not responding to other stakeholders' decisions promptly, and the isolation of AM system components due to inadequate information-sharing procedures or protocols. Information overload can also occur when a stakeholder is overwhelmed by the data and information resources accessible, and they have difficultly identifying the correct information needed to make a decision and therefore they do not make a decision or make a poor decision using the wrong information. Both situations may lead to a systemic risk situation as hindered or wrong decisions may instigate other connected AM system components using missing or incorrect information to make decisions and further cascading the dysfunctionality throughout the AM system. An example of a systemic risk in an AM system it that if the engineering component of the AM system were to become dysfunctional because it used incorrect future data projections, then the AM system would cease to operate as it was originally intended, as other components within the AM system are reliant on the information provided by the engineering component (e.g., capital project delivery and maintenance).

Numerous resources outline the main components that form a typical AM system (Federation of Canadian Municipalities and National Research Council 2005; Global Forum on Maintenance and Asset Management 2014; Institute for Asset Management 2015) while others highlight the importance of organizing AM as a system of connected components (First State Investments 2010; Hodkiewicz 2015; Institute for Asset Management 2015; United Nations 2021). However, to the best of the author's knowledge, a connected AM system network model to identify systemic risk associated with information asymmetry and information overload within an AM system does not exist. Therefore, the identification and reduction of systemic risk within an AM system necessitates: *i)* the development of a typical AM system model as a network of connected components (i.e., nodes and links), which is essential for the identification of systemic risks within the AM system; and *ii)* the deployment of analytics strategies to mitigate the information asymmetry and information overload effects within a typical AM system.

1.4. ASSET MANAGEMENT SYSTEM MODEL

There is not a one-size-fits-all approach to infrastructure AM, but the Institute for Asset Management (IAM) developed the guideline *Asset Management – An Anatomy* to help individuals better understand the AM discipline, where to find more information, and what to do next in their AM journey (Institute for Asset Management 2015). In addition, *Asset Management – An Anatomy* was developed to help organizations with deciding whether to adopt AM and/or to improve their

AM capability (Institute for Asset Management 2015). The IAM is the international professional body for AM professionals that generates AM knowledge and resources, best practice guidelines, and awareness of the benefits of the AM discipline for individuals, organizations, and wider society (Institute for Asset Management 2021).The IAM's approach to view infrastructure AM as a system of connected subject areas is widely deployed within infrastructure owning, operating, and managing organizations around the world (Zuashkiani et al. 2014; CIGRE WG C1.38 2020).

The IAM's *Anatomy* can be represented as the *six-box model* as shown in Figure 1-1. Each component of the *six-box model* is a major division of an AM system within an organization (i.e., *Strategy & Planning*, *Asset Management Decision-Making*, *Lifecycle Delivery*, *Asset Information*, *Organization & People*, and *Risk & Review*). *Strategy & Planning* develops a plan which aligns the organization's AM activities and all AM stakeholders toward a common goal. *Asset Management Decision-Making* makes AM strategies and evaluates challenges related to each aspect of an asset's lifecycle including asset acquisition/creation, operation, maintenance, and end of life disposal, decommissioning, or renewal. *Lifecycle Delivery* involves the delivery of an asset's service and value over its lifespan, from acquisition/creation, through operation/maintenance, and finally, to end of life disposal, decommissioning, or renewal. *Asset Information* involves the standards, strategy, management, and systems used for information and data that is employed by the other AM divisions. *Organization & People* involves a review of the organizational structure, roles, responsibilities, and contractual relationships. *Risk & Review* identifies the risks related to an asset's lifecycle delivery, understands and manages such risks, and continues the improvement and development of AM activities according to the strategy and plan.

The six major divisions are further broken down by the *Global Forum's Asset Management Landscape* into 39 subject areas (i.e., Figure 1-2) that describe the breadth of activities within the scope of AM, the relationships between activities, the need to integrate such relationships, and the critical role for AM to align with and deliver the strategic goals of an asset-intensive organization (Global Forum on Maintenance and Asset Management 2014; Institute for Asset Management 2015). It is these AM subject areas that form the main components of a typical AM system within an organization.

Figure 1-2. 39 AM subject areas illustrating the extent of AM activities within an AM system.

1.5. RESILIENCE

For the purposes of this thesis, resilience of infrastructure components will be discussed and quantified with respect to their goals (i.e., robustness and rapidity) as well as their means (i.e., redundancy and resourcefulness) (Bruneau et al. 2003; Panteli and Mancarella 2015b; Gholami et al. 2018; Salem et al. 2020). In this respect, *robustness* is defined as the ability of a component to maintain operation while experiencing disruptions; *rapidity* is the time taken to recover from such disruptions and return to the normal (or near normal) operation levels; *redundancy* is the capability to deliver the intended function provided that some components have experienced a disruption, degradation, or loss of functionality; and finally, *resourcefulness* is the capacity to restore service such that the component could return to a normal operation level (Bruneau et al. 2003; Bie et al. 2017). The relationship among the resilience goals is presented in Figure 1-3, where the *resilience trapezoid* indicates the robustness as the percentage of the remaining functioning components and the rapidity as the total time to return to the pre-disruption operation level (Jufri et al. 2019).

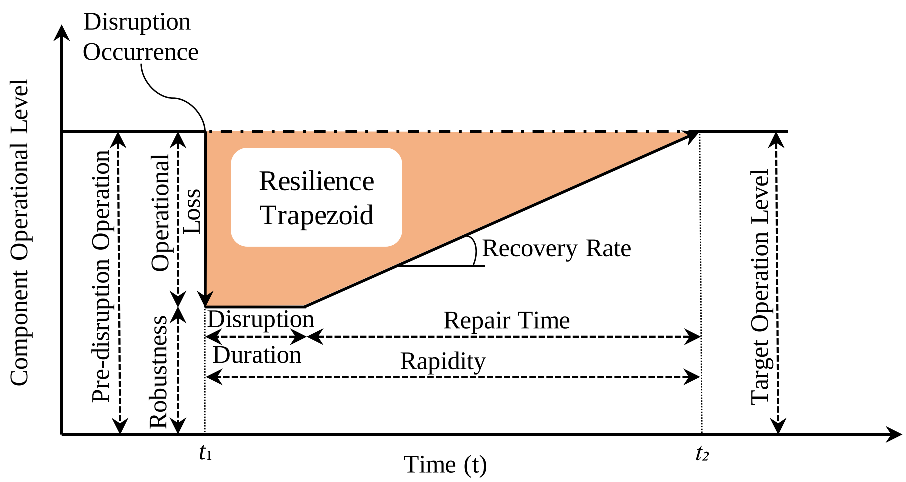


Figure 1-3. Resilience trapezoid.

1.6. Analytics in Infrastructure Asset Management

Analytics involves the systematic computational analysis of data for the discovery, interpretation, and communication of meaningful insights to enable effective decision-making (Barker et al. 2017; Delen and Ram 2018; Haggag et al. 2021; Tsai et al. 2015). In recent years, organizations have transformed their decision-making from traditional experience-based to data-driven in an attempt to drive improved value for their organizations (Delen et al. 2018; O'Neill and Brabazon 2019; Scheibe et al. 2019). Leveraging data-driven analytics methods for decision-making has allowed AM maintenance processes to employ predictive or preventative maintenance as opposed to reactive maintenance (Bertling Tjernberg 2018). This means that maintenance decisions are driven by the data the asset is producing and the condition of the asset as opposed to a standardized schedule or when the asset fails (Mehairjan 2017; Qiu et al. 2013). Using and applying analytics effectively to make useful decisions requires good data sources as erroneous data can lead to bad decisions even when the same analytics methods are employed (Koziel et al. 2021). Therefore, the first step all organizations looking to implement analytics for improved data-driven decision-making need to make is to invest time and resources to ensure the data they use is high quality and is measuring the correct features (Levene et al. 2018; de Sousa et al. 2018). Once a quality data source is obtained, analytics is often described as either descriptive, predictive, or prescriptive and each will be described in the following sections (IBM Watson IoT 2017). This thesis primarily focuses on power infrastructure analytics applications in AM.

1.6.1. Descriptive Analytics

Descriptive analytics collects, organizes, and presents current or historical data in a way that is easily understood (Vesset 2018). Primarily descriptive analytics focuses on what has occurred in the past or what is currently happening with the data (Delen et al. 2018). Organizations typically have key performance indicators (KPIs) that illustrate their critical metrics used to measure the performance of their organizations and the assets they own, manage, and operate. Often descriptive analytics leverages visual applications to display such KPIs, allowing for the resulting insights to be accessed by a wide audience (IBM Watson IoT 2017). Insights obtained from descriptive analytics illustrate past performance and allow an organization to measure and investigate such past performance to identify areas that are performing well or areas that need improvement or change (Delen and Ram 2018). These performance insights should be easy to obtain within a well designed descriptive analytics application as opposed to searching for results through tables or poorly designed figures (Levene et al. 2018). For example, within power infrastructure applications, descriptive analytics has been deployed to investigate the reliability metrics of transmission power assets (Bian et al. 2014; Ekisheva et al. 2016, 2018; Ekisheva and Gugel 2015a; b; Papic et al. 2014, 2017, 2018; Schaller 2012; Schaller and Ekisheva 2016), the statistical properties of transformer forced outages (Abdelfatah et al. 2013), and weather event influence on power infrastructure outages (Black et al. 2018; Mukherjee et al. 2018). Within each of the power infrastructure application examples, descriptive analytics was deployed to identify historical KPI performance to support decisions where investment was needed to improve performance and identify assets that were exceeding or performing according to expectations.

1.6.2. PREDICTIVE ANALYTICS

Predictive analytics is the process by which trends and patterns in historical data are formulated in a mathematical model to predict future outcomes (IBM Watson IoT 2017). Primarily, predictive analytics builds machine learning models that use past historical performance and other influencing feature data to forecast possible outcomes and the likelihood of such outcomes (Goyal et al. 2016; Heng et al. 2016). These machine learning models are generally classified as supervised (i.e., developing a mathematical function that maps the relationship between input-output pairs) or unsupervised (i.e., categorizing the dataset based on similarity, without specifying any certain outputs) (Zumel and Mount 2020). A predictive analytics methodology is known as classification if the output feature is a categorical feature, otherwise it is known as regression (e.g., numerical output features) (Delen and Ram 2018). Complete and informative input data is necessary to train a machine learning model to predict a specified numerical or categorical output feature (Hastie et al. 2009). Such input data includes informative contributing features to the output feature (e.g., climatic conditions, economic conditions, geographic locations, asset characteristics, and maintenance history) and past historical performance of related KPI values (Haggag et al. 2021). For example in power infrastructure applications, predictive analytics has been deployed to predict specific KPI including the number of power outages (Dokic et al. 2019; Nateghi 2018), the number of customers without power (Nateghi 2018), the distribution infrastructure outage durations (Eskandarpour et al. 2017), and the power outage severity (Mukherjee et al. 2018). Each study used a unique set of input features, based on the accessible data, within their machine learning model to predict the indicated output feature. The predictions from a machine learning model can be used by decision-makers as proactive information on potential future scenarios, enabling a more data-driven approach to future decision-making (IBM Watson IoT 2017).

1.6.3. PRESCRIPTIVE ANALYTICS

Prescriptive analytics is the process that explores a set of possible actions and decisions and suggests the optimal result based on the analysis of the available complex data (Barker et al. 2017). Specifically, prescriptive analytics leverages the advanced capabilities of optimization and mathematical models to propose recommended actions and decisions and the reasons why such actions and decisions are recommended and any implications resulting from implementation (IBM Watson IoT 2017). Prescriptive analytics is the most complex of the analytics options, but when implemented correctly, it can provide the greatest value to decision-makers (Delen and Ram 2018). For example in AM infrastructure applications, prescriptive analytics has primarily been used to optimize the schedule, risk, and cost for repair, restoration, or rehabilitation of infrastructure assets (Abu-Samra et al. 2020; Goyal et al. 2016) and in maintenance planning and scheduling (Chen et al. 2015). Prescriptive analytics has the potential to offer the most benefit to AM decision-makers by sifting through solution spaces that are too large or complex to be understood without an analytics application, but it also requires the highest quality data to obtain results that can be relied upon to support decision-making (IBM Watson IoT 2017).

1.7. RESEARCH OBJECTIVES

The main goal of the work presented in this dissertation is to present analytics strategies that can be used to mitigate the information asymmetry and information overload systemic risks that plague an AM system and to enhance the resilience of infrastructure in response to unplanned disruptions. These analytics strategies describe a descriptive analytics approach to improve the presentation of data and information while also ensuring stakeholders are not overwhelmed by large volumes of information, a predictive analytics approach to forecast a resilience KPI for future scenario planning ensuring that all AM stakeholders make decisions using common forecasted models, and a prescriptive analytics approach for minimizing the exposure of an AM system to systemic risk by adding additional information connections throughout the AM system. As such, the scope of the thesis has been set to achieve the following objectives:

- Develop a network model including the major subject areas of a typical AM system and present descriptive analytics for the critical AM subject areas and associated KPIs to reduce the exposure of AM stakeholders to information asymmetry and information overload.

- Develop a predictive analytics framework that could be deployed to enhance the resilience of power infrastructure organizations by classifying the resilience metric rapidity of a power infrastructure component forced outage soon following the outage occurrence.
- Develop a prescriptive analytics model that minimizes the exposure of an AM system to systemic risk from information asymmetry and information overload by increasing the information connections throughout the AM system network.

1.8. THESIS ORGANIZATION

This section summarizes the content of each of the five chapters in this dissertation:

- Chapter 1 provides the background required for this research, an overview of the objectives, and a description of the thesis organization.
- Chapter 2 introduces a toolbox that can be used by asset-owning organizations in the development and implementation of their AM system to enhance the capability of the AM system and reduce the expose of the AM system to systemic risk. The five tools are: 1) Dependence identification and network modelling; 2) Network centrality analysis; 3) Descriptive analytics of critical subject area paired KPI; 4) KPI-based predictive analytics; and 5) Prescriptive analytics for optimal network configuration. The utility of the developed toolbox is shown for Tools 1, 2,

and 3 using a real AM system network and the KPI from power transmission infrastructure outages. The chapter shows how the toolbox improves the information symmetry within the AM system—subsequently mitigating the dependence-induced systemic risk.

- Chapter 3 outlines a framework that describes the process of employing a consistent data source to predict a resilience KPI, ensuring that all decision-makers use the same forecasted information in future decision-making. Specifically, the framework could be deployed by power utilities to classify the KPI rapidity of power infrastructure component forced outages soon after their occurrence. To demonstrate the framework applicability, it is deployed using transmission power infrastructure forced outage data from the Canadian Electricity Association (CEA). Rapidity classification results are obtained specific to five individual power utilities annually for five predicted years. The validity of the framework is confirmed as an all categorical input feature set was able to accurately classify the categorical rapidity output feature.
- Chapter 4 presents a method for reduction in systemic risk within an AM system caused by information asymmetry and information overload. The AM system is represented as a network of connected components (i.e., nodes and links) with the functionality of each component (i.e., link weight) as the number of ISO 55001 clause connections to each component. Optimized AM system network configurations are found for multiple

centrality measures based on a specified number of link additions allowable using a genetic algorithm approach. The optimal network configurations reduce the exposure of an AM system to systemic risk by increasing the resilience of the network.

- Chapter 5 provides a summary of this research, the overall contributions, and suggestions for future work.

It should be noted that Chapters 2, 3, and 4 represent standalone manuscripts that are already published, submitted, or prepared for publishing as journal articles. These chapters collectively describe the research body as outlined in the research objectives of the introduction chapter of this dissertation. However, some overlap may exist between chapters for the completeness of each manuscript for its acceptance as a standalone journal paper.

1.9. ACRONYMS

ASCE	American Society of Civil Engineers
AM	Asset Management
CEA	Canadian Electricity Association
KPI	Key Performance Indicator
IAM	Institute for Asset Management
ISO	International Organization for Standardization

1.10. REFERENCES

Abdelfatah, M., El-Shimy, M., and Ismail, H. M. (2013). “Outage data analysis of utility power transformers based on outage reports during 2002-2009.” International Journal of Electrical Power and Energy Systems, Elsevier, 47(1), 41–51.

Abu-Samra, S., Ahmed, M., and Amador, L. (2020). “Asset Management Framework for Integrated Municipal Infrastructure.” Journal of Infrastructure Systems, 26(4), 04020039.

Alzoor, F. S., Ezzeldin, M., Mohamed, M., and El-Dakhakhni, W. (2021). “Prioritizing Bridge Rehabilitation Plans through Systemic Risk-Guided Classifications.” Journal of Bridge Engineering, 26(7), 04021038.

American Society of Civil Engineers. (2021). 2021 Report Card for America’s Infrastructure.

Barker, K., Lambert, J. H., Zobel, C. W., Tapia, A. H., Ramirez-Marquez, J. E., Albert, L., Nicholson, C. D., and Caragea, C. (2017). “Defining resilience analytics for interdependent cyber-physical-social networks.” Sustainable and Resilient Infrastructure, Taylor & Francis, 2(2), 59–67.

Bergh, D. D., Ketchen, D. J., Orlandi, I., Heugens, P. P. M. A. R., and Boyd, B. K. (2019). “Information Asymmetry in Management Research: Past Accomplishments and Future Opportunities.” Journal of Management, 45(1), 122–158.

Bertling Tjernberg, L. (2018). Infrastructure Asset Management with Power System Applications. Taylor & Francis Group, Boca Raton.

Bian, J. J., Ekisheva, S., and Slone, A. (2014). “Top risks to transmission outages.” IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Bie, Z., Lin, Y., Li, G., and Li, F. (2017). “Battling the Extreme: A Study on the Power System Resilience.” Proceedings of the IEEE, 105(7), 1253–1266.

Black, J., Hoffman, A., Hong, T., Roberts, J., and Wang, P. (2018). “Weather Data for Energy Analytics: From Modeling Outages and Reliability Indices to Simulating Distributed Photovoltaic Fleets.” IEEE Power and Energy Magazine, 16(3), 43–53.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O’Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., and Von Winterfeldt, D. (2003). “A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities.” Earthquake Spectra, 19(4), 733–752.

Canadian Infrastructure Report Card. (2019). Monitoring the State of Canada’s Core Public Infrastructure: The Canadian Infrastructure Report Card 2019.

Canadian Network of Asset Managers. (2018). Asset Management 101 - The What, Why, and How For Your Community. Ottawa.

Chen, L., Henning, T. F. P., Raith, A., and Shamseldin, A. Y. (2015). “Multiobjective Optimization for Maintenance Decision Making in Infrastructure Asset Management.” Journal of Management in Engineering, 31(6), 04015015.

CIGRE WG C1.38. (2020). Valuation as a comprehensive approach to asset management in view of emerging developments. C1: Power system development and economics.

Delen, D., Moscato, G., and Toma, I. L. (2018). “The impact of real-time business intelligence and advanced analytics on the behaviour of business decision makers.” 2018 International Conference on Information Management and Processing, IEEE, 49–53.

Delen, D., and Ram, S. (2018). “Research challenges and opportunities in business analytics.” Journal of Business Analytics, Taylor & Francis, 1(1), 2–12.

Dokic, T., Pavlovski, M., Gligorijevic, D., Kezunovic, M., and Obradovic, Z. (2019). “Spatially Aware Ensemble-Based Learning to Predict Weather-Related Outages in Transmission.” Proceedings of the 52nd Hawaii International Conference on System Sciences, Hawaii International Conference on System Sciences, 1–10.

Ekisheva, S., and Gugel, H. (2015a). “North American transformer outage rates and durations in assessment of transmission system reliability and availability.” IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Ekisheva, S., and Gugel, H. (2015b). “North American AC circuit outage rates and durations in assessment of transmission system reliability and availability.” IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Ekisheva, S., Lauby, M. G., and Gugel, H. (2016). “North American transformer outages initiated by transmission equipment failures and human error.” IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Ekisheva, S., Papic, M., Clemons, M. S., Quest, R., Pakeltis, M. J., and Weisman, K. (2018). "Outage statistics, reliability and availability of DC circuits in North American Bulk Power System." IEEE Power and Energy Society General Meeting, IEEE Computer Society, 1–5.

Eskandarpour, R., Khodaei, A., and Arab, A. (2017). "Improving power grid resilience through predictive outage estimation." 2017 North American Power Symposium (NAPS), IEEE, Morgantown, WV, USA, 1–5.

Ezzeldin, M., and El-Dakhakhni, W. E. (2019). "Robustness of Ontario power network under systemic risks." Sustainable and Resilient Infrastructure, Taylor & Francis, 1–20.

Federation of Canadian Municipalities and National Research Council. (2005). Managing Infrastructure Assets: Decision Making and Investment Planning. InfraGuide.

First State Investments. (2010). Best Practice in Infrastructure Asset Management. (J. Altmann, ed.), PEI Media Ltd., London.

Gholami, A., Shekari, T., Amirioun, M. H., Aminifar, F., Amini, M. H., and Sargolzaei, A. (2018). "Toward a consensus on the definition and taxonomy of power system resilience." IEEE Access, IEEE, 6, 32035–32053.

Global Forum on Maintenance and Asset Management. (2014). The Asset Management Landscape. Global Forum on Maintenance and Asset Management.

Goyal, A., Aprilia, E., Janssen, G., Kim, Y., Kumar, T., Mueller, R., Phan, D., Raman, A., Schuddebeurs, J., Xiong, J., and Zhang, R. (2016). "Asset health management using predictive and prescriptive analytics for the electric power grid." IBM Journal of Research and Development, 60(1.4), 1–14.

Haggag, M., Yorsi, A., El-Dakhakhni, W., and Hassini, E. (2021). "Infrastructure performance prediction under Climate-Induced Disasters using data analytics." International Journal of Disaster Risk Reduction, Elsevier Ltd., 56(November 2020), 102121.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning. Springer Series in Statistics, Springer New York, New York, NY.

Heng, F.-L., Zhang, K., Goyal, A., Chaudhary, H., Hirsch, S., Kim, Y., Lavin, M. A., and Raman, A. (2016). "Integrated analytics system for electric industry asset management." IBM Journal of Research and Development, 60(1.2), 1–12.

Hodkiewicz, M. R. (2015). "The development of ISO 55000 series standards." Lecture Notes in Mechanical Engineering, 19, 427–438.

IBM Watson IoT. (2017). Descriptive, predictive, prescriptive : Transforming asset and facilities management with analytics. Somers, NY.

Institute for Asset Management. (2015). Asset Management – an Anatomy. Bristol, UK.

Institute of Asset Management. (2014). The IAM Competences Framework Requirements.

International Organization for Standardization. (2014). ISO 55000 — Asset management — Overview, principles and terminology. Geneva.

Jufri, F. H., Widiputra, V., and Jung, J. (2019). "State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies." Applied Energy, 239(January), 1049–1065.

Koziel, S., Hilber, P., Westerlund, P., and Shayesteh, E. (2021). "Investments in data quality: Evaluating impacts of faulty data on asset management in power systems." Applied Energy, Elsevier Ltd., 281(November 2020).

de la Pena, C., Gonzalez Fernandez, D., and Rodriguez Gonzalez, J. (2016). "How analytics can improve asset management in electric power networks." McKinsey & Company Electrical Power & Natural Gas, <https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/> (Jun. 14, 2018).

Levene, J., Litman, S., Schillinger, I., and Toomey, C. (2018). How advanced analytics can benefit infrastructure capital planning,. New Jersey.

Lloyd, C. (2010). Asset management - Whole-life management of physical assets. Thomas Telford Limited, London.

López-Espinosa, G., Moreno, A., Rubia, A., and Valderrama, L. (2015). "Systemic risk and asymmetric responses in the financial industry." Journal of Banking and Finance, 58, 471–485.

Mehairjan, R. P. Y. (2017). "Risk-Based Maintenance for Electricity Network Organizations."

Miller, B. B. (2017). “Information Asymmetries Conceal Fraud and Systemic Risks in the U.S. Banking Industry.” The Harvard Law School Forum on Corporate Governance, The Harvard Law School Forum on Corporate Governance, <https://corpgov.law.harvard.edu/2017/08/19/information-asymmetries-conceal-fraud-and-systemic-risks-in-the-u-s-banking-industry/> (Jul. 29, 2021).

Mukherjee, S., Nateghi, R., and Hastak, M. (2018). “A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S.” Reliability Engineering and System Safety, Elsevier Ltd, 175, 283–305.

Nateghi, R. (2018). “Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems.” IEEE Access, IEEE, 6, 13478–13489.

O’Neill, M., and Brabazon, A. (2019). “Business analytics capability, organisational value and competitive advantage.” Journal of Business Analytics, Taylor & Francis, 2(2), 160–173.

Panteli, M., and Mancarella, P. (2015). “The Grid: Stronger, Bigger, Smarter?” IEEE Power and Energy Magazine, IEEE, 13(May/June), 58–66.

Papic, M., Agarwal, S., Allan, R. N., Billinton, R., Dent, C. J., Ekisheva, S., Gent, D., Jiang, K., Li, W., Mitra, J., Pitto, A., Schneider, A., Singh, C., Vadlamudi, V. V., and Varghese, M. (2017). “Research on Common-Mode and Dependent (CMD) Outage Events in Power Systems: A Review.” IEEE Transactions on Power Systems, IEEE, 32(2), 1528–1536.

Papic, M., Agarwal, S., Bian, J., Billinton, R., Dent, C., Dobson, I., Jirutitijaroen, P., Li, W., Menten, T., Mitra, J., Schneider, A., Singh, C., and Vadlamudi, V. V. (2014). "Effects of dependent and common mode outages on the reliability of bulk electric system - Part II: Outage data analysis." IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Papic, M., Clemons, M., Ekisheva, S., Gent, D., Norris, J., Pakeltis, M., Peacock, M., and Till, D. (2018). "Performance analysis of North American AC circuits on common structures using TADS and CEA outage statistics." 2018 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), IEEE, 1–6.

Pell, R., Svoboda, R., Eagar, R., Ondko, P., and Kirschnick, F. (2015). Effective Infrastructure Asset Management- A holistic approach to transformation. Arthur D. Little - Prism, Hong Kong.

Qiu, Q., Fleeman, J. A., Ball, D. R., Rackliffe, G., Hou, J., and Cheim, L. (2013). "Managing critical transmission infrastructure with advanced analytics and smart sensors." IEEE Power and Energy Society General Meeting, IEEE, 1–6.

Salem, S., Siam, A., El-Dakhakhni, W., and Tait, M. (2020). "Probablistic resilience-guided infrastructure risk management." Journal of Management in Engineering, 36(6), 1–15.

Schaller, J. (2012). "Common mode event perspectives from the canadian electricity association equipment reliability information system." IEEE Power and Energy Society General Meeting, IEEE, 1–6.

Schaller, J., and Ekisheva, S. (2016). "Leading causes of outages for transmission elements of the North American bulk power system." IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Scheibe, K. P., Nilakanta, S., Ragsdale, C. T., and Younie, B. (2019). "An evidence-based management framework for business analytics." Journal of Business Analytics, Taylor & Francis, 2(1), 47–62.

de Sousa, R., Gonzalez, D., Rodriguez Gonzalez, J., and Tai, H. (2018). Harnessing the power of advanced analytics in distribution asset management. London.

Tsai, C. W., Lai, C. F., Chao, H. C., and Vasilakos, A. V. (2015). "Big data analytics: a survey." Journal of Big Data, Springer International Publishing, 2(1), 1–32.

Uddin, W., Hudson, W. R., and Haas, R. (2013). Public Infrastructure Asset Management. McGraw-Hill Education, New York.

United Nations. (2021). Managing Infrastructure Assets for Sustainable Development - A Handbook for Local and National Governments. (United Nations Department of Economic Affairs and United Nations Capital Development Fund, eds.), New York.

Vesset, D. (2018). "Descriptive analytics 101: What happened?" IBM, <https://www.ibm.com/blogs/business-analytics/descriptive-analytics-101-what-happened/> (Mar. 11, 2020).

Zuashkiani, A., Schoenmaker, R., Parlikad, A., and Jafari, M. (2014). "A critical examination of asset management curriculum in Europe, North America and Australia." Asset Management Conference 2014, IET, London, 1–6.

Zumel, N., and Mount, J. (2020). Practical Data Science with R. Manning Publications Co., Shelter Island.

Chapter 2

NETWORK ANALYTICS FOR INFRASTRUCTURE ASSET MANAGEMENT SYSTEMIC RISK ASSESSMENT

ABSTRACT

The ever-increasing investment gap in deteriorating infrastructure has necessitated the development of more effective asset management (AM) strategies. However, information asymmetry between AM stakeholder silos has been recognized as a key challenge in implementing effective AM strategies. The connectivity within the AM system introduces systemic risks (possibility of dependence-induced cascade failure) to the entire AM system operation when information asymmetry occurs. This study describes a toolbox to enable AM stakeholders to assess such systemic risks through a network analytics approach. The network, representing the AM system, is examined through its centrality measures to identify the most critical AM subject areas within the AM system. These subject areas are subsequently paired with assets' key performance indicators (KPIs). Within the developed toolbox, *descriptive* analytics provide transferrable KPI insights between stakeholders to reduce key asset information asymmetry. In parallel, *predictive* analytics forecast KPIs, ensuring stakeholder awareness of future asset performance to allow for appropriate preparation. Subsequently, *prescriptive* analytics employs heuristic-based optimization for optimal configuration of the AM network. The five tools are: 1) Dependence identification and network modelling; 2) Network centrality analysis; 3) Descriptive analytics of critical subject area paired KPI; 4) KPI-based predictive analytics; and 5) Prescriptive analytics for optimal network configuration. The utility of the developed toolbox is demonstrated for Tools 1-3 using a real AM system network and KPIs associated with power transmission infrastructure outages. Based on the analyses, managerial insights are drawn to illustrate the utility of the developed approach in improving the information asymmetry within the AM system—subsequently mitigating dependence-induced systemic risks.

KEYWORDS: analytics, information asymmetry, infrastructure asset management, key performance indicators, network analysis, systemic risk, transmission outages.

2.1. INTRODUCTION

The infrastructure assets in Canada and the United States continue to deteriorate each year, widening the gap in infrastructure spending needed to improve the asset conditions to serviceable levels (Infrastructure Canada 2018; McBride and Moss 2020). The American Society of Civil Engineers (ASCE) (2021) report gave America's infrastructure an overall grade of C-, up from a D+ in 2017. Although this improvement was partially attributed to asset owners adopting asset management (AM) techniques to prioritize spending constrained by the limited funding resources, the lingering low grade is attributed to massive maintenance backlogs, deteriorating infrastructure condition, and a lack of comprehensive asset inventory and consistent condition data (ASCE 2021). The *Canadian Infrastructure Report Card* (2019) states that most infrastructure used daily by Canadians is more than 20 years old and deteriorating rapidly. This report also outlined effective AM plan implementation and operationalization within asset-intensive organizations as critical to maximizing the impact of limited resources. In this respect, Uddin et al. (2013) concluded that infrastructure AM encompasses the systematic and coordinated planning and programming of investments, design, construction, maintenance, operation, and in-service evaluation of physical infrastructure and associated components. Additionally, Ross (2019) concluded AM can be described as the collective term for the structured decision-making and execution of plans to optimize a balance between infrastructure performance, efforts, and risk through the use of available and procurement of future assets.

2.1.1. ASSET MANAGEMENT SYSTEM MODEL

The Institute for Asset Management (IAM) developed the conceptual AM model in 2014 as a guide for AM professionals to implement and operate an AM approach in their organizations (Institute for Asset Management 2015). The IAM is the international professional body for AM professionals and generates AM knowledge, best practice guidance, and awareness of the benefits of the AM discipline for individuals, organizations, and wider society (Institute for Asset Management 2021). The AM system model has also been referred to as the *Six Box Model* as there are six connected AM divisions (i.e., *Strategy & Planning*, *Asset Management Decision-Making*, *Life Cycle Delivery*, *Asset Information*, *Organization & People*, and *Risk & Review*), as shown in Figure 2-1 (Institute for Asset Management 2015). *Strategy & Planning* aligns the organization's AM activities to fit within a consistent plan that has been developed and approved by all stakeholders. *Asset Management Decision-Making* reviews the challenges and makes decisions regarding how each of these stages occurs within the main areas of an asset's life: asset acquisition/creation; operation; maintenance; and end of life disposal, decommissioning, or renewal. *Life Cycle Delivery* involves the entire lifespan of the asset, from acquisition/creation, through operation/maintenance, and finally, end of life disposal, decommissioning, or renewal. *Asset Information* is typically input to an AM process, created or modified by a process, or the output of a process. *Organization & People* involves a review of the organizational structure, roles, responsibilities, and contractual relationships. *Risk & Review* identifies the risks related to an asset's life cycle delivery, understands and manages such risks, establishes a feedback mechanism within the organization to allow for input on the AM objectives, strategy and plan, and supports the continued improvement and development of AM activities.

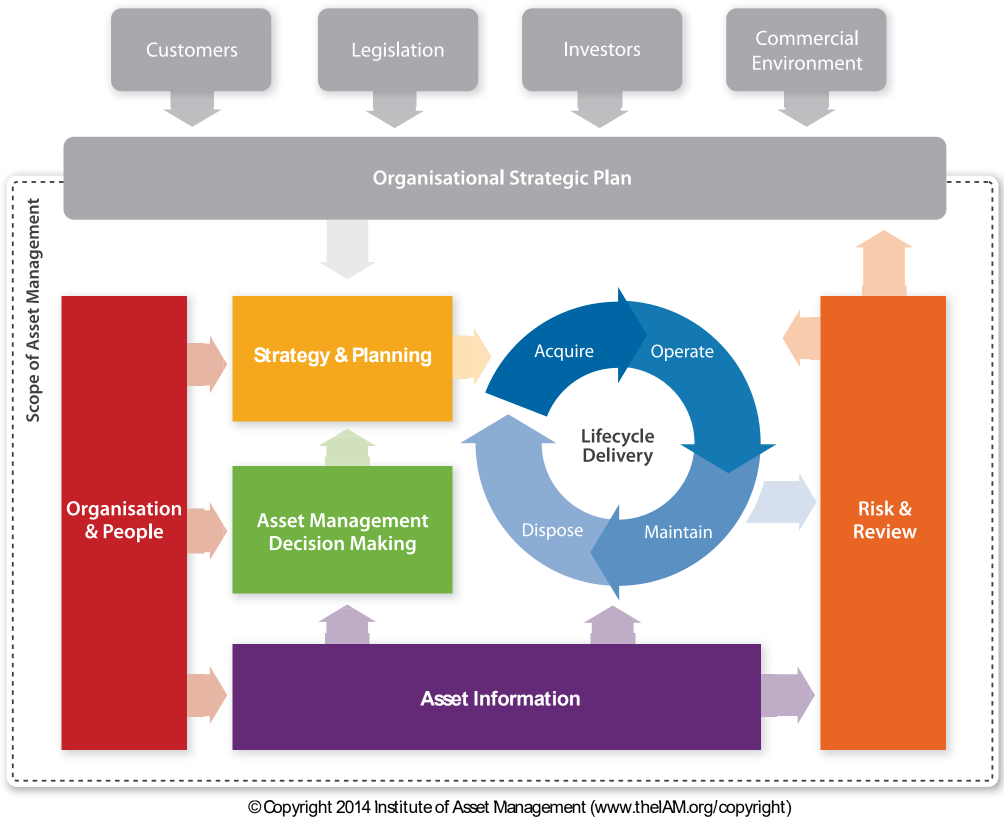


Figure 2-1. IAM conceptual model from (Institute for Asset Management 2015).

Overall, there are 39 AM system subject areas, outlined by the *Global Forum's Asset Management Landscape* (2014), across the six AM system divisions, as shown in Figure 2-2. The 39 subject areas were designed to illustrate the breadth of activities within the scope of AM, the interrelationships between activities and the need to integrate them, and the critical role for AM to align with and deliver the strategic plan goals of an asset-intensive organization (Institute for Asset Management 2015). The connectivity among the AM subject areas within the AM system introduces systemic risks (possibility of dependence-induced cascade failure). This might occur when functionality failure(s) through either one or more AM subject areas, nodes, or information flow, links, cascade throughout the remaining functional AM subject areas, thus hindering relevant decision making abilities.

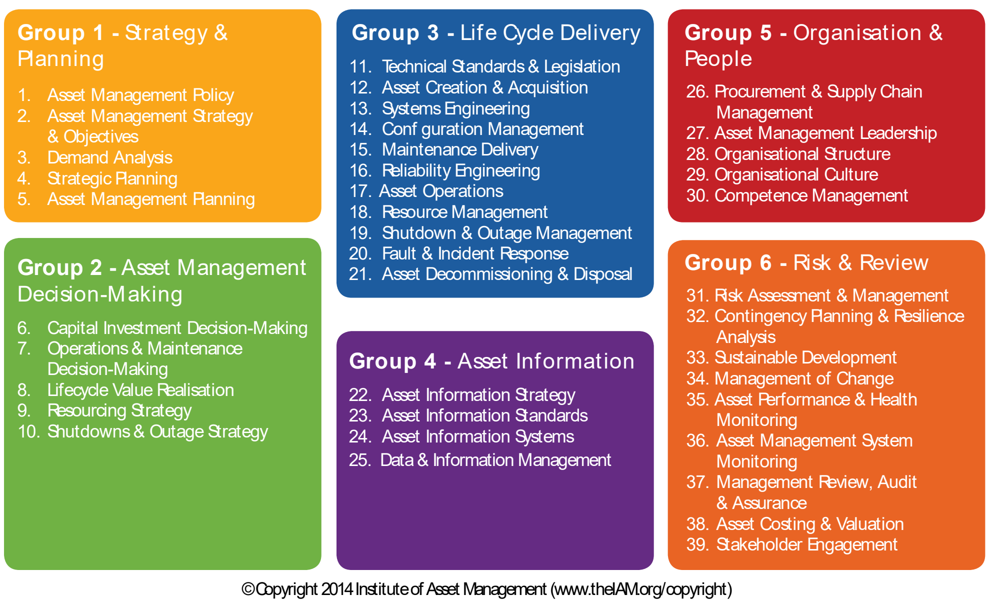


Figure 2-2. IAM conceptual model subject areas from (Institute for Asset Management 2015).

2.2. BACKGROUND

Key to effective AM is the *collective and coordinated* effort that involves collaboration between multiple stakeholders (e.g., engineering department, operations department, AM department, finance department, project management department, and owner). These AM stakeholders form silos in the absence of necessary collaboration. Such stakeholder silos have been shown to be the major hurdle in the implementation and operation of an effective AM system within an asset-intensive organization (Pell et al. 2015; de la Pena et al. 2016; Golightly et al. 2018). The AM implementation and operation failures typically occur when stakeholders experience information asymmetry because of inadequate information-sharing protocols and/or not readily sharing key information that could mutually benefit their infrastructure's AM system (Brunetto et al. 2014; Xerri et al. 2015; Golightly et al. 2018).

Information asymmetry occurs when one party in a relationship has more or better quality real-time or historical information than another (Bergh et al. 2019). Such information asymmetry creates systemic (i.e., dependence-induced) risks within the AM system due to stakeholders not timely realizing (and thus responding to) the impact of different stakeholders' decisions within the implementation and operation of their AM system. In other words, they are precluded from using real-time or historical information to support their decision-making (Bergh et al. 2019). This, in turn, causes isolation of an AM subject area node or breaks some of its information flow links to other nodes, potentially inducing a cascade failure throughout the AM system network. An example of systemic risk in a different industry is in international banking where in 2008, the failure of the Lehman Brothers caused the collapse of the global banking sector and, without government bailouts to other major banking institutions, the cascade collapse would have been far greater (López-Espinosa et al. 2015; Miller 2017). In this example, there was information asymmetry in the mortgage-backed securities where risky mortgages were packaged as high-quality debt leading to the seller having better information than the buyer (Tarver 2020). In this study through a network analytics lens, the systemic risks involved with implementing and operating an AM system pertain to information asymmetry due to either node failure(s), which represent a specific AM subject area losing functionality and thus its ability to contribute to the overall (AM system) network, or link failure(s), which represent an interruption in the information flow between AM subject area nodes.

It is also important to understand that organization structure can be either decentralized (i.e., the organization is divided into smaller teams in charge of specific aspects of the organization and decision-making occurs at various levels within the organization (Graybeal et al. 2018)) or centralized (i.e., one or a select few individuals make the important decisions (e.g., resource allocation) and provide the strategic direction for the organization (Graybeal et al. 2018)). A decentralized structure offers many benefits in quick decision and response time and skilled and specialized management. Johnson & Johnson, for example, has successfully adopted this management structure across their over 200 operating companies (Weldon 2008; Mohamad et al. 2017). However, drawbacks of a decentralized structure include coordination issues between teams working towards a company’s strategic goal, and each team prioritizing their own goals over the organization's goals (i.e., teams operating as silos) (Vantrappen and Wirtz 2017). The benefits of a centralized organization structure include clarity in decision-making, streamlined implementation of policies and initiatives, and control over the strategic direction of the organization (e.g., Apple) (Graybeal et al. 2018). The disadvantages include employees having difficulty providing feedback on operations and the limited flexibility among the lower management levels to influence changes (Vantrappen and Wirtz 2017). Decentralized decision-making in infrastructure restoration was shown to be an effective approach by Crowther (2008) as well as Talebiyan and Dueñas-Osorio (2020). The AM system is typically decentralized with AM stakeholders in charge of and making decisions pertaining to specific AM subject areas (Golightly et al. 2018). Thus, the current study views and analyzes the AM stakeholders as a decentralized system while still proposing a centralized information database solution that addresses the main challenge of information asymmetry between AM stakeholders and ensures AM stakeholders are not overwhelmed with too much information.

A decentralized system can be represented as a network consisting of connected nodes and links, representing a web of connected components (Barabási 2016). The nodes simulate the components of a system, whereas the links represent the dependency between these nodes. Networks are often analyzed using specific measures related to either system components (i.e., node-based or link-based) or the entire connected system (i.e., network-based). Node-based measures focus on centrality analysis as it relates to the node's importance in the network by assessing the connectedness of that node to other network nodes. There are different centrality measures applied in a wide variety of applications (Derrible 2012; Lee et al. 2013; Estrada and Knight 2015; Das et al. 2018; Ezzeldin and El-Dakhakhni 2019; Goforth et al. 2020).

Analytics facilitates the realization of business objectives through reporting of data to analyze trends (i.e., descriptive analytics), creating prediction models for forecasting (i.e., predictive analytics), and optimizing processes to enhance performance (i.e., prescriptive analytics) (Tsai et al. 2015; Delen and Ram 2018). Analytics have been applied in various studies within multiple infrastructure industries to improve AM processes. *Descriptive* analytics applications focus on deriving insights into performance trends, from complex data, mainly through visualizations (Abdelfatah et al. 2013; Barker et al. 2017; Black et al. 2018; Mukherjee et al. 2018). *Predictive* analytics applications use historical data within machine learning models to predict an output (e.g., health index, condition, outage severity, or asset remaining life) (Zhou et al. 2016; Dehghanian et al. 2019; Yang et al. 2019; Piryonesi and El-Diraby 2020). *Prescriptive* analytics applications attempt to optimize intervention and maintenance planning and scheduling (Qiu et al. 2013; Chen et al. 2015; Heng et al. 2016; Abu-Samra et al. 2020). Such applications showed the benefits of employing analytics to improve the specific subsets of AM, but there remains a disconnect between the use of the analytics and the bigger picture view that considers systemic risks within the infrastructure AM system.

In addition, although studies have shown that analytics provides a competitive edge when integrated into business processes (Delen et al. 2018; Scheibe et al. 2019; O'Neill and Brabazon 2019; Hassan 2019) it is critical to first identify the key hubs within an organizational structure, through which information flows, for the organization to operate effectively (McDowell et al. 2016). This concept has been studied in organizational networks identifying key stakeholders (e.g., companies, people, or departments) that are critical to the functionality and effective operation of the organization (Barão et al. 2017; Ujwary-Gil 2019; Eisenberg et al. 2020). Nonetheless, to the best of the authors' knowledge, the identification of AM systemic risks with a method to reduce the information asymmetry within the AM system subject areas (i.e., hubs) is yet to be developed. As such, in this study, a toolbox is created to integrate network analysis and data analytics for an AM system model that incorporates the decentralized nature of the AM stakeholders and subject areas and presents a centralized database whereby AM subject area-specific information will be displayed for the necessary stakeholders responsible for such AM subject areas. This approach will ensure consistency across all AM subject areas while also preventing AM stakeholders from being overloaded with too much information, allowing them to focus on only the information necessary to make decisions within the AM subject areas they are responsible for (Herrera et al. 2011; Prajogo et al. 2018).

This paper first outlines the study goals and objectives, followed by a description of the considered network measures. Subsequently, the developed toolbox is presented to describe five distinct tools that identify critical AM system subject areas using network analysis and employ analytics with infrastructure asset key performance indicators (KPIs) to reduce the systemic risks caused by information asymmetry between dependent AM subject areas. In the current study, due to data restrictions, the utility of the toolbox is demonstrated using only the first three of five tools considering an AM conceptual model developed by the IAM and power transmission infrastructure outage KPIs. Finally, managerial insights are drawn to illustrate how asset managers can reduce the systemic risks within an AM system.

2.3. Study Goal and Objectives

The study goal is to mitigate systemic risks, within an AM system, created by information asymmetry between dependent AM subject areas. This study attempts to break down the silos that infrastructure stakeholders operate within, allowing decisions to be made using the same information ensuring cohesiveness among stakeholders working towards their AM goals and objectives with a set of described tools. The described tools will enable AM stakeholders to identify the network structure of an AM system—achieved by modelling the complex connections within such a system. In addition, the described tools will allow for AM stakeholders to analyze the resulting network to identify dependence-induced systemic risks to implementing and operating an effective AM system—achieved by identifying the critical subject areas within the network using network measures (e.g., node- and link-based centralities) specific to the AM system network structure. Finally, the described tools will employ descriptive, predictive, and prescriptive analytics using infrastructure KPIs (e.g., average outage duration, bridge condition index, and the number of water main failures per 1000 km) and the AM network structure, specific to the organization owning, managing and operating the infrastructure assets to reduce the information asymmetry—thus ensuring risk-informed and effective decision-making. This concept, applied on a manufacturing operational performance study by Prajogo et al. (2018), illustrated that good information management practices within an organization can have a significant impact on the overall business performance. The latter was accomplished through sharing information, using information technology tools within an organization, and sharing information with supply chain partners. As such, Prajogo et al. (2018) emphasized that organization management must look for ways to facilitate the sharing and centralized management of information across internal and external organizational boundaries.

2.4. NETWORK MEASURES

Complex network theory allows for the modelling of complex system connections through a network of nodes and links (Boccaletti et al. 2006; Barabási 2016; Salama et al. 2020). This section provides a background of some relevant node- and network-based measures. Within the context of this study, nodes represent the main AM system subject areas and links represent the connections between the subject areas. The links within the network are directed, indicating information, documentation, knowledge, and/or policy being transferred from a source node to a target node, and unweighted, as each connection is viewed as equally important (unless otherwise specified) to the overall operation of the AM system. The level of connectedness within an AM system necessitates a network-based model to understand each node's importance to the AM system. An adjacency matrix (**A**) can be formed that describes the connectivity and disconnection between the AM system network nodes. Each element of the adjacency matrix, A_{ij}, is either 1, illustrating a direct connection between nodes i and j $(i \neq j)$, or 0 otherwise (Barabási 2016). Specific node-based centrality measures that relate to the AM network model include:

Betweenness centrality identifies nodes that play a central role in connecting to other nodes in the network (Freeman 1977). The betweenness centrality of node (BC_i) measures the total number of shortest paths passing through node *i* as expressed in Eqn. 1.

$$BC_i = \sum_{j \neq i \neq k} \frac{\rho_{jk}(i)}{\rho_{jk}} \tag{2-1}$$

where ρ_{jk} = number of shortest paths connecting node *j* to node k and $\rho_{jk}(i)$ = number of shortest paths connecting node *j* to node *k* that traverse node *i* in the network.

Closeness centrality represents how close a node is to all other network nodes (Estrada and Knight 2015). The closeness centrality of node *i* (CC_i) is determined by finding the shortest path using either weighted or unweighted links for node *i* as:

$$CC_i = \frac{N-1}{\sum_j d(i,j)} \tag{2-2}$$

where $d(i,j)$ are the shortest path distances between nodes *i* and *j* and *N* is the total number of network nodes.

Degree Centrality assesses the relative influence of nodes as the number of degrees (links) that a node directly shares with other nodes (Estrada and Knight 2015). As such, the degree centrality of a node i (DC_i) is defined using the adjacency matrix $\boldsymbol{A} = (a_{i,j})$ as:

$$DC_i = \sum_{j=1}^{N} a_{ij} \tag{2-3}$$

This centrality measures the direct influence of a node on its connected nodes.

Eigenvector centrality quantifies the extent of node connectedness to other important (i.e., high degree centrality) nodes (Thai and Pardalos 2012). The relative centrality score of node *i* (x_i) for adjacency matrix $\boldsymbol{A} = (a_{i,j})$ is:

$$x_i = \frac{1}{\lambda}\sum_j a_{i,j}x_j \tag{2-4}$$

where λ is a constant eigenvalue and the equation could be rearranged in vector notation as the eigenvector equation $\boldsymbol{Ax} = \lambda\boldsymbol{x}$. This centrality indicates node importance as its connection to other important nodes and non-connection to unimportant nodes in the network.

In addition to node failures, links might also fail, representing information asymmetry between nodes. Therefore, it is important to understand the importance of each link to the functionality of the network. A centrality metric related to the importance of information flow in a network is the link betweenness centrality (LBC) (Teixeira et al. 2016), defined as the number of shortest paths that traverse the link (Freeman 1977). Practicality, the LBC is a measure of how central a link is to the network, and in the case of the AM system network, it measures the criticality of a specific link to information asymmetry. High ranking links contribute systemic risks to the network as their failure would lead to a cascading failure throughout the network. Although, the links within this study are specified as unweighted, future extensions of this toolbox might incorporate link types and weights that relate to the type or criticality of information that is passed between nodes (e.g., raw, pre-processed, figures, decisions, or connections to AM standards).

In addition to the aforementioned node- and link-based centrality measures, there are also network-based measures that quantify the connectedness of the overall network structure (Estrada and Knight 2015; Opdyke et al. 2017; Valentin et al. 2018). The most relevant to the AM network include:

Network density (ND) represents the ratio of actual links within a network to the potential links that could be formed within the network if the network were fully connected (Barabási 2016). It can be calculated using the following equation for a directed network:

$$ND = \frac{l}{N(N-1)} \tag{2-5}$$

where l is the number of links in the network and N is the number of nodes in the network. The network density is a measure of the network's health and effectiveness. The ratio has values ranging from 0 (i.e., a completely unconnected network) to 1 (i.e., a fully connected network). In AM applications, it can assess the level of connectedness between all subject areas in sharing information and indicate the susceptibility to network failure (for low values).

Average degree centrality (ADC) is the ratio of the summation of the degree centrality values for all nodes to the total number of nodes in the network (Barabási 2016). It can be calculated using the following equation for all i nodes:

$$ADC = \frac{\sum DC_i}{N} \tag{2-6}$$

where DC_i is the degree centrality for node i and N is the total number of nodes in the network. This measure indicates how quickly disruptions can diffuse throughout the network. Within the context of AM, this measure refers to network dependence and highlights the systemic risk due to the cascading effects of failed AM system subject areas.

2.5. NETWORK ANALYTICS TOOLBOX

To address the study goal and objectives, the following five tools were developed, as shown in Figure 2-3: 1) Dependence identification and network modelling – where the AM network structure is identified and modeled; 2) Network centrality analysis – to identify the critical AM subject areas causing systemic risk; 3) Descriptive analytics of critical subject area paired KPI – to develop targeted visualizations that focus on only the necessary information for decision-making relevant to specific AM subject areas; 4) KPI-based predictive analytics – to forecast KPI metrics to enable more proactive decision-making; and 5) Prescriptive analytics for optimal network configuration – to minimize AM network systemic risks. Each of these tools will be further described in detail later in the paper as to how each tackles the study goal and objectives.

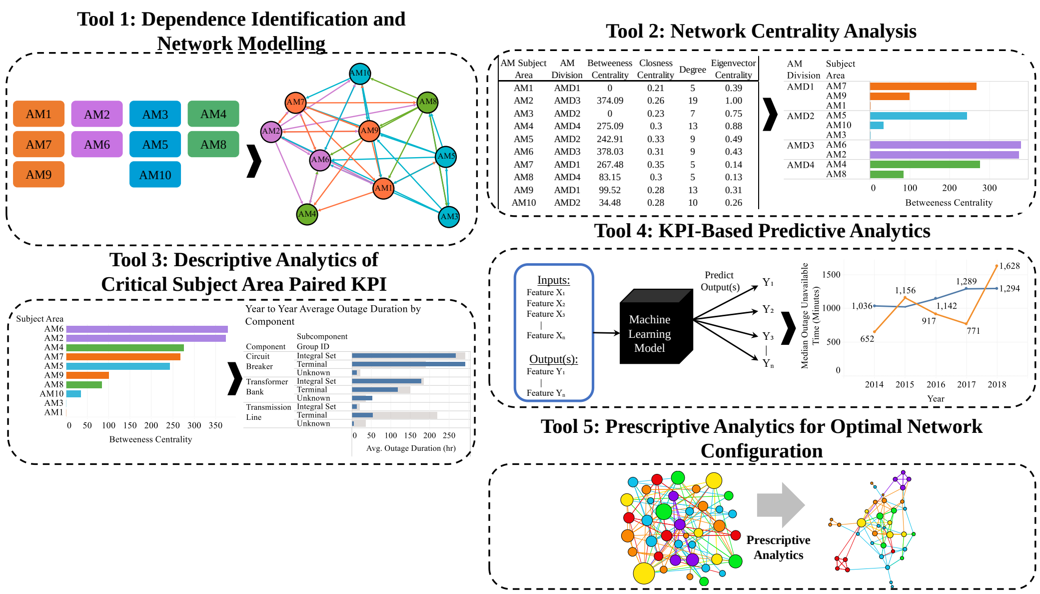


Figure 2-3. Overview of the asset management network analytics toolbox.

2.5.1. Tool 1: Dependence Identification and Network Modeling

Tool 1 describes the process for identifying the network structure of an AM system. The first step in implementing Tool 1 involves modeling an organization's AM system in terms of its specific subject areas that describe the implementation and operation of its AM system as a network. In this respect, the connections between subject areas are identified based on expert AM knowledge. The links between nodes are representative of a decision, information or data transfer, strategy, or policy that is passed from one node (i.e., source node) to another (i.e., target node). It is these links that define the dependence between subject areas to form the AM system network. These links are indicated in the adjacency matrix, as illustrated in Figure 2-4, where a link presence is indicated by a 1 and its absence is indicated by a 0. As mentioned earlier, the links do not have an associated weight value as the link represents the presence of a connection in the form of a decision, information or data transfer, strategy, or policy that is passed from one source node to another target one. The adjacency matrix can then be used to visualize the network, as illustrated in Figure 2-4. For example, Figure 2-4 shows ten AM subject areas in the adjacency matrix and an illustrative network of a potential AM network model with nodes representing subject areas and links representing the connections between subject areas.

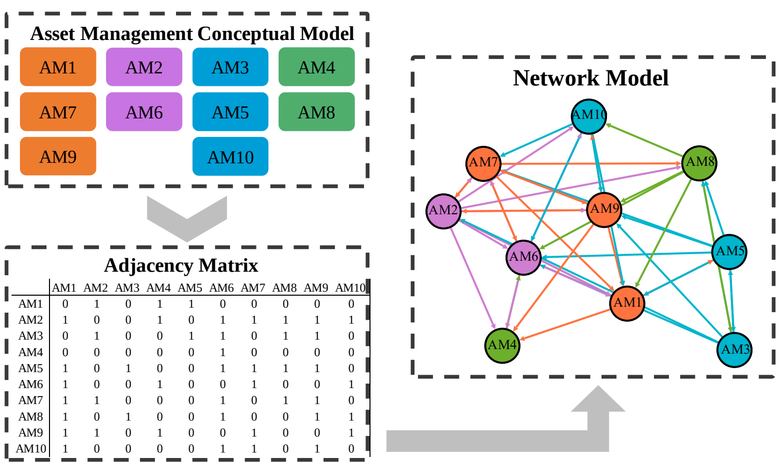
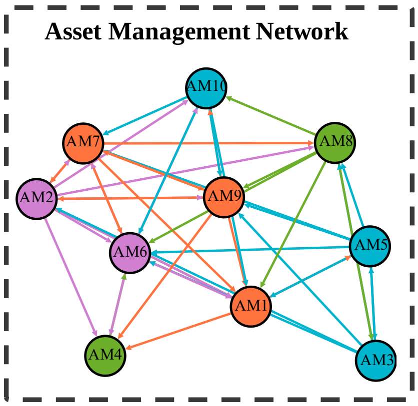


Figure 2-4. Tool 1: Dependence identification and network modelling of asset management subject areas.

2.5.2. Tool 2: Network Centrality Analysis

Tool 2 describes the process for analyzing the resulting network to identify dependence-induced systemic risks to implementing and operating an effective AM system. The AM network layout, as generated using Tool 1, is employed by Tool 2 to calculate the centrality measures—identifying the importance of each AM subject area within an AM system. Such centralities, illustrated in Figure 2-5, highlight potential node/link systemic risk in the AM system where the centralities are converted to a ranked list of node/link importance. In this respect, the betweenness centrality is a measure of the importance of the subject area to the overall implementation of the AM system within an organization. The closeness centrality is a measure of indirect AM information flow between the not-directly connected nodes. The degree centrality is a measure of the criticality of AM subject areas to the dependent subject areas. The eigenvector centrality is a measure of node connectedness and importance to other highly connected nodes, identifying subject areas that have a strong influence on other important AM subject areas. An organization would need to determine the most relevant centrality measure to their implementation strategy. For example, an organization conducting a preliminary screening of their AM structure would utilize the eigenvector centrality to determine subject areas that influence other highly influential subject areas, allowing the organization to focus their attention on a few subject areas to maximize the impact of improvement in their AM system. When the calculated node- or link-based centrality measures are high, there is a greater likelihood of the AM network failing if such an important subject area or link were to become dysfunctional. Therefore, the importance ranking identifies the most critical subject areas and links exposed to the dependence-induced systemic risk involved with the implementation and operation of an AM system.



Network Measures Table

AM Subject Area	AM Division	Betweeness Centrality	Closness Centrality	Degree	Eigenvector Centrality
AM1	AMD1	0	0.21	5	0.39
AM2	AMD3	374.09	0.26	19	1.00
AM3	AMD2	0	0.23	7	0.75
AM4	AMD4	275.09	0.3	13	0.88
AM5	AMD2	242.91	0.33	9	0.49
AM6	AMD3	378.03	0.31	9	0.43
AM7	AMD1	267.48	0.35	5	0.14
AM8	AMD4	83.15	0.3	5	0.13
AM9	AMD1	99.52	0.28	13	0.31
AM10	AMD2	34.48	0.28	10	0.26

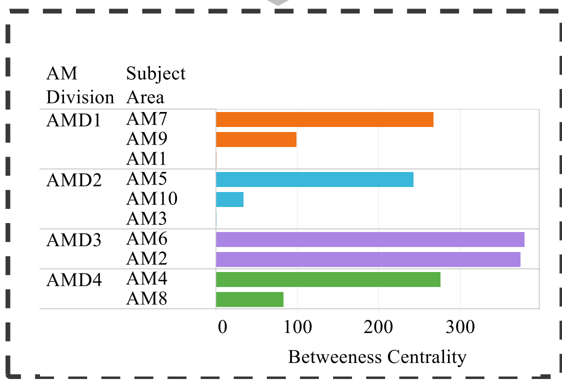


Figure 2-5. Tool 2: Asset management network centrality analysis and subject area importance ranking.

2.5.3. Tool 3: Descriptive Analytics of Critical Subject Area-Paired KPI

Tool 3 describes the process for implementing descriptive analytics for infrastructure asset KPIs. The subject areas most exposed to systemic risk, as identified from the centrality measures from Tool 2, can be further analyzed using

Tool 3 when paired with infrastructure-industry-specific KPIs. Tool 3 visualizations are designed to draw their information from a centralized database and only display KPI information directly related to an AM subject area. Tool 3 visualization are designed to ensure stakeholders focus only on the pertinent information necessary to make decisions within that AM subject area instead of being overwhelmed by all AM information from the centralized database. For example, the subject area *outage management* is paired with the KPI *avg. outage duration* year-over-year and the trend in the *avg. outage duration* as shown in Figure 2-6. Figure 2-6 presents an example of two KPIs and their evolving values with time. It should be noted that only the KPI *values* are expected to continue to change with time (i.e., dynamic) as new information becomes available, whereas their pairing to the systemic risk-critical AM subject areas is expected to remain largely the same (i.e., static) as the AM system network is not expected to change with time.

Infrastructure KPIs are metrics of a specified asset or overall infrastructure network performance. KPIs can be continuous or discrete, and can also be qualitative (e.g., low, medium, high) or quantitative (e.g., 50-70%) in nature. Although the KPI are paired to the AM subject areas, making them static in terms of their evaluation approach, their values are nonetheless expected to be dynamic as they continuously change with time under different conditions (e.g., climate). For example, the pavement industry uses pavement condition index and international roughness index, the bridge industry uses bridge condition index, the power industry uses system average interruption frequency index, system average interruption duration index and mean outage duration, and the water and wastewater industry uses the number of breaks per year, the number of failures per 1000 km, and leakage of water per year (Alzoor et al. 2021; Uddin et al. 2013). As the KPIs differ between infrastructure industries, Tool 3 pairs the critically dependent subject areas of the AM system with the relevant KPIs. Descriptive analytics can then be used to illustrate these paired KPIs, as illustrated in Figure 2-6. Descriptive analytics often includes building a KPI-tailored dashboard that allows user interaction to gain useful KPI insights (Wexler et al. 2017). Tool 3 facilitates clear dashboards to be circulated between stakeholders to ensure every stakeholder would be informed on the KPIs related to the important AM subject areas. Huang et al. (2019) showed that when stakeholders can see the impact of their work, they are more likely to develop trust in the management processes and therefore share information internally more readily. Therefore, Tool 3 facilitates clear visualizations to be circulated among stakeholders that manage the key AM subject areas, ensuring that these stakeholders only see the necessary information related to the decisions they need to make within the AM subject area they are responsible for. Tool 3 applications, in turn, allow for AM stakeholders to monitor their impact on the AM system for their specific AM subject areas and ensure all decisions specific to each AM subject area are made using consistent information.

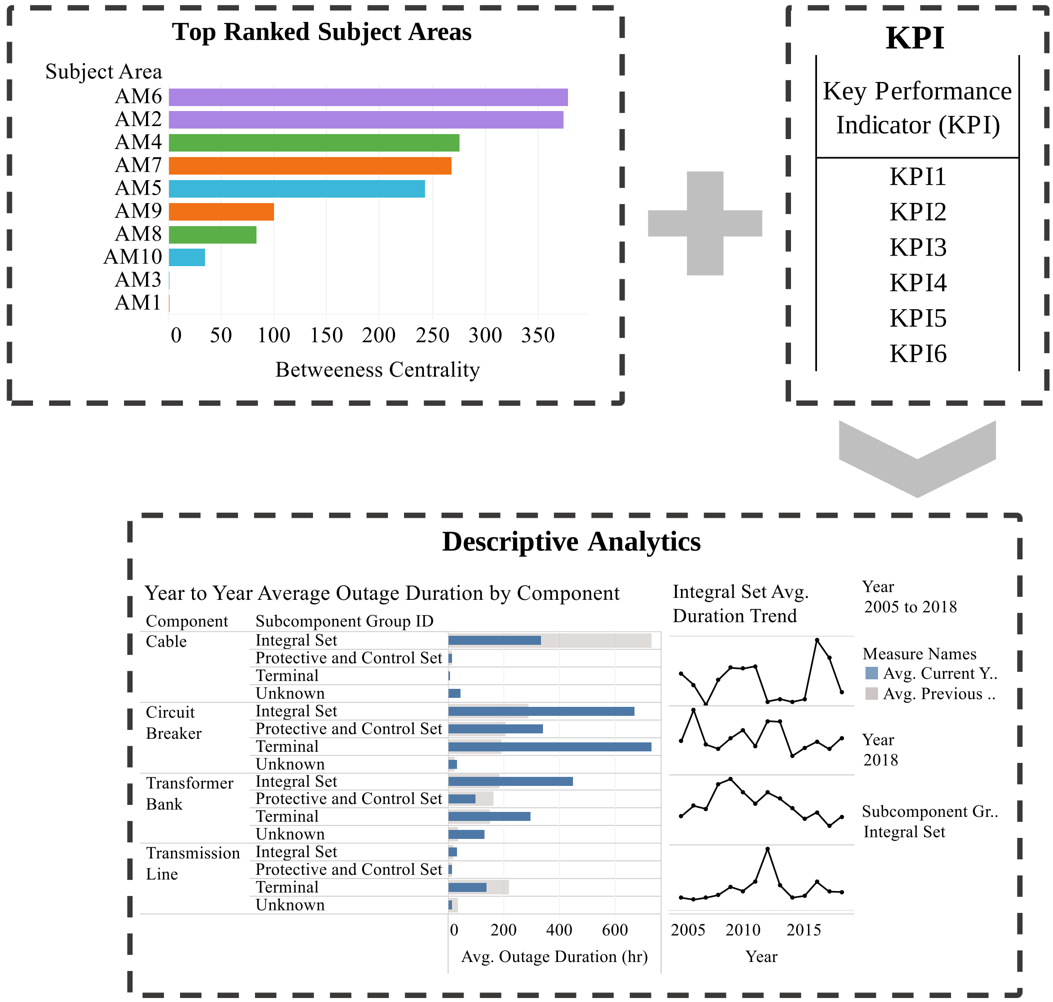


Figure 2-6. Tool 3: Descriptive analytics to inform critical asset management subject areas using paired KPIs.

2.5.4. Tool 4: KPI-Based Predictive Analytics

Tool 4 describes the process for implementing predictive analytics for infrastructure KPIs. Building on the descriptive analytics of Tool 3, Tool 4 is focused on developing a predictive analytics model for the aforementioned KPIs. Figure 2-7 outlines the process for this tool by including historical KPI performance within a machine learning model to output forecasted KPI metrics. Machine learning models are typically classified as either supervised (i.e., developing a mathematical function that maps the relationship between specific input-output pairs) or unsupervised (i.e., categorizing the dataset based on similarity, without pre-specifying outputs) (Zumel and Mount 2020). Examples of machine learning models include decision trees, artificial neural networks, and support vector machines (Aggarwal 2015).

Input data is employed to train the machine learning model to predict a numerical or categorical output based on the provided contributing features (Hastie et al. 2009). This allows a decision-maker to predict a KPI output value based on contributing features (e.g., climatic conditions, economic conditions, geographic location, asset characteristics, time, and maintenance history) and past historical KPI values (Haggag et al. 2021). Any additional input features would be infrastructure industry-specific, therefore the organization would need to establish which features would be accessible before building a KPI predictive analytics model. Other research studies have successfully predicted specific infrastructure KPI within different industries in isolation (Zhou et al. 2016; Dehghanian et al. 2019; Yang et al. 2019; Piryonesi and El-Diraby 2020), therefore a summary of some key techniques deployed in those studies to meet the goals of Tool 4 is provided below.

For example, Figure 2-7 is presented as an illustration of this tool to show a forecasted 5-year period for the *average outage duration* KPI used by the power industry. The input features include contributions to the outage (e.g., outage cause, failure mode, and climatic information) and component or system characteristics (e.g., voltage, affected component/system, time of the outage, and geographic location). As to be expected, the outputs from Tool 4 applications are only good for AM decision-making if the information used for inputs is of high quality and pertain to meaningful data. Koziel et al. (2021) investigated the impact of using faulty data in AM decision-making and found that there were significant implications on optimal replacement schedules. Therefore, high-quality data must be gathered related to each AM subject area-paired KPI to facilitate reaching the most effective AM decisions. The two lines in Figure 2-7 (i.e., orange and blue) indicate different organizations and their forecasted KPI. The KPI predictive analytics model would forecast future performance, allowing stakeholders to be informed and able to prepare plans for more effective AM. Tool 4 applications improve the information asymmetry in that all stakeholders are aware of and striving towards a clear performance goal for their critical AM KPIs and ensure consistency among stakeholders that manage the AM subject areas in that they make decisions based on consistent predictive models built specifically to each AM subject area-paired KPI.

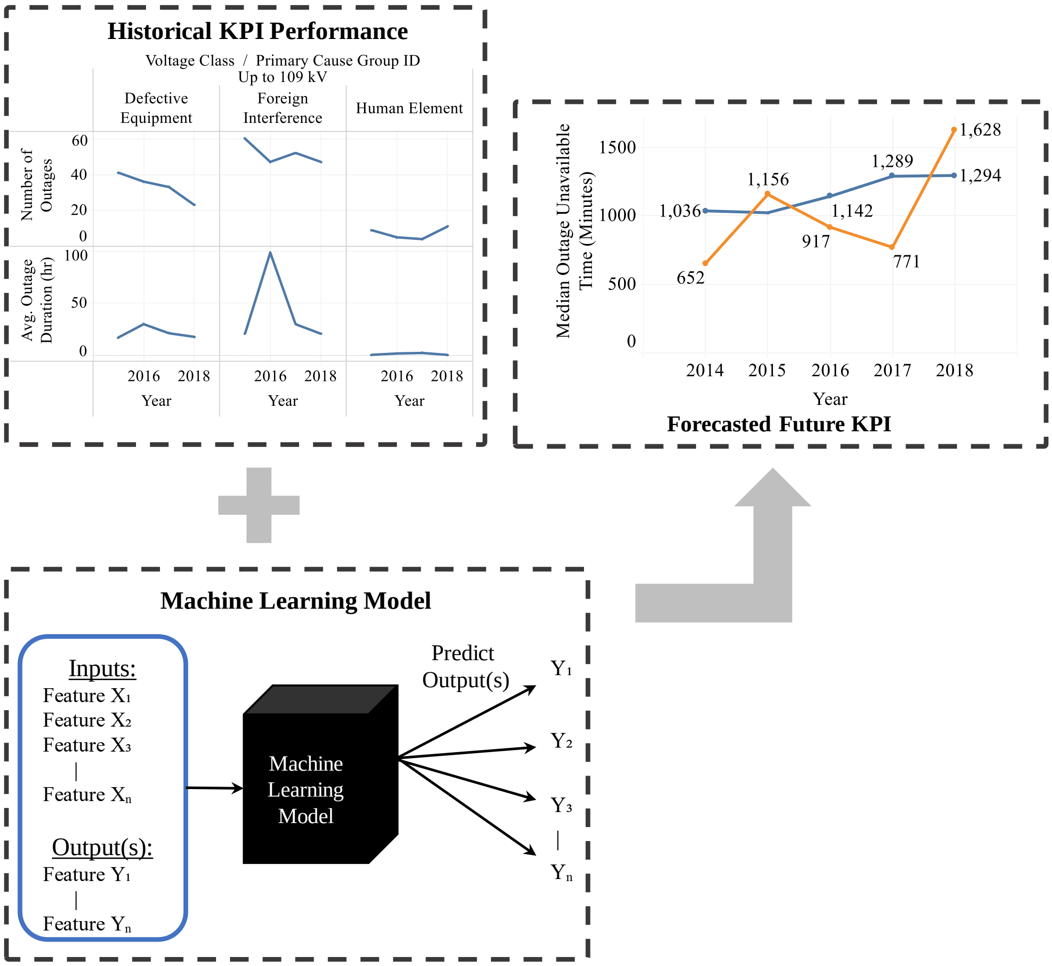


Figure 2-7. Tool 4: Predictive analytics to forecast future KPI performance using historical input features.

2.5.5. TOOL 5: PRESCRIPTIVE ANALYTICS FOR OPTIMAL NETWORK CONFIGURATION

Tool 5 describes the process for implementing prescriptive analytics by optimizing the AM subject area network through adding links to the original network to minimize the average centrality measure of choice (e.g., betweenness, closeness, degree, or eigenvector) or multiple centralities, depending on the objective function

(Thai and Pardalos 2012). The optimization problem would include a constraint on the number of allowable links to add to the network before it becomes too centralized or non-functional and there would be a cost per link addition in terms of new information, policy, or decision that would be transferred. The optimization would minimize the systemic risks in the AM network by reducing the impact of failure related to high centrality nodes or links through the addition or subtraction of links in the network while still maintaining the functionality of the AM system network. Such optimization would employ, for example, genetic algorithms or other heuristics, where a population of solutions is generated and evolves until a (near) optimal solution is obtained (Goldberg 1989). Each solution within the population represents a single realization of the input features (i.e., individual). New individuals are reproduced through special evolutionary operators including: 1) *elitism*, where individuals with greater fitness are replicated; 2) *crossover*, where sets of two individuals (i.e., parents) are selected based on predefined criteria (e.g., random selection or a selection based on the fitness value) and subsequently mixed to produce new individuals; and 3) *mutation*, where single parents are altered randomly to produce new individuals (Nearchou 2004; Scrucca 2013; Yosri et al. 2021). Within the AM system network optimization, each link would be represented as a feature within the individual and the values would be either 1, indicating a link presence, or 0, indicating link absence. The application of Tool 5 would present an optimized configuration of the AM system such that the systemic risks would be minimized, as illustrated in the new network configuration shown in Figure 2-8.

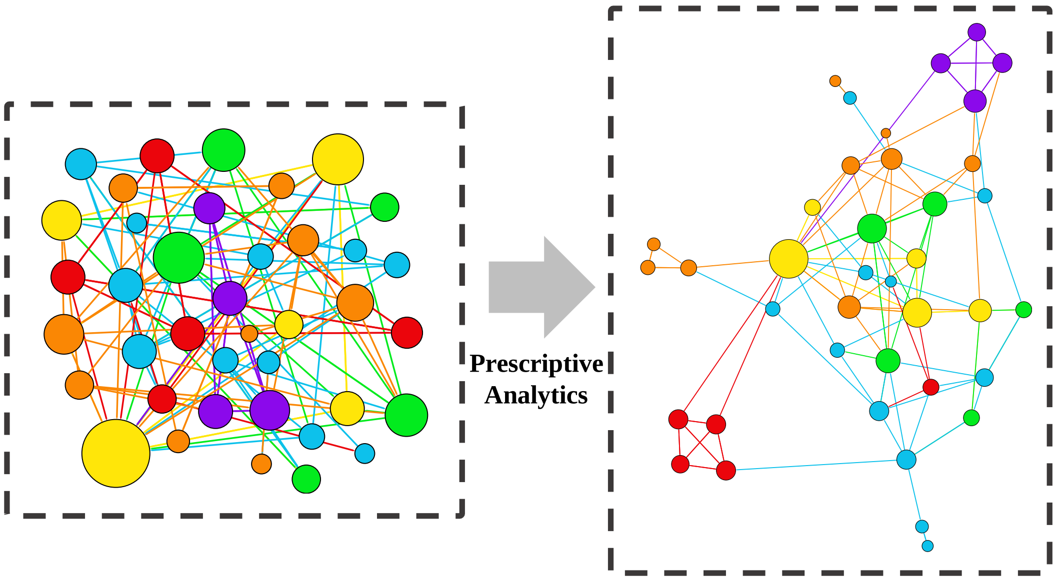


Figure 2-8. Tool 5: Prescriptive analytics for optimal AM system network configuration.

2.6. Demonstration Application: Power Transmission Infrastructure

The developed toolbox was demonstrated using the IAM's conceptual model for an AM network and transmission infrastructure asset outage data gathered by the Canadian Electricity Association (CEA). The IAM is the international professional body for AM professionals and develops AM knowledge, best practice guidelines, and generates awareness of the benefits of the AM discipline for individuals, organizations, and the wider society (Institute for Asset Management 2021). The CEA membership includes generation, transmission, and distribution power utilities and industrial partners from across Canada (CEA 2020). The toolbox was applied in this setting to display its application within the asset-intensive transmission power industry.

2.6.1. Project Description

The IAM conceptual model was used for building the infrastructure AM network. To show the connection between important AM subject areas and industry-specific KPIs, a transmission infrastructure asset outage dataset was obtained from the CEA, covering the period from 1978-2018. The transmission infrastructure network is critical to the reliable delivery of power from generators to substations and ultimately customers. Therefore, the effective and efficient management of transmission infrastructure assets is critical for safe and reliable power delivery. The transmission equipment outage data is for equipment operating at high voltages of 60 kV and above (CEA 2018). The outages are recorded for transmission infrastructure components including transmission lines, cables, transformer banks, circuit breakers, synchronous compensators, static compensators, shunt reactor banks, shunt capacitor banks, and series capacitor banks. The KPIs recorded and published by the CEA in their annual report are shown in Table 2-1 along with a definition of each KPI metric. This demonstration of the toolbox will involve the application of Tools 1 to 3 only, as the data needed for implementation of Tools 4 and 5 is restricted by transmission infrastructure owner/operator for their internal use. As such, the demonstration will focus on describing the utility of the toolbox for the identification of AM subject areas that are most critical to induce systemic risk within an AM system.

Table 2-1. KPIs calculated and published by the CEA in their annual report.

Key Performance Indicator (KPI)	Definition
Frequency (per 100 km.a)	The number of outages divided by kilometre years divided by 100.
Frequency (per a)	The number of outages divided by component years
Number of Outages	The number of major component-related forced outages.
Total Outage Duration (h)	Total forced unavailable time (i.e., the time required to completely restore a component to service) of the component-related outages.
Average Outage Duration (h)	Total outage duration divided by the number of outages.
Median Outage Duration (h)	50% of the forced unavailability times are greater than this value.
Unavailability (%)	The product of frequency and average outage duration in years. It is expressed as a percentage of the component's population.

2.6.2. Network Analysis

Based on available details from the IAM's conceptual model subject areas and the connections between subject areas as outlined by the Global Forum's *Asset Management Landscape*, the adjacency matrix, shown in Appendix S1, was developed (Global Forum on Maintenance and Asset Management 2014). The connections were specified within the report for each subject area as related subjects and artefacts. The colors of the subject areas in the adjacency matrix (Appendix S1) correspond to the AM divisions from Figure 2-2. Tool 1 uses the adjacency matrix to develop the network model shown in Figure 2-9. A transmission utility AM system would typically include the AM subject areas from Figure 2-2 and Figure 2-9. The node colour refers to the AM division of the subject area. The network is directed as typically subject areas pass knowledge, information, and policy in only one direction (i.e., from source to target nodes). The link colour is the same as the source node. Using colour as a distinguishing feature allows for the identification of clusters of AM division-based subject areas. This is shown in Figure 2-9, where the *Asset Information* (purple) and *Organization & People* (red) division subject area nodes are highly interconnected within their clusters. Conversely, the *Strategy & Planning* (gold) division subject areas are not only clustered amongst themselves but instead are highly connected with other subject areas.

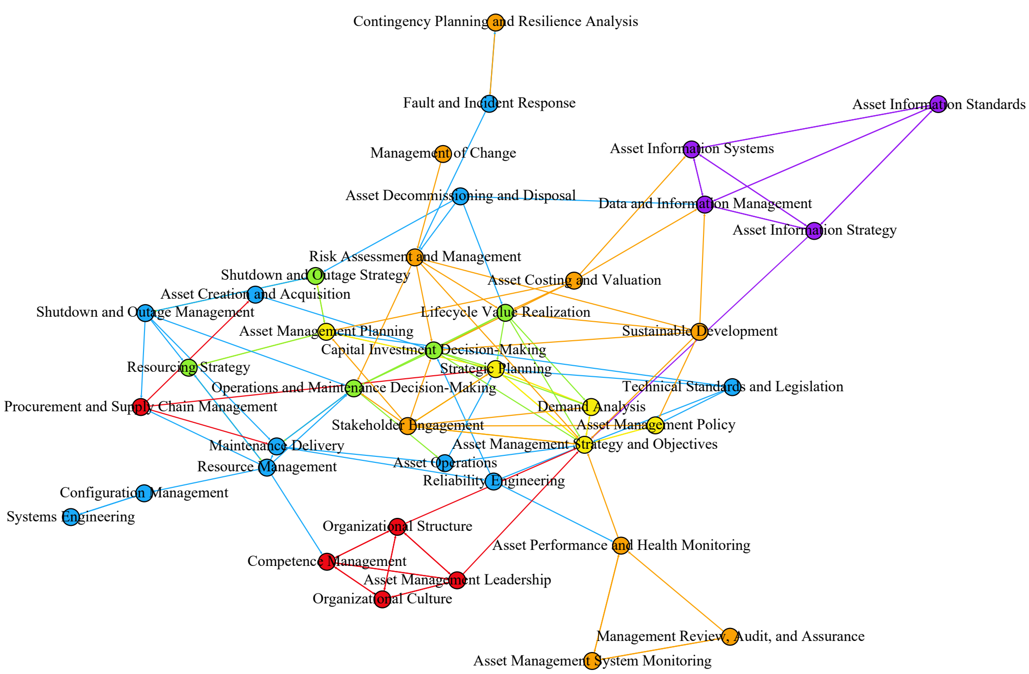


Figure 2-9. AM system network.

Network modelling is useful for viewing node connections, whereas node-, link-, and network-based centrality analyses are needed to identify the highly dependent nodes/links that induce systemic risk to the AM network. Figure 2-10 shows the top ten subject areas for each of the previously described centrality measures. Of note are the *Strategy & Planning* division subject areas of *Asset Management Strategy and Planning*, *Asset Management Planning*, and *Strategic Plan*. These AM subject areas all rank high for betweenness centrality, degree centrality, and eigenvector centrality. This indicates that for an organization to implement an effective AM system, it must have a strong AM plan and objective targets. In addition, the *Operation and Maintenance Decision-Making* and *Resourcing Strategy* of the *Asset Management Decision-Making* division were high-ranking subject areas among the centrality measures.

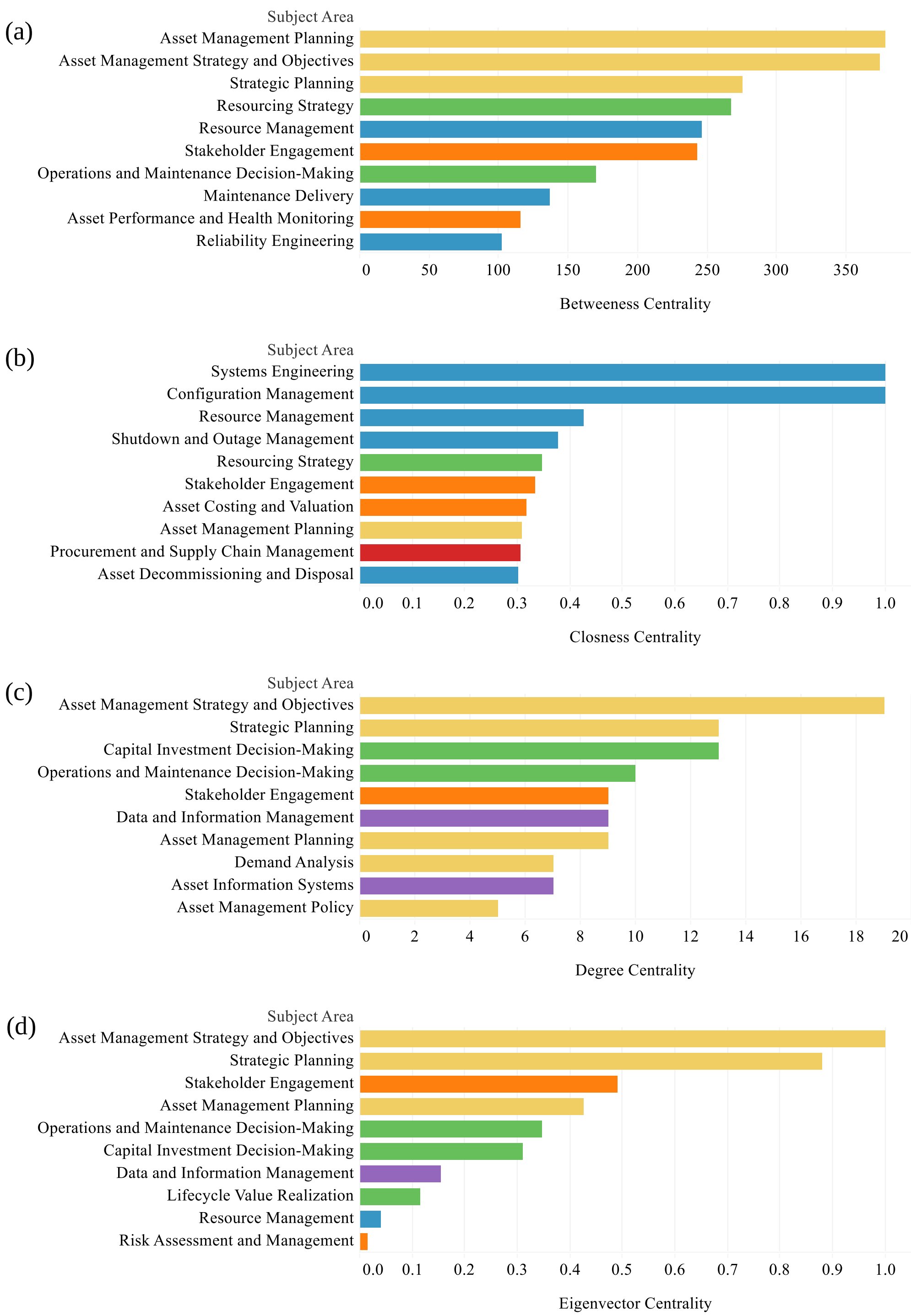


Figure 2-10. Top ten asset management subject areas ranked by centrality measures: (a) Betweenness Centrality, (b) Closeness Centrality, (c) Degree Centrality, and (d) Eigenvector Centrality.

The links shown in Table 2-2 illustrate the critical links contributing to systemic risks within the AM system. The betweenness centrality of each link was found as per the network measure previously described. The links are ordered based on their criticality, which indicates their importance to the network functionality. Of note within Table 2-2, the *Resourcing Strategy* and *Asset Management Planning* nodes have multiple important links suggesting that it is particularly important for these AM subject areas to have excellent communication with the connected AM subject areas.

Table 2-2. Top ten AM network links by betweenness centrality.

Source Node	Target Node	Betweenness Centrality
Asset Management Planning	Resourcing Strategy	283
Resourcing Strategy	Resource Management	282
Strategic Planning	Asset Management Planning	248
Asset Management Strategy and Objectives	Stakeholder Engagement	241
Maintenance Delivery	Reliability Engineering	138
Operations and Maintenance Decision-Making	Maintenance Delivery	133
Resource Management	Competence Management	133
Stakeholder Engagement	Operations and Maintenance Decision-Making	114
Asset Management Strategy and Objectives	Strategic Planning	112
Stakeholder Engagement	Asset Management Planning	110

In addition to the node-based centrality measures, the network-based measures are important to evaluate the overall resilience of the network to potential failures (Barabasi, 2016). The average degree centrality of the AM subject area network is 3.15 and the network density is 0.08, meaning that only 8% of the potential links of a fully connected network connect the AM subject areas. This implies that the network is vulnerable to systemic risk because if one or more of the previously identified critical nodes/links were to be disrupted, the AM system would be greatly impacted.

2.6.3. DESCRIPTIVE ANALYTICS

Three of the critically dependent subject areas, as determined from the node-based centrality analysis in Figure 2-10, are used to illustrate the use of descriptive analytics for subject area-paired KPI analysis. *Asset Management Strategy and Objectives*, *Asset Management Planning*, and *Operations and Maintenance Decision-Making* were chosen for illustration as these subject areas ranked high in the centrality importance measures previously analyzed. Three of the thirty-nine subject areas were chosen to illustrate the use of Tool 3 for the sake of brevity, but organizations should employ descriptive analytics to pair each subject area to at least one infrastructure AM KPI. It should be noted that insights into multiple subject areas can be taken from the same figure, as illustrated below.

Figure 2-11 illustrates the *Asset Management Strategy and Objectives* subject area focused on developing a long-term plan for managing an organization's infrastructure assets (Institute for Asset Management 2015). Figure 2-11 shows how the organization can see their AM KPIs (e.g., *median outage duration* in hours and *number of outages*) compared to other organizations. In this context, the term *organization,* shown in Figure 2-11, indicates anonymized transmission utilities that contributed outage data for all shown years. Figure 2-11 contains three main sections where a stakeholder can glean information after selecting the asset type in the view from the menu (i.e., titled: Component): 1) The top bar graph indicates the KPIs from the period as selected on the side slider menu along with the median KPI value among all organizations over the selected period (e.g., 2014-2018); 2) The line graph shows the changing KPI for each organization (colour) over the period indicated by the side slider menu; and 3) The bottom point graph is based on the organization selected by the menu (titled: Organization) and is broken down by voltage class to highlight the long-duration outage events with an option to select a specific point to display the outage features. Descriptive analytics in this application will allow the key stakeholders to track the objective KPIs while also investigating the long-duration outages to understand and remedy them.

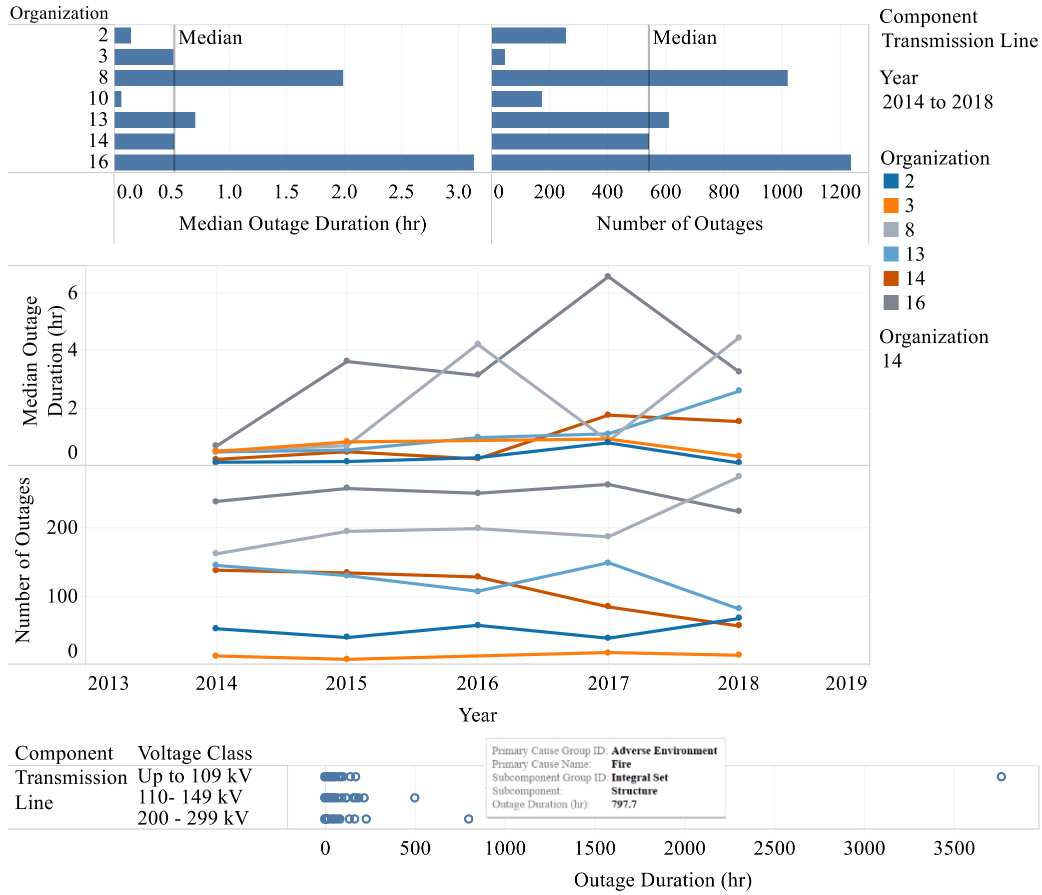


Figure 2-11. Descriptive analytics for *Asset Management Strategy & Objectives* subject area including *Median Outage Duration* and *Number of Outages* KPIs for transmission line assets.

Figure 2-12 illustrates the KPI related to two important subject areas in *Asset Management Planning* and *Operations and Maintenance Decision-Making*. The *Asset Management Planning* subject area focuses on achieving the AM objectives and identifying related risks arising from previous asset failures whereas the *Operations and Maintenance Decision-Making* subject area focuses on ensuring a predictable and acceptable level of service throughout the asset's life (Institute for Asset Management 2015). Figure 2-12 illustrates both subject areas and contains two sections: 1) The bar graph shows current year performance (blue) compared to previous year performance (grey), which can be changed using the arrow selector on the right side of the dashboard; and 2) The sparkline shows the specified subcomponent (selected from the drop-down menu: Subcomponent Group) KPI over the years specified in the side menu (titled: Year). This descriptive analytics application allows an organization to monitor the year-over-year changes in *average outage duration* to analyze previous asset failures and develop mitigation plans to ensure the issues do not arise again. In addition, the service level and trending performance of the specified asset subcomponent can be monitored to ensure that an acceptable level of service is being provided by that subcomponent and to view the impact of operations and maintenance decisions. For example, the trending performance, indicated by the sparkline, for transformer banks shows a decrease in *average outage duration*. This would indicate that the operations and maintenance decisions being made are positively affecting the performance. This descriptive analytics application would then be distributed to all AM stakeholders to see the positive impact of their coordinated effort in improving the operations and maintenance decisions.

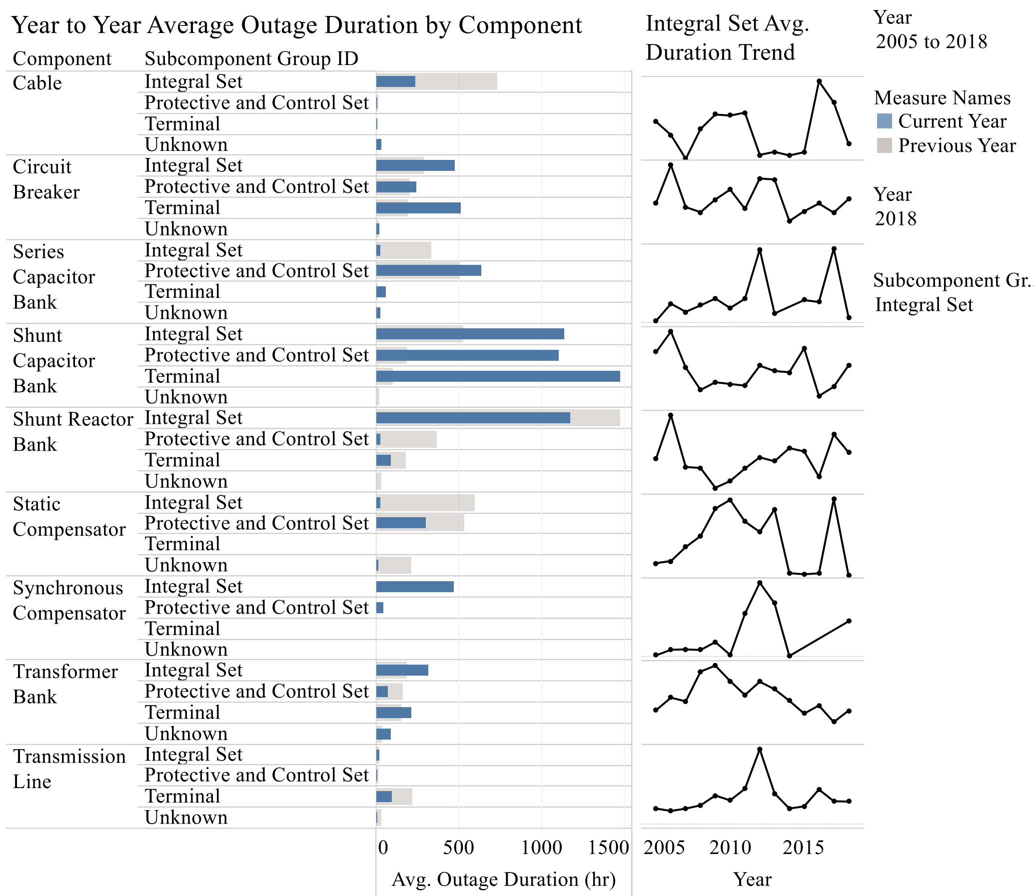


Figure 2-12. Descriptive analytics for *Asset Management Planning* and *Operations and Maintenance Decision-Making* subject area and *Average Outage Duration* KPI for each transmission asset component.

2.7. MANAGERIAL INSIGHTS

The described network analytics toolbox and the subsequent demonstration yields managerial insights that can reduce the information asymmetry between AM subject areas by targeting corresponding systemic risks within the AM system network. The following subsections coincide with the main AM divisions *Strategy & Planning*, *Asset Management Decision-Making*, *Lifecycle Delivery*, *Asset Information*, *Organization & People*, and *Risk & Review*. The insights will be presented from the viewpoint of an asset manager within an infrastructure asset-intensive organization and are transferable across infrastructure AM industries. The application of Tools 1 and 2 would be organization-specific whereas Tools 3 to 5 would have industry-specific KPIs, and additional features, as outlined in the respective tool descriptions.

2.7.1. Strategy & Planning

Based on the centrality measures analyzed in the demonstration of the AM network, it was shown that some *Strategy & Planning* subject areas were very important to the successful implementation and operation of an AM system. This can be used as evidence that the most critical aspect of a successful AM system is the development of a clear and precise strategy and plan. This ensures information symmetry between stakeholders as all stakeholders are guided by the same clearly defined strategy and plan and not overwhelmed by too much information. This will allow all other dependent lifecycle stages and decisions that follow to be guided by a clear strategy.

2.7.2. Asset Management Decision-Making

Key subject areas, based on the centrality measures, important to the effective decision-making within an AM system are *Resourcing Strategy*, *Capital Investment Decision-Making*, *Operations and Maintenance Decision-Making*, and *Lifecycle Value Realization* as shown in Figure 2-10. These critical AM subject areas highlight the necessity for all AM stakeholders to be making decisions based on the same information. Descriptive analytics applications showed how stakeholders could stay informed of current information and see the effects of their decisions on the AM KPIs. Figure 2-12 illustrated this concept by showing the KPI variation related to the specified AM subject areas, allowing stakeholders to view the impact of their AM decisions on the KPI of the assets.

2.7.3. Lifecycle Delivery

The lifecycle of assets includes acquisition, operation, maintenance, and disposal. This cycle is continuously operating within an infrastructure asset-intensive organization as infrastructure assets are at different stages of their lifecycle. As is shown in Figure 2-10, the *Lifecycle Delivery* subject areas are important for the closeness centrality measure indicating that these subject areas are important to the indirect information flow within the AM system. This means that information is not passed through a direct connection to a node, but through one or multiple other nodes. The nodes that pass information to other nodes typically process such information so that it can be readily used by the following node. For example, Figure 2-10 shows that the subject area *Fault and Incident Response* processes the information from the *Contingency Planning and Resilience Analysis* subject area before passing it onto the *Risk Assessment and Management* subject area.

2.7.4. Asset Information

Digital data related to the KPIs for infrastructure assets is critical to implement the network analytics toolbox. Asset information is most valuable in a digital format so that it can be used to track KPIs and implement the tools to improve the value assets provide throughout their lifecycle. The *Asset Information* subject areas are highly connected, as shown in Figure 2-9 and Table S1, indicating that if one were to be disrupted, then the others would also be disrupted. An asset manager should note that the design and development of a robust asset information collection, reporting, and storage system is critical to the development of informative results on asset performance, which in turn is necessary to evaluate the effectiveness of the AM plan being implemented. Throughout the lifecycle of an asset, the KPIs associated with an asset's performance is critical to be collected and stored consistently so that the KPIs can be monitored on the timeline tailored to a specific infrastructure asset (e.g., monthly, yearly, or every 5 years). This will enable Tools 3-5 to be deployed effectively to improve the AM system. In addition, having a consistent asset information reporting method allows stakeholders to have access to the same information, therefore reducing the information asymmetry.

2.7.5. Organization & People

The *Organization & People* subject areas were also clustered in the AM network, suggesting that most subject areas within this division are connected and if there is a disturbance in one then it will affect all the others. The asset manager should use this insight to ensure all stakeholders are clear of the AM strategy, goals, and implementation plan, to reduce the potential for information asymmetry between stakeholder silos. The stakeholder buy-in to implementing an AM system is critical. There needs to be strong organizational management so that stakeholders can see the positive effects of information symmetry on infrastructure asset KPIs following the implementation of an AM system.

2.7.6. RISK & REVIEW

The risk and review process is critical to the evaluation of the effectiveness of the AM plan within an organization. The use of descriptive analytics allows asset managers to efficiently evaluate the KPIs for the organization's assets. The targeted descriptive analytics applications to specified critical AM subject areas allow asset managers to concentrate on detailed information quickly to minimize the time needed to search for the result they are looking for. Descriptive analytics also allows for rapid consultation amongst stakeholders, therefore improving the review process and necessary collaboration. Deploying Tool 3 also allows for automatic updates to occur in the figures so that a stakeholder does not need to continuously update figures for use in reports. This allows all stakeholders to have access to the same information, allowing them to make decisions with the most accurate and up-to-date information.

2.8. CONCLUSION

Global infrastructure assets are continuously deteriorating, and the current condition of infrastructure is poor in both Canada and the United States. To maximize the value of each dollar spent on infrastructure for repair, rehabilitation, replacement, and maintenance, effective and efficient asset management (AM) practices are needed. One of the main challenges in implementing and operating an effective AM system within an organization is the systemic risks caused by information asymmetry between dependent AM system subject areas. This study presents a network analytics toolbox to identify the systemic risks induced within an AM system and reduce the information asymmetry by using key performance indicator (KPI) analytics paired with the critical AM subject areas. The five tools described include: 1) Dependence Identification and Network Modelling; 2) Network Centrality Analysis; 3) Descriptive Analytics of Critical Subject Area Paired KPI; 4) KPI-Based Predictive Analytics; and 5) Prescriptive Analytics for Optimal Network Configuration. Tool 1 describes how to build an AM network from an organization's AM system. The connections between the AM system subject areas are used to develop an adjacency matrix and the adjacency matrix is then used to build an AM system network. Tool 2 employs node- and network-based centrality measures to determine the most critical nodes to the operation of the AM system network. Tool 3 takes the critical AM subject areas, identified from the node-based centrality measures, and uses descriptive analytics to track KPIs that directly relate to the important AM subject areas. Tool 4 uses historical KPI values and additional influencing features within a machine learning model to predict future KPIs. Tool 5 uses the existing AM network structure and applies optimization to generate the optimal AM system network configuration to minimize the systemic risks.

The toolbox was subsequently deployed to demonstrate three of the five tools using the Institute for Asset Management's conceptual model and transmission infrastructure asset outage KPIs. Critical AM subject areas were identified through the node- and link-based centrality measures and descriptive analytics was deployed so that transmission utilities would be able to track their KPIs as they directly relate to the important subject areas. The AM subject areas were described using descriptive analytics applications. Key managerial insights for systemic risk identification associated with the AM system and reduction in information asymmetry between AM stakeholders were highlighted as they related to each of the major AM divisions.

Understandably, the implementation of the tools described within the study requires active participation among all AM stakeholders to be effective. As in all data-driven models, high-quality input data is necessary to achieve a useful output. Specific, and expected, limitations of Tool 1 relate to its dependency on an organization’s record-keeping of its AM system and/or an organization’s level of understanding of how the AM subject areas are linked together. For Tool 2, if any component (i.e., node or link) changes due to organization restructuring, then all centrality values would need to be revised. Tool 3 requires an infrastructure industry-specific expert to pair AM subject areas with relevant KPIs and there needs to be relevant data to generate KPIs within the existing database. Tool 4 is influenced by the features available for inputs to the machine learning model (e.g., if all available features are categorical then there is a limited number of machine learning models that can be used, and the output will also be categorical). Finally, the ability of Tool 5 to provide an exact (or near exact) solution might be affected by the complexity of the objective function and constraints, indicating that users might resort to heuristics, for example, to reach a solution. In addition, it is expected that organizations would implement the described tools in sequential order as indicated and become comfortable with using each tool implementation before implementing the next tool. By adopting the toolbox presented in this study, it is expected that stakeholders can reduce the systemic risks within an AM system using AM subject area-specific Tool 3 outputs that display information from a centralized database, thus ensuring that an AM subject area's information is not siloed from the overall AM system. This also ensures that AM stakeholders make decisions using a consistent information source, reducing the likelihood of stakeholders acting in silos and causing information asymmetry.

2.9. ACKNOWLEDGEMENTS

The data for this study was provided by the Canadian Electricity Association (CEA) and their support in the development of this study has been greatly appreciated. The financial support for the study was provided through the Canadian Nuclear Energy Infrastructure Resilience under Systemic Risk (CaNRisk) – Collaborative Research and Training Experience (CREATE) program of the Natural Science and Engineering Research Council (NSERC) of Canada. The support of the INTERFACE Institute and the INViSiONLab is also acknowledged in the development of this study. In addition, the authors would like to thank the

anonymous reviewers for their detailed comments in helping to clarify the systemic risk definition.

2.10. NOTATIONS

A_{ij}	Adjacency matrix with elements for nodes *i* and *j*
BC_i	Betweenness centrality of node *i*
ρ_{jk}	Number of shortest paths connecting node *j* to node *k*
$\rho_{jk}(i)$	Number of shortest paths that connect node *j* to node *k* that traverse node *i*
CC_i	Closeness centrality of node *i*
$d(i,j)$	Shortest path distance between nodes *i* and *j*
N	Total number of network nodes
DC_i	Degree centrality of node *i*
x_i	Relative eigenvector centrality score
λ	Eigenvalue
ND	Network density
l	Number of links in the network
ADC	Average degree centrality

2.11. ACRONYMS

AM	Asset management
ASCE	American Society of Civil Engineers
CEA	Canadian Electricity Association
IAM	Institute for Asset Management
KPI	Key performance indicator
LBC	Link betweenness centrality

2.12. Supplemental Materials

Table S1. Adjacency matrix indicating interconnections between asset management subject areas.

Target Source	Asset Management Policy	Asset Management Strategy and Objectives	Demand Analysis	Strategic Planning	Asset Management Planning	Capital Investment Decision-Making	Operations and Maintenance Decision-Making	Lifecycle Value Realization	Resourcing Strategy	Shutdown and Outage Strategy	Technical Standards and Legislation	Asset Creation and Acquisition	Systems Engineering	Configuration Management	Maintenance Delivery	Reliability Engineering	Asset Operations	Resource Management	Shutdown and Outage Management	Fault and Incident Response	Asset Decommissioning and Disposal	Asset Information Strategy	Asset Information Standards	Asset Information Systems	Data and Information Management	Procurement and Supply Chain Management	Asset Management Leadership	Organizational Structure	Organizational Culture	Competence Management	Risk Assessment and Management	Contingency Planning and Resilience Analysis	Sustainable Development	Management of Change	Asset Performance and Health Monitoring	Asset Management System Monitoring	Management Review, Audit, and Assurance	Asset Costing and Valuation	Stakeholder Engagement
Asset Management Policy		1																																					
Asset Management Strategy and Objectives	1		1	1																																			1
Demand Analysis		1		1																																			
Strategic Planning		1	1		1																																		1
Asset Management Planning				1					1	1																													
Capital Investment Decision-Making		1	1	1			1	1																															
Operations and Maintenance Decision-Making						1									1		1																						
Lifecycle Value Realization		1	1	1		1	1																																
Resourcing Strategy					1													1								1													
Shutdown and Outage Strategy					1														1							1													
Technical Standards and Legislation	1	1		1	1																																		
Asset Creation and Acquisition						1																																	
Systems Engineering														1																									
Configuration Management													1																										
Maintenance Delivery							1						1			1	1																						
Reliability Engineering		1				1							1																						1				
Asset Operations		1		1																																			
Resource Management							1		1					1	1											1				1									
Shutdown and Outage Management							1		1	1					1											1													
Fault and Incident Response																															1	1							
Asset Decommissioning and Disposal								1		1															1						1								
Asset Information Strategy	1																						1	1	1														
Asset Information Standards																						1		1	1														
Asset Information Systems																						1	1		1														
Data and Information Management																						1	1	1															
Procurement and Supply Chain Management				1								1			1																								
Asset Management Leadership		1																										1	1	1									
Organizational Structure		1																									1		1	1									
Organizational Culture																											1	1		1									
Competence Management																											1	1	1										
Risk Assessment and Management		1				1	1	1																															
Contingency Planning and Resilience Analysis																				1																			
Sustainable Development	1	1				1		1																	1						1								
Management of Change																															1								
Asset Performance and Health Monitoring		1																																		1			
Asset Management System Monitoring																																			1		1		
Management Review, Audit, and Assurance																																			1	1			
Asset Costing and Valuation					1	1		1																1	1														
Stakeholder Engagement	1	1	1	1	1	1	1																																

2.13. References

Abdelfatah, M., El-Shimy, M., and Ismail, H. M. (2013). "Outage data analysis of utility power transformers based on outage reports during 2002-2009." International Journal of Electrical Power and Energy Systems, Elsevier, 47(1), 41–51.

Abu-Samra, S., Ahmed, M., and Amador, L. (2020). "Asset Management Framework for Integrated Municipal Infrastructure." Journal of Infrastructure Systems, 26(4), 04020039.

Aggarwal, C. C. (2015). Data Mining: The textbook. Springer International Publishing, New York.

Alzoor, F. S., Ezzeldin, M., Mohamed, M., and El-Dakhakhni, W. (2021). "Prioritizing Bridge Rehabilitation Plans through Systemic Risk-Guided Classifications." Journal of Bridge Engineering, 26(7), 04021038.

American Society of Civil Engineers. (2021). 2021 Report Card for America's Infrastructure.

Barabási, A.-L. (2016). Network Science. Cambridge University Press, Cambridge, UK.

Barão, A., de Vasconcelos, J. B., Rocha, Á., and Pereira, R. (2017). "A knowledge management approach to capture organizational learning networks." International Journal of Information Management, Elsevier, 37(6), 735–740.

Barker, K., Lambert, J. H., Zobel, C. W., Tapia, A. H., Ramirez-Marquez, J. E., Albert, L., Nicholson, C. D., and Caragea, C. (2017). “Defining resilience analytics for interdependent cyber-physical-social networks.” Sustainable and Resilient Infrastructure, Taylor & Francis, 2(2), 59–67.

Bergh, D. D., Ketchen, D. J., Orlandi, I., Heugens, P. P. M. A. R., and Boyd, B. K. (2019). “Information Asymmetry in Management Research: Past Accomplishments and Future Opportunities.” Journal of Management, 45(1), 122–158.

Black, J., Hoffman, A., Hong, T., Roberts, J., and Wang, P. (2018). “Weather Data for Energy Analytics: From Modeling Outages and Reliability Indices to Simulating Distributed Photovoltaic Fleets.” IEEE Power and Energy Magazine, 16(3), 43–53.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U. (2006). “Complex networks: Structure and dynamics.” Physics Reports, 424(4–5), 175–308.

Brunetto, Y., Xerri, M., and Nelson, S. (2014). “Building a Proactive, Engagement Culture in Asset Management Organizations.” Journal of Management in Engineering, 30(4), 04014014.

Canadian Infrastructure Report Card. (2019). Monitoring the State of Canada’s Core Public Infrastructure: The Canadian Infrastructure Report Card 2019.

CEA. (2018). Instruction Manual for Reporting Component Forced Outages for Transmission Equipment. Ottawa, ON.

CEA. (2020). “Overview - Canadian Electricity Association.” <https://electricity.ca/about-cea/overview/#about> (Apr. 16, 2020).

Chen, L., Henning, T. F. P., Raith, A., and Shamseldin, A. Y. (2015). "Multiobjective Optimization for Maintenance Decision Making in Infrastructure Asset Management." Journal of Management in Engineering, 31(6), 04015015.

Crowther, K. G. (2008). "Decentralized risk management for strategic preparedness of critical infrastructure through decomposition of the inoperability input-output model." International Journal of Critical Infrastructure Protection, Elsevier, 1(C), 53–67.

Das, K., Samanta, S., and Pal, M. (2018). "Study on centrality measures in social networks: a survey." Social Network Analysis and Mining, Springer Vienna, 8(1), 1–11.

Dehghanian, P., Zhang, B., Dokic, T., and Kezunovic, M. (2019). "Predictive Risk Analytics for Weather-Resilient Operation of Electric Power Systems." IEEE Transactions on Sustainable Energy, 10(1), 3–15.

Delen, D., Moscato, G., and Toma, I. L. (2018). "The impact of real-time business intelligence and advanced analytics on the behaviour of business decision makers." 2018 International Conference on Information Management and Processing, IEEE, 49–53.

Delen, D., and Ram, S. (2018). "Research challenges and opportunities in business analytics." Journal of Business Analytics, Taylor & Francis, 1(1), 2–12.

Derrible, S. (2012). "Network centrality of metro systems." PLoS ONE, 7(7), 1–10.

Eisenberg, D. A., Park, J., and Seager, T. P. (2020). "Linking Cascading Failure Models and Organizational Networks to Manage Large-Scale Blackouts in South Korea." Journal of Management in Engineering, 36(5), 04020067.

Estrada, E., and Knight, P. (2015). A first course in network theory. Oxford University Press.

Ezzeldin, M., and El-Dakhakhni, W. E. (2019). “Robustness of Ontario power network under systemic risks.” Sustainable and Resilient Infrastructure, Taylor & Francis, 1–20.

Freeman, L. C. (1977). “A Set of Measures of Centrality Based on Betweenness.” Sociometry, 40(1), 35–41.

Global Forum on Maintenance and Asset Management. (2014). The Asset Management Landscape. Global Forum on Maintenance and Asset Management.

Goforth, E., Ezzeldin, M., El-Dakhakhni, W., Wiebe, L., and Mohamed, M. (2020). “Network-of-Networks Framework for Multimodal Hazmat Transportation Risk Mitigation: Application to Used Nuclear Fuel in Canada.” Journal of Hazardous, Toxic, and Radioactive Waste, 24(3).

Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Optimization, and Machine Learning, Addison-Wesley Longman Publishing Co. Inc., Boston, MA.

Golightly, D., Kefalidou, G., and Sharples, S. (2018). “A cross-sector analysis of human and organisational factors in the deployment of data-driven predictive maintenance.” Information Systems and e-Business Management, Springer Berlin Heidelberg, 16(3), 627–648.

Gondia, A., Ezzeldin, M., and El-Dakhakhni, W. (2021). "A Dynamic Network Toolbox for Resilience-Driven Management of Infrastructure Projects." Automation in Construction, submitted, 1–29.

Graybeal, P., Franklin, M., and Cooper, D. (2018). Principles of Accounting, Volume 2: Managerial Accounting. (OpenStax, ed.), OpenStax , Houston, TX.

Haggag, M., Yorsi, A., El-Dakhakhni, W., and Hassini, E. (2021). "Infrastructure performance prediction under Climate-Induced Disasters using data analytics." International Journal of Disaster Risk Reduction, Elsevier Ltd., 56(November 2020), 102121.

Hassan, N. R. (2019). "The origins of business analytics and implications for the information systems field." Journal of Business Analytics, Taylor & Francis, 2(2), 118–133.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning. Springer Series in Statistics, Springer New York, New York, NY.

Heng, F.-L., Zhang, K., Goyal, A., Chaudhary, H., Hirsch, S., Kim, Y., Lavin, M. A., and Raman, A. (2016). "Integrated analytics system for electric industry asset management." IBM Journal of Research and Development, 60(1.2), 1–12.

Herrera, F., Chan, G., Legault, M., Kassim, R. M., and Sharma, V. (2011). The digital workplace: Think, share, do. Deloitte & Touche LLP.

Huang, K., Li, M., and Markov, S. (2019). "The Information Asymmetry between Management and Rank-and-File Employees: Determinants and Consequences." AAA 2019 Management Accounting Section (MAS) Meeting, 1–53.

Infrastructure Canada. (2018). Investing in Canada: Canada’s long-term infrastructure plan. Ottawa.

Institute for Asset Management. (2015). “Asset Management – an Anatomy Ver3.” <theIAM.org/AMA>.

Institute for Asset Management. (2021). “About the IAM.” <https://theiam.org/about-us/> (May 25, 2021).

Koziel, S., Hilber, P., Westerlund, P., and Shayesteh, E. (2021). “Investments in data quality: Evaluating impacts of faulty data on asset management in power systems.” Applied Energy, Elsevier Ltd., 281(November 2020).

de la Pena, C., Gonzalez Fernandez, D., and Rodriguez Gonzalez, J. (2016). “How analytics can improve asset management in electric power networks.” McKinsey & Company Electrical Power & Natural Gas, <https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/> (Jun. 14, 2018).

Lee, S. H., Choi, J. Y., Yoo, S. H., and Oh, Y. G. (2013). “Evaluating spatial centrality for integrated tourism management in rural areas using GIS and network analysis.” Tourism Management, Elsevier Ltd, 34, 14–24.

López-Espinosa, G., Moreno, A., Rubia, A., and Valderrama, L. (2015). “Systemic risk and asymmetric responses in the financial industry.” Journal of Banking and Finance, 58, 471–485.

McBride, J., and Moss, J. (2020). “The State of U.S. Infrastructure.” Council on Foreign Relations, <https://www.cfr.org/backgrounder/state-us-infrastructure> (Apr. 6, 2021).

McDowell, T., Horn, H., and Witkowski, D. (2016). Organizational Network Analysis.

Miller, B. B. (2017). “Information Asymmetries Conceal Fraud and Systemic Risks in the U.S. Banking Industry.” The Harvard Law School Forum on Corporate Governance, The Harvard Law School Forum on Corporate Governance, <https://corpgov.law.harvard.edu/2017/08/19/information-asymmetries-conceal-fraud-and-systemic-risks-in-the-u-s-banking-industry/> (Jul. 29, 2021).

Mohamad, A., Zainuddin, Y., Alam, N., and Kendall, G. (2017). “Does decentralized decision making increase company performance through its Information Technology infrastructure investment?” International Journal of Accounting Information Systems, 27(October), 1–15.

Mukherjee, S., Nateghi, R., and Hastak, M. (2018). “A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S.” Reliability Engineering and System Safety, Elsevier Ltd, 175, 283–305.

Nearchou, A. C. (2004). “The effect of various operators on the genetic search for large scheduling problems.” International Journal of Production Economics, 88, 191–203.

O’Neill, M., and Brabazon, A. (2019). “Business analytics capability, organisational value and competitive advantage.” Journal of Business Analytics, Taylor & Francis, 2(2), 160–173.

Opdyke, A., Lepropre, F., Javernick-Will, A., and Koschmann, M. (2017). “Inter-organizational resource coordination in post-disaster infrastructure recovery.” Construction Management and Economics, Routledge, 35(8–9), 514–530.

Pell, R., Svoboda, R., Eagar, R., Ondko, P., and Kirschnick, F. (2015). Effective Infrastructure Asset Management- A holistic approach to transformation. Arthur D. Little - Prism.

Piryonesi, S. M., and El-Diraby, T. E. (2020). "Data Analytics in Asset Management: Cost-Effective Prediction of the Pavement Condition Index." Journal of Infrastructure Systems, 26(1), 04019036.

Prajogo, D., Toy, J., Bhattacharya, A., Oke, A., and Cheng, T. C. E. (2018). "The relationships between information management, process management and operational performance: Internal and external contexts." International Journal of Production Economics, Elsevier Ltd, 199, 95–103.

Qiu, Q., Fleeman, J. A., Ball, D. R., Rackliffe, G., Hou, J., and Cheim, L. (2013). "Managing critical transmission infrastructure with advanced analytics and smart sensors." IEEE Power and Energy Society General Meeting, IEEE, 1–6.

Ross, R. (2019). Reliability Analysis for Asset Management of Electric Power Grids. John Wiley & Sons Ltd., Singapore.

Salama, M., Ezzeldin, M., El-Dakhakhni, W., and Tait, M. (2020). "Temporal networks: a review and opportunities for infrastructure simulation." Sustainable and Resilient Infrastructure, Taylor & Francis, 1–16.

Scheibe, K. P., Nilakanta, S., Ragsdale, C. T., and Younie, B. (2019). "An evidence-based management framework for business analytics." Journal of Business Analytics, Taylor & Francis, 2(1), 47–62.

Scrucca, L. (2013). "GA: A package for genetic algorithms in R." Journal of Statistical Software, 53(4), 1–37.

Talebiyan, H., and Duenas-Osorio, L. (2020). "Decentralized Decision Making for the Restoration of Interdependent Networks." ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering, 6(2), 04020012.

Tarver, E. (2020). "How Financial Markets Exhibit Asymmetric Information." Investopedia, <https://www.investopedia.com/ask/answers/042915/how-do-financial-market-exhibit-asymmetric-information.asp> (Aug. 30, 2021).

Teixeira, A. S., Santos, F. C., and Francisco, A. P. (2016). "Spanning edge betweenness in practice." Studies in Computational Intelligence, 644, 3–10.

Thai, M. T., and Pardalos, P. M. (2012). Handbook of Optimization in Complex Networks: Communication and Social Networks. (M. T. Thai, P. M. Pardalos, and Laboratory of Algorithms and Technologies for Networks Analysis, eds.), Springer Science+Business Media, New York.

Tsai, C. W., Lai, C. F., Chao, H. C., and Vasilakos, A. V. (2015). "Big data analytics: a survey." Journal of Big Data, Springer International Publishing, 2(1), 1–32.

Uddin, W., Hudson, W. R., and Haas, R. (2013). Public Infrastructure Asset Management. McGraw-Hill Education, New York.

Ujwary-Gil, A. (2019). "Organizational network analysis: A study of a university library from a network efficiency perspective." Library and Information Science Research, Elsevier, 41(1), 48–57.

Valentin, V., Naderpajouh, N., and Abraham, D. M. (2018). “Integrating the Input of Stakeholders in Infrastructure Risk Assessment.” Journal of Management in Engineering, 34(6), 04018042.

Vantrappen, H., and Wirtz, F. (2017). “When to Decentralize Decision Making, and When Not To.” Harvard Business Review, <https://hbr.org/2017/12/when-to-decentralize-decision-making-and-when-not-to> (Jul. 29, 2021).

Weldon, W. (2008). “Johnson & Johnson CEO William Weldon: Leadership in a Decentralized Company.” 2008 Wharton Leadership Conference, Philadelphia, PA.

Wexler, S., Shaffer, J., and Cotgreave, A. (2017). The Big Book of Dashboards. John Wiley & Sons, Inc., Hoboken.

Xerri, M. J., Nelson, S., and Brunetto, Y. (2015). “Importance of Workplace Relationships and Attitudes toward Organizational Change in Engineering Asset-Management Organizations.” Journal of Management in Engineering, 31(5), 04014074.

Yang, S., Zhou, W., Zhu, S., Wang, L., Ye, L., Xia, X., and Li, H. (2019). “Failure probability estimation of overhead transmission lines considering the spatial and temporal variation in severe weather.” Journal of Modern Power Systems and Clean Energy, Springer Berlin Heidelberg, 7(1), 131–138.

Yosri, A., Elleathy, Y., Hassini, S., and El-Dakhakhni, W. (2021). “Genetic Algorithm-Markovian Model for Predictive Bridge Asset Management.” Journal of Bridge Engineering, 26(8), 04021052.

Zhou, K., Fu, C., and Yang, S. (2016). "Big data driven smart energy management: From big data to big insights." Renewable and Sustainable Energy Reviews, Elsevier Ltd, 56, 215–225.

Zumel, N., and Mount, J. (2020). Practical Data Science with R. Manning Publications Co., Shelter Island.

Chapter 3

RAPIDITY PREDICTION FOR POWER INFRASTRUCTURE FORCED OUTAGES: A DATA-DRIVEN APPROACH

ABSTRACT

Power infrastructure is essential for the operation of almost all other critical infrastructure systems, including water, transportation, and telecommunication. Recently, there has been an increase in forced power outage frequency and extent due to infrastructure aging, extreme weather events, and deliberate attacks. To combat forced power outage risks, researchers have been focusing their attention on improving the resilience of different power infrastructure systems. A key aspect of infrastructure resilience is the rapidity, defined as the time required to return to normal operation levels following functionality disruptions. This study develops a machine learning-based framework to predict the rapidity of power infrastructure following forced outages. The framework includes classification models such as bagging, random forests, and artificial neural networks to accommodate the categorical nature of typical power infrastructure component outage features. The framework also includes a genetic algorithm for optimized selection of such features in order to facilitate the model’s best prediction performance. The utility of the developed framework is demonstrated using actual transmission line forced outages data. Within the demonstration application, the rapidity is split into two classes indicating short and extended outages and the random forest classification model is found to have the best rapidity prediction performance. In addition, key features pertaining to outage classification are explored using partial dependence analysis. Finally, insights for resilience-guided asset management are presented. The developed framework enables infrastructure stakeholders to predict forced outage rapidity classes soon after the occurrence of the former —subsequently enabling rapid identification of appropriate resources needed to promptly restore infrastructure functionality and thus ensuring infrastructure resilience.

KEYWORDS: classification, machine learning, outage analysis, power infrastructure, rapidity, resilience

3.1. INTRODUCTION

Continuously operating power infrastructure is crucial for the functionality of almost all other critical infrastructure systems (e.g., water, transportation, and telecommunication) (Haggag et al. 2020). The frequency and severity of major outages impacting power infrastructure have been consistently increasing over the past few decades due to aging, operation errors, severe climatic changes, and deliberate attacks (Bhusal et al. 2020; Haggag et al. 2021b). As such, enhancing the resilience of power infrastructure continues to be on the forefront of infrastructure research efforts (Idaho National Laboratory 2010; Preston et al. 2016; Canadian Electricity Association 2016; Raoufi et al. 2020). In power system applications, resilience has been defined using different aspects, but most definitions include the ability to quickly recover/rebound from adverse events (Panteli and Mancarella 2015a; Ciapessoni et al. 2019; Gholami et al. 2018).

For the purposes of this study, resilience will be discussed and quantified with respect to its goals (i.e., robustness and rapidity) as well as its means (i.e., redundancy and resourcefulness) (Bruneau et al. 2003; Panteli and Mancarella 2015b; Gholami et al. 2018; Salem et al. 2020). In this respect, *robustness* is defined as the ability of a component to maintain operation while experiencing disruptions; *rapidity* is the time taken to recover from such disruptions and return to the normal (or near normal) operation levels; *redundancy* is the capability to deliver the intended function provided that some components have experienced a disruption, degradation, or loss of functionality; and finally, *resourcefulness* is the capacity to restore service such that the component could return to a normal operation level (Bruneau et al. 2003; Bie et al. 2017). The relationship between the resilience goals is presented in Figure 3-1, where the *resilience trapezoid* indicates the robustness as the percentage of the remaining functioning power components and the rapidity as the total time to recover following disruptions (Jufri et al. 2019). Improving power infrastructure resilience therefore involves minimizing the area of the resilience trapezoid through maximizing the robustness and minimizing the rapidity following forced outages. However, the development of an effective resilience enhancement strategy necessitates first the accurate estimation of the robustness and rapidity following such forced outages (Bhusal et al. 2020). This study focuses on predicting the resilience metric of rapidity, using predictive analytics (i.e., data-driven models), which is key to enhance the response of the utility owner immediately following outages through quick allocation of the necessary resources to restore service.

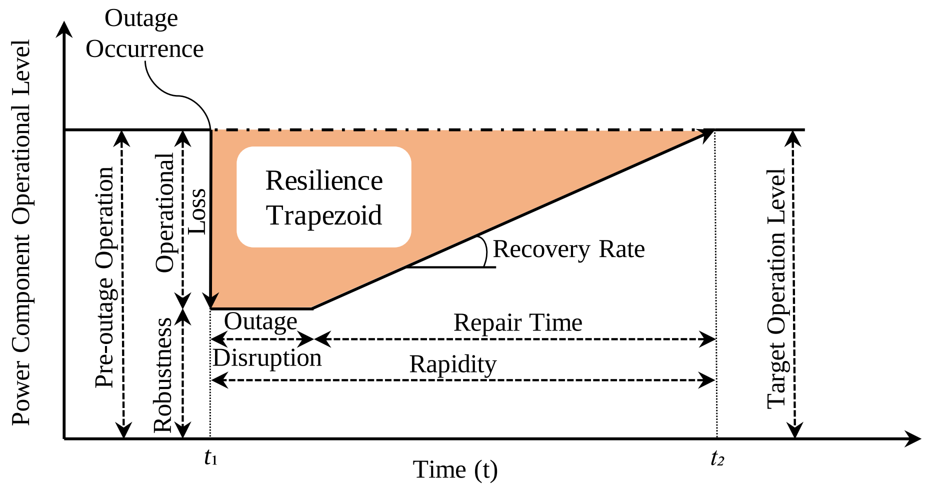


Figure 3-1. Component resilience concept and goal metrics.

Analytics involves collecting, cleaning, processing, visualizing, and analyzing related datasets, and subsequently gaining informative insights to make effective decisions (Aggarwal 2015). Owing to the large volume of spatio-temporal data generated by utility companies, it is challenging to develop pertinent usable insights for predicting power infrastructure key performance indicators (Zhou et al. 2016; Zhang et al. 2019). *Predictive analytics* has shown promise in deriving insights from data through the development of machine learning models (Goyal et al. 2016; Delen and Ram 2018; Davila Delgado et al. 2020). Such models are generally classified as supervised (i.e., developing mathematical functions that map the relationship between input-output pairs) or unsupervised (i.e., categorizing dataset observations based on similarity without specifying any certain outputs) (Zumel and Mount 2020).

The application of predictive analytics to estimate resilience goal metrics for power infrastructure has mainly focused on predicting the outage occurrence and severity (i.e., robustness) under natural disasters and adverse events. For example, Nateghi (2018) built an ensemble model for a power distribution network in the Central Gulf Coast Region of the U.S. to predict the number of outages, customers without power, and cumulative outage duration in each distribution cell under a simulated hurricane Katrina scenario. Eskandarpour et al. (2017) developed a support vector machine (SVM) model to predict distribution power grid outages due to hurricanes. Mukherjee et al. (2018) integrated climatic, electricity consumption, customers served, economic characteristics, population, and land and water mass information within an SVM model to classify the distribution power outage severity and subsequently used a random forest model to identify major risk factors. Dokic et al. (2019) used logistic regression with input features of weather data and spatial characteristics of transmission lines and substations to predict the outage occurrences. Omran and El Houby (2020) classified the type of electrical disturbances in North America by integrating feature selection with five machine learning models for classification, including k-nearest neighbour, artificial neural network, decision tree, logistic regression, and naïve Bayes. Xie et al. (2020) presented a review of machine learning methods used to assess and control the power network stability and restore the system to operation level following adverse events.

Although the above studies highlighted the suitability of machine learning approaches for the prediction of outage occurrences and severity (i.e., robustness) for power infrastructure components, a systematic machine learning-based framework to predict the rapidity metric of resilience is yet to be developed and tested. Therefore, this study presents a framework that can be deployed using power infrastructure component forced outage data, if such data is available, to classify the rapidity of a forced outage based on its contributing features that are known soon after an outage occurrence. The paper first describes the study goal and objectives, then outlines the features of the framework and describes the unique nature of the features within the databases. Subsequently, the framework utility is illustrated using actual transmission line forced outage data. A classification model is developed to predict the rapidity classes. The importance of the contributing features is further explored as they relate to extended forced outages. Finally, managerial insights are drawn with respect to the resilience-guided asset management based on the results of the feature importance and rapidity classification.

3.2. Study Goal and Objectives

The study goal is to present a framework to predict the rapidity of forced power infrastructure outages using key contributing features known soon after the onset of such outages. To achieve this goal, the study objectives include: 1) Developing machine learning models to classify the rapidity following a forced outage using only categorial input features that would be known by a utility soon after the onset of the outage; and 2) Identifying key rapidity-critical features, allowing utilities to adjust their asset management strategy to mitigate the impact of such risks in the future; and 3) Employing historical datasets pertaining to transmission infrastructure forced outage events to demonstrate the applicability of the aforementioned objectives and draw relevant managerial insights.

3.3. FRAMEWORK DEVELOPMENT

The developed framework, shown in Figure 3-2, summarizes the steps of a data-guided strategy to classify the rapidity of power infrastructure components based on existing databases (Step 1). For example, two of the main transmission outage databases in North America are produced by the Canadian Electricity Association (CEA) and the North American Electric Reliability Corporation's Transmission Availability Data System (NERC TADS). CEA and NERC TADS started recording outage events in 1978 and 2008, respectively (Papic et al. 2016), and several studies have employed these databases to estimate the reliability metrics of power transmission assets (Schaller 2012; Papic et al. 2017; Bian et al. 2014; Papic et al. 2014; Ekisheva and Gugel 2015a; Ekisheva and Gugel 2015b; Ekisheva et al. 2016; Schaller and Ekisheva 2016; Papic et al. 2018; Ekisheva et al. 2018). The CEA database includes a record of all forced outage events from participating utilities across Canada for transmission components with an operating voltage of 60 kV and above (CEA 2018). Specific components include: transmission lines, cables, circuit breakers, synchronous and static compensators, as well as transformer-, shunt reactor-, shunt capacitor-, and series capacitor-banks. Similarly, the NERC TADS database includes a record of forced outage events for transmission components that operate at AC and DC voltages ≥ 200 kV (North American Electric Reliability Corporation 2007). The transmission components include AC circuits (overhead and underground), transformers, AC/DC back-to-back converters, and DC circuits. The described framework from Figure 3-2 can be deployed on any power infrastructure forced outage database that includes details similar to those described in the previously presented examples.

Step 2 of the Figure 3-2 framework presents feature processing that may include data cleaning, imputation, and preparation as the historical databases are not readily available in a format that can be directly employed within a suitable machine learning classification model. Data cleaning involves the removal of missing observations or replacing them with a representative statistic (e.g., mean or median) (Fujikawa and Ho 2002). The latter process is typically referred to as data imputation and can be applied in an unsupervised fashion through clustering the data and replacing the missing values accordingly (Li et al. 2005; Patil et al. 2010). Following data cleaning and missing values imputation, the data should be organized using appropriate feature labels so that the features can be tracked throughout the modeling process. Finally, the feature values must be in a format compatible with the machine learning models to be developed (e.g., categorical or numerical).

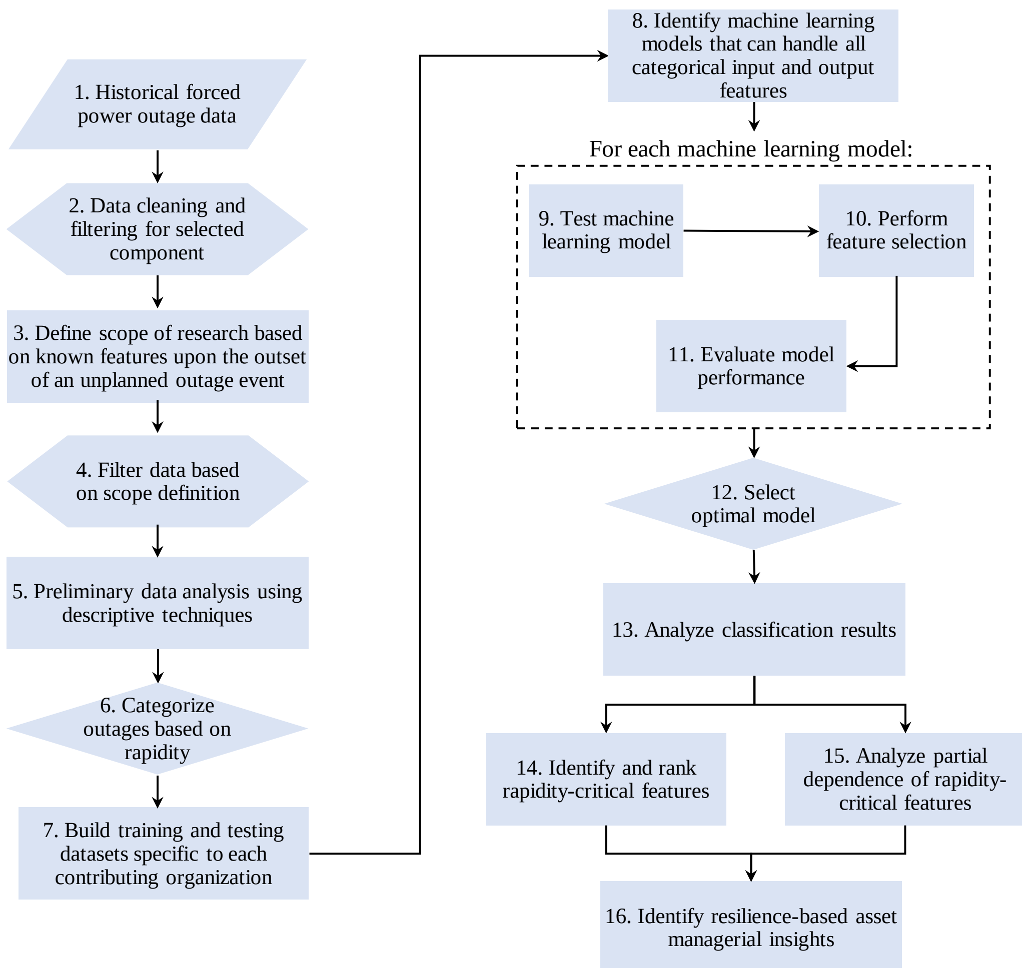


Figure 3-2. Framework for developing machine learning classification models for forced outage rapidity prediction considering categorial input and output features.

Step 3 in the Figure 3-2 framework is the scope definition, considering that the framework was developed so that utility companies are able to predict the rapidity of outages soon after their occurrence. Step 4 of the framework involves filtering the data based on the scope definition. Some features recorded for each outage are not influencing inputs per se, and therefore those features should be excluded when the classification model is developed. For example, the year of the outage is an irrelevant input feature for classification as each year starts a new set of outages and there is no correlation between a given year to the real-time prediction of the outage. Instead, the month, day, and hour can be included as time-indication input features.

Typical power infrastructure forced outage databases (e.g., CEA and NERC TADS) are distinctive in that they include mostly categorical features in conjunction with the actual duration of the forced outage. This type of dataset presents a unique challenge as correlation between features cannot be evaluated through classical measures (e.g., Pearson, Kendall, and Spearman correlation coefficients), as such measures are not applicable when categorical features are of interest. Therefore, Step 5 in the framework performs exploratory data analysis using descriptive analytics to provide the analyst with preliminary insights on feature values and trends in the data based on associated categorical features. For example, descriptive analytics might involve investigating the population descriptions and distribution of the output feature (e.g., rapidity density curve) or analyzing the association between the outage characteristics and contributing conditions (e.g., cause, location, failure, and fault) (Abdelfatah et al. 2013; Barker et al. 2017; Mukherjee et al. 2018; Black et al. 2018).

Once the input feature set has been filtered and descriptive analysis has been applied on the output feature, the output feature can be categorized into classes related to the rapidity of each forced outage instance. Databases typically do not directly define a rapidity metric, nonetheless the outage duration can be used to reflect the rapidity as it reflects the time spent by the utility to restore the service of the component following forced outage events. Therefore, the numerical values associated with forced outage durations are mapped to rapidity classes because all other features in typical power infrastructure forced outage databases are categorical. Step 6 of the framework involves identifying the number of classes for the rapidity feature with specific reference to the project objective. For example, the framework application demonstration to follow splits the rapidity into two classes, with 75% of the data in one class and 25% in the other, as the demonstration goal is to correctly predict the outages in the second class (i.e., the extended forced outages). While predicting the actual rapidity value is crucial, this is challenging particularly when the input features are entirely categorical (e.g., CEA and NERC TADS databases), as inputs would need to be mapped into continuous dummy counterparts in order to relate them to rapidity in a regression fashion. Alternatively, as adapted in the current study, forced outage durations (i.e., rapidity following a forced outage) can be classified into short or extended classes, where an accurate prediction of the outage class can help the service providers to efficiently allocate their resources and thus reduce the expected rapidity.

Step 7 in the Figure 3-2 framework requires identifying the training and testing subsets for each data-contributing organization. This is a key step as typical power infrastructure forced outage datasets contain multiple organizations' outage records, and thus it is important to separate each organization's outages in order to develop more relevant machine learning models. This step is also important as each utility operates in a different geographic area with different climatic and operation conditions. In addition, the infrastructure assets may have been constructed at different times with different systems, manufacturers, and maintenance practices—yielding key differences between the data-contributing organizations that subsequently necessitates the creation of a model specific to each organization.

3.3.1. Machine learning methods for classification

Step 8 in the Figure 3-2 framework specifies that machine learning classification models be identified that can handle the entirely categorical input and output feature set. Classification is a supervised learning method where input and output pairs are used for developing, testing, and validating the machine learning model. Specifically, a classification model aims at using input features to predict the output feature values (e.g., rapidity) (Aggarwal 2015). This framework presents decision tree-based ensemble machine learning models (i.e., bagging and random forest) and artificial neural network models as methods for handling categorical feature sets. It should be highlighted that other machine learning methods (e.g., support vector machines, Naïve Bayes, and boosting) can be used for classification modelling using categorical feature sets; however, the methods employed in the present study were chosen because their performance has been confirmed in similar applications involving categorical feature sets (Chi et al. 2012; Aggarwal 2015; Gondia et al. 2020; Zumel and Mount 2020; Haggag et al. 2021a; Haggag et al. 2021b).

Decision trees rely on hierarchical tree-like decisions based on the values of the input features (Zumel and Mount 2020). Each decision tree split is based on the full set of input feature values and is specific to one or multiple input feature values. On their own, decision trees do not often capture the output feature values accurately; therefore, ensemble methods are often used to improve the performance of a decision tree-based classification model (James et al. 2013). The framework presented in this study discusses only bagging and random forests ensemble methods for brevity; however, other ensemble methods are provided in detail in Aggarwal (2015). Bagging employs sampling with replacement from the training data to generate *t* bootstrapped classifiers (Zumel and Mount 2020). The final classification result is the average of the k classification processes. Random forests are an improvement to bagging as they de-correlate the resulting prediction by only allowing the model to split the decision tree using a specified number of features, *m* (Zumel and Mount 2020). For classification problems, the initial value of *m* is $\sqrt{q}$, where *q* is the number of features in the dataset. This initial *m* value is then adjusted iteratively in order to improve the model performance.

Artificial neural networks are machine learning models developed based on the human nervous system. A network is composed of an input layer, one or more hidden layers, and an output layer and a set of neurons are allocated to each hidden layer and are connected to subsequent and following layers through synapses. Each neuron receives inputs from subsequent neurons, performs computations on these inputs, and then passes them along to the following neurons (Aggarwal 2015). The computation occurring at each neuron is defined by the input connection weights, each of which can be seen as the strength of the synapse connection. A training dataset is typically used to incrementally change the weights whenever incorrect classifications occur to ultimately produce a model that has the best classification performance (Hastie et al. 2009). It is noteworthy that the greater the number of hidden layers and neurons used, the greater the risk of model overfitting (Aggarwal 2015). Overfitting refers to the model being able to efficiently replicate the training subset albeit with limited generalizability to other subsets. Therefore, it is highly recommended to start developing a neural network model with one hidden layer containing two or three neurons and gradually increase the network complexity only if the model performance is not satisfactory to the project goal.

3.3.2. MACHINE LEARNING MODEL SELECTION

Step 9 involves testing machine learning models, suitable to handle categorial features, employing the previously selected input features to predict the identified output feature for observations not used during the model training. A feature selection process is then applied in Step 10 to identify the most important input features. Feature selection is the process by which the number of input features is reduced, minimizing the redundant and noisy information produced by correlated and unnecessary inputs, respectively (Aggarwal 2015). Several feature selection approaches have been developed over the past few decades and are typically classified into filters and wrappers. Filter approaches require predefining an evaluation measure (e.g., correlation coefficient between the inputs and outputs) and subsequently eliminating the inputs based on a specified threshold. On the other hand, in wrapper approaches, a search technique is adopted to identify the most important input features based on the performance of the corresponding data-driven model. The developed framework employs a genetic algorithm wrapper technique, the application of which starts by randomly assuming a population of solutions that evolves continuously until an optimal solution is obtained (Goldberg 1989). Each solution (i.e., individual) within the population represents a single realization of the set of input features used within the machine learning model and is assigned a fitness value based on the model performance. New individuals are reproduced through special evolutionary operators including: 1) *elitism*, where individuals with greater fitness are replicated; 2) *crossover*, where sets of two individuals (i.e., parents) are selected based on a predefined criteria (e.g., random selection or a selection based on the fitness value) and subsequently mixed to produce new individuals; and 3) *mutation*, where single parents are altered randomly to produce new individuals (Nearchou 2004; Scrucca 2013; Yosri et al. 2021). Individuals are reproduced continuously until a predefined termination criterion is met, such as the number of generations, computational time, an acceptable fitness threshold, or a cost function that combines some or all of the aforementioned criteria.

In some cases, quantification of the overall performance of a machine learning model may not adequately reflect large errors if they occur in only small portions of the data set. Therefore, a cross-validation process should be applied to assess the generalizability of a model to an independent data set (Rohani et al. 2018). Multiple cross-validation methods have been developed to date (e.g., holdout, k-fold, and bootstrap) and each method has its own limitations, biases, and computational costs (Kohavi 1995). The current study uses a k-fold cross-validation during the feature selection process as described by Borra and Di Ciaccio (2010). In addition, the classification model's performance was assessed using annual datasets specific to each data-contributing organization (i.e., using multiple smaller subsets of the whole dataset).

3.3.3. MACHINE LEARNING MODEL PERFORMANCE METRICS

Step 11 of the Figure 3-2 framework specifies that the classification model performance should be evaluated, for example by using the following confusion matrix-based metrics:

$$Confusion\ Matrix = \begin{array}{cc|cc} & & \multicolumn{2}{c}{Predicted\ Class} \\ & & True & False \\ \hline \multirow{2}{*}{Actual\ Class} & True & TP & FN \\ & False & FP & TN \end{array} \tag{3-1}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3-2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3-3}$$

$$Recall = \frac{TP}{TP + FN} \tag{3-4}$$

where TP (True Positive) is the number of instances correctly classified as positive, TN (True Negative) is the number of instances correctly classified as negative, FP (False Positive) is the number of instances incorrectly classified as positive (i.e., Type I error), and FN (False Negative) is the number of instances incorrectly classified as negative (i.e., Type II error).

The *accuracy* is the ratio of correctly predicted instances to the total number of observations and is a valuable metric for consideration only when the dataset is symmetrical (i.e., there are an equal number of instances in each class). The *precision* is the ratio of correctly predicted positive instances to the total number of positive predicted instances. The *recall* is the ratio of correctly predicted positive instances to the total number of instances in its actual class. Although the accuracy, precision, and recall are all important to evaluating the classification model performance, the final classification model should be selected (Step 12) using the performance metrics most important to the utility stakeholder's project goal. For example, a model with the greatest recall value might be selected if the project goal is to minimize Type II error and a model with the greatest precision might be selected if the project goal were to minimize Type I error. Once the optimal machine learning model is selected, Step 13 includes analyzing the organization-specific classification model results using the performance metrics described through Equations 3-1 to 3-4.

3.3.4. RAPIDITY-CRITICAL FEATURE IMPORTANCE

Step 14 highlights the importance of understanding how the input features affect the classification of the output feature. This is especially important with ensemble models where the decision tree structure can no longer be visualized clearly (James et al. 2013). Feature importance analyses are typically employed to identify the features that can significantly improve the model performance (i.e., enhance the predictability of the output feature). Several measures can be used to express the feature importance (e.g., node impurity, mean decrease in accuracy), and a feature importance plot for such measures can be used to rank the input features in the order of their contribution to correctly classify the output. This study adopts the *mean decrease in accuracy* as a feature importance measure as this metric is based on the efficiency of predicting the out of bag samples when the selected input feature is excluded from the model (Hastie et al. 2009). The feature importance plot would rank the input features in order of their contribution for outage rapidity classification.

Step 15 of the Figure 3-2 framework outlines that to further understand the effect of input features, a partial dependence plot (PDP) is generated to examine the relative influence of input feature values to the output classification prediction (Molnar 2021). For classification models, the relative influence of a specified input feature value x, controlling for all the other input features, is found using the following equation:

$$f_k(x) = \log[p_k(x)] - \frac{1}{K}\sum_{j=1}^{K}\log\left[p_j(x)\right], \qquad j = 1,2,\dots,K \tag{3-5}$$

where $f_k(x)$ is the relative influence on the log probability values for the input feature *x* (Hastie et al. 2009), $p_k(x)$ and $p_j(x)$ are the probabilities of input feature *x* for classes *k* and *j*, respectively, and *K* is the number of classes in the output feature. It should be emphasized that is the $f_k(x)$ values are used to develop PDPs for classification models (Greenwell 2017). The PDP axis for the value of Equation 3-5 presents how the log-odds for class *k* (i.e., feature influence) depends on different subsets of the predictor features (Greenwell 2017). The PDP, in turn, indicates a positive, negative, or neutral feature influence on the correct classification of the output feature.

Following the development of PDPs, the final step of the Figure 3-2 framework is to identify resilience-guided asset managerial insights specific to the influence of feature values to the extended outages. The insights would present actionable items that could be performed by a utility to address the atypical performance of key feature values. Typically, identifying such insights would involve investigation into some aspects of the utility's asset management plan in order to minimize the extended outage key contributing features in the future (e.g., reviewing maintenance practices for a specific location on a transmission line or proposing a shorter timeframe between tree trimmings).

3.4. FRAMEWORK APPLICATION DEMONSTRATION

To demonstrate the utility of the developed framework shown in Figure 3-2, it was deployed using transmission line outages from the CEA database, as per Step 1 of the framework. Following Step 2 of the framework, 24 features are collected by the CEA for each transmission line outage as shown in Table 3-1. The study goal (i.e., to predict the rapidity of an outage soon following its occurrence) dictated that only 14 useful features (indicated by * in Table 3-1) could be employed as inputs to the machine learning classification model, as per Step 3 of the framework. Step 4 of the framework included extracting the dataset for the years 2005-2018 as the CEA included two additional descriptive features for each outage starting in 2005. However, these two additional features also created a challenge as they each included more than 53 separate sub-categorical levels. Some levels of the *primary cause name* and *subcomponent name* features were thus merged to reduce the complexity of the classification model and enable the development of a meaningful model using available packages in the *R* language (Liaw and Wiener 2002). The subcomponents of a transmission line component include external elements associated with the major component. For example, the subcomponents of a *wood pole* major component are a *conductor*, *structure*, *joints and dead-ends*, *insulation system*, *ground wire*, and *hardware*. These updates were primarily used to change a level name that included the word "other" at the end of the label to be added to the main heading (e.g., "Equipment Failure Other" was converted to "Equipment Failure"). In addition, only the sustained outages (i.e., outages with a duration greater than one minute) were included in the dataset as the focus of this demonstration is to predict forced outages with a measurable outage duration.

Table 3-1. The 24 features collected by the CEA for transmission line forced outages (* indicates selected as model input feature).

Feature	Type
Year	Numeric
Outage period	Binary
Contributor	Categorical
Event ID	Categorical
Common mode	Categorical
Common tower	Categorical
Major component code	Categorical
Voltage *	Categorical
Outage duration	Numeric
Outage start date	Time/Date
Primary cause group *	Categorical
Primary cause name *	Categorical
Subcomponent group *	Categorical
Subcomponent name *	Categorical
Type of failure *	Categorical
Type of fault *	Categorical
Component	Categorical
Conductors per phase *	Categorical
Ground wires *	Categorical
Structure *	Categorical
Circuit per tower *	Categorical
Month *	Categorical
Day *	Categorical
Hour *	Categorical

Step 5 of the framework included preliminary data analysis of the output feature. The statistical behavior of rapidity was investigated, as shown in Figure 3-3, in order to identify an appropriate rapidity *class split* (i.e., the time after which an outage is considered an extended event). Rapidity was found to exhibit a right-skewness which shows that the forced outages in transmission lines are mostly characterized by shorter durations. Accordingly, 480 minutes (i.e., eight hours) was chosen as the limit of extended outage events as approximately three-quarters of the durations are less than 480 minutes and one-quarter are longer. This limit was also chosen as it represents a typical full-workday for repair crews and any value longer than eight hours would limit the ability of a utility to assign appropriate resources within a typical shift. Following Step 6 of the framework, the output feature considered herein is the rapidity class (i.e., classified as either *1-480 minutes* or *>480 minutes*). The distribution between forced outage classes is shown in Figure 3-4.

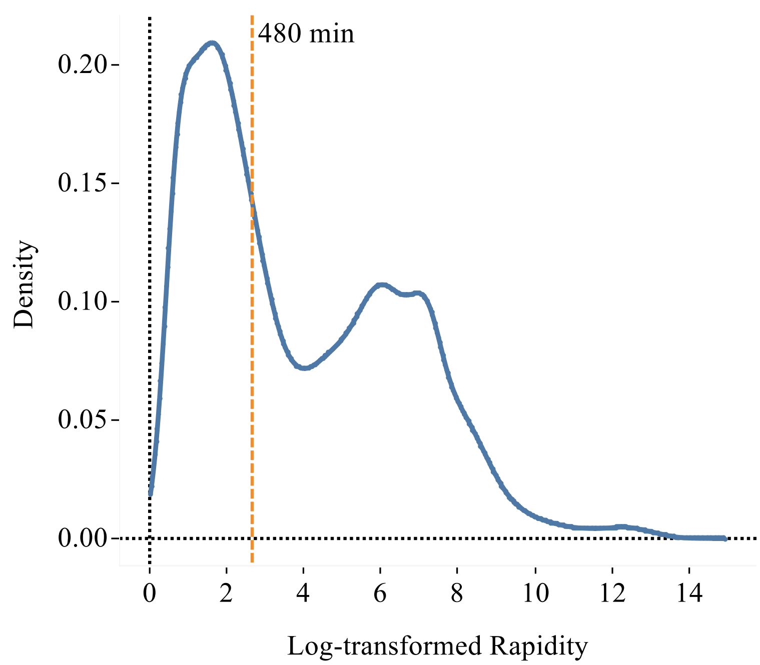


Figure 3-3. Density plot for log-transformed forced outages.

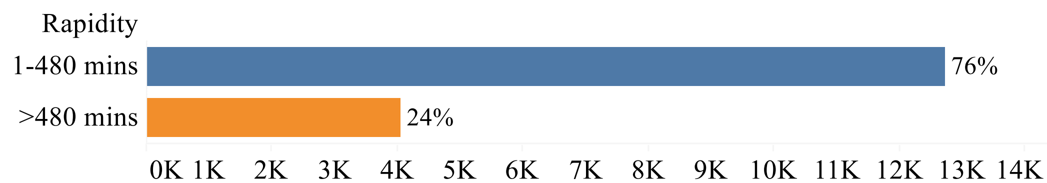


Figure 3-4. Population split of forced outages among all contributing organizations between rapidity classes from 2005-2018.

3.4.1. Classification Results

To follow Step 7 of the framework, the forced outage dataset from 2005 to 2018 was divided into training (from 2005 to 2013) and (from 2014 to 2018) testing subsets. Five of the seven utility organizations provided outage data for all the years from 2005-2018, therefore these five organization-specific datasets were used in the following model development. The training and testing subsets were created separately for those datasets, as displayed in Table 3-2. Consistent with Step 8 of the framework, the selected features of the CEA database are all categorical. As explained earlier, this situation presents a unique challenge limiting the machine learning classification models that could be deployed. The current demonstration application specifically employed bagging, random forests, and artificial neural networks as the machine learning classification techniques.

Table 3-2. Organization-specific training and testing data sets indicated as the number of forced outage instances in each set.

Organization	Training (2005-2013)	Testing (2014-2018)
Org 1	357 (59%)	253 (41%)
Org 2	2,521 (71%)	1,018 (29%)
Org 3	1,261 (67%)	612 (33%)
Org 4	697 (56%)	540 (44%)
Org 5	2,254 (65%)	1,236 (35%)

A schematic for the classification process is presented in Figure 3-5, where the inputs and output have been previously described in detail. The hyperparameters of a machine learning model represent critical components as their values control the learning process, model performance, and time to converge (Feurer and Hutter 2019). Selecting such values is typically carried out through an iterative process, either manually (Kuhn 2019) or using an optimization technique (Claesen and De Moor 2015). In this study, the hyperparameters of the bagging, random forest, and single hidden layer artificial neural network were adjusted manually until highly performing models were obtained (Kuhn 2019). Three machine learning models for classification were developed in the present study, as per Step 9 of the framework, including: 1) Model 1: bagging (500 trees and $m = 14$); 2) Model 2: random forest (500 trees and $m = 4$); and 3) Model 3: artificial neural network (200 iterations, 4 neurons, and 5×10^{-4} weight decay). A genetic algorithm-based feature selection was also embedded within the classification models, as per Step 10 of the framework, with the goal to maximize the recall; however, the model performance was not significantly improved by reducing the number of input features. During the feature selection process, a 5-fold cross-validation was incorporated to assess the result generalizability on independent data sets. Embedding the 5-fold cross validation within the feature selection process ensured that the most important features were obtained to generate the best performing and most generalizable classification model. This implied that all of the 14 features are important in distinguishing extended forced outages from shorter ones.

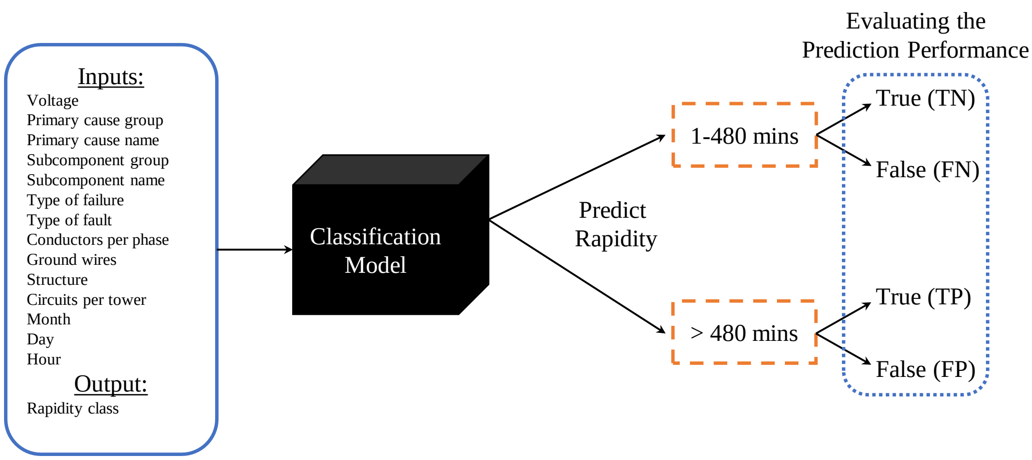


Figure 3-5. Classification model schematic showing implementation of framework previously described.

According to Step 11 of the framework, confusion matrix metrics (i.e., precision and recall) were compared considering the testing subset using Equations (3-2) and (3-3), as shown in Figure 3-6. The accuracy was not included in the comparison as the output feature is not symmetrical. It can clearly be seen from Figure 3-6 that Model 3 (i.e., artificial neural network) did not perform well in comparison to Models 1 and 2. The precision and recall values from Model 3 were consistently lower than Models 1 and 2, except in the case of Org 1, where all models classified the outages almost perfectly. The better performance of Models 1 and 2 compared to Model 3 might be attributed to: 1) the ensemble nature of the bagging and random forest approaches that enable combining multiple classifiers in order to produce a superior one (Polikar 2012); 2) the nonparametric nature of bagging and random forest approaches, which enhances their ability to capture the relationship between categorical features (Molnar 2021); and, 3) the typical procedures of developing artificial neural networks that includes the optimization of a set of numerical weights and biases, which is difficult to achieve for categorical inputs and outputs (Molnar 2021). There is less than 1% difference between the precision and recall values of Models 1 and 2. Therefore, to accomplish Step 12 of the framework, Model 2 was chosen as the preferred model, as a random forest model requires lower computational resources, which further reduced the variance, as the model is also more resistant to noise and outliers (Aggarwal 2015).

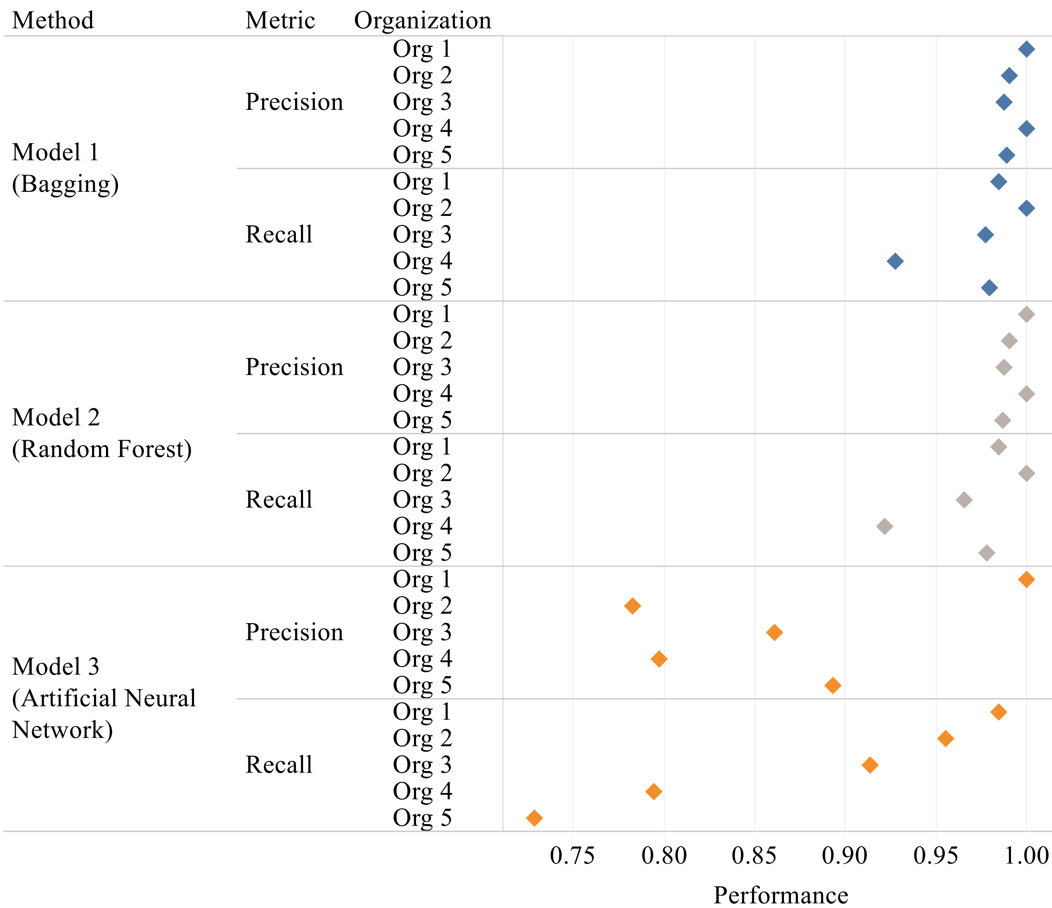


Figure 3-6. Average performance metrics for Models 1, 2, and 3 as implemented on the testing dataset for each of the five organizations.

Following Step 13 of the framework, Figure 3-7 was developed to better understand the annual classification performance (i.e., recall and precision) specific to each organization, viewing extended outages correctly predicted by Model 2 as true positives. Each performance metric (i.e., precision and recall) is important since outage misclassification would result in either excessive/unnecessary resources being sent to an outage location (i.e., precision) or inadequate resources/quick enough response provided (i.e., recall). Each tested year used the classification model trained based on the 2005-2013 outage data specific to each organization. The validity of Model 2 to correctly classify extended outages is demonstrated through the high precision and recall values for each organization-specific tested year, as shown in Figure 3-7. *Org 1's* classification results can be seen to be perfect from 2014-2017 with a reduction in recall in 2018 (7.7% misclassification), as shown in Figure 3-7. The model for *Org 2* had perfect recall from 2014-2018 whereas there was 1.4% and 3.4% precision error in 2016 and 2018, respectively, otherwise it yielded perfect performance. *Org 3* had the same precision and recall values from 2014-2017 indicating that there were the same number of misclassifications between rapidity classes in those years. *Org 4* had perfect precision for all tested years whereas the recall values were the lowest among all five organizations. This would suggest that within *Org 4*, the features recorded do not fully capture the differences between short and extended outages and additional years would thus be necessary to be evaluated as perfect classification occurred in 2018. *Org 5* had precision and recall values ranging from 95% to 100%, indicating that the selected features made it feasible to distinguish the two rapidity classes. These classification results prove the viability of applying the developed framework to classify future forced outages recorded in the CEA database soon after their occurrence. The accurate classification of forced outages would allow utilities to quickly respond with appropriate resources to return the transmission line component back to service rapidly.

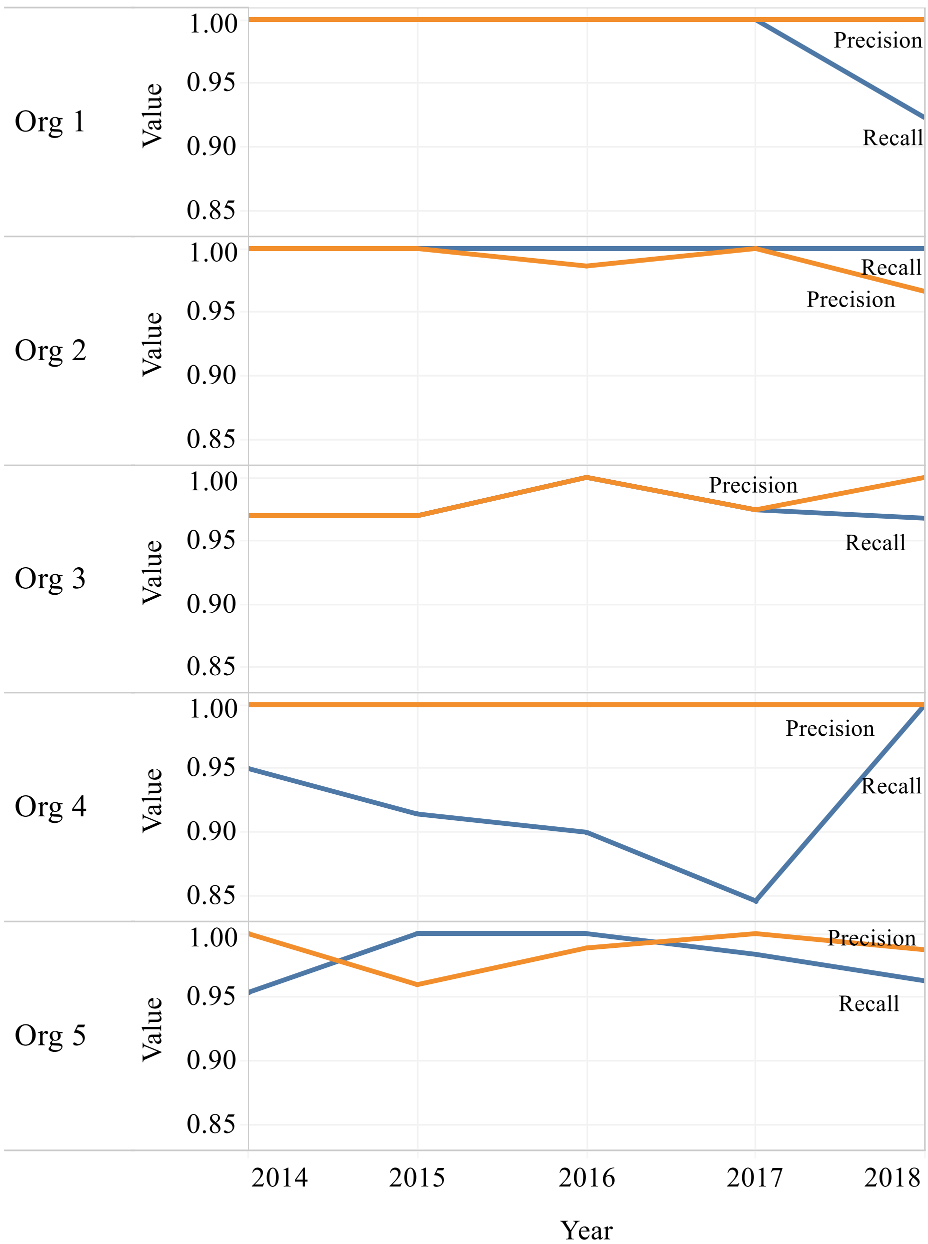


Figure 3-7. Annual Model 2 performance using testing dataset specific to each organization according to Table 3-2.

3.4.2. Feature Importance

Following Step 14 of the framework, Figure 3-8 ranks the input feature list in terms of importance, indicating the usefulness of features in distinguishing between the rapidity classes. The list uses the mean decrease in accuracy as the ranking metric as described previously. As indicated in Figure 3-8, the two most important features for all but *Org 2* are *primary cause name* and *subcomponent name*. *Org 2* has *primary cause group* as their second ranked input feature. These features, because of their importance ranking, can be viewed as critical to the correct classifications of the extended forced outages. To further investigate these features, PDPs were generated to view the importance of the feature values to the extended forced outage classification, as per Step 15 of the framework. PDPs are presented only for *Org 1* and *Org 4* for illustration, as *Org 1* had near perfect classification and *Org 4* had the greatest misclassification in terms of recall.

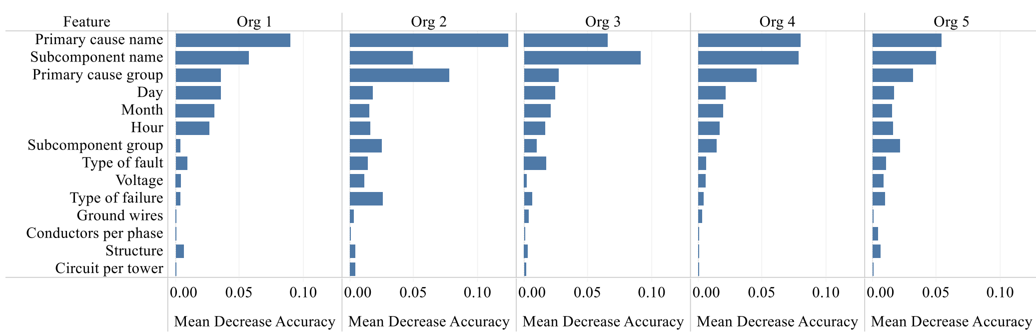


Figure 3-8. Input feature importance to extended forced outage classification for each organization.

Figure 3-9 shows the PDP investigating the feature values for *primary cause name* for *Org 1* and *Org 4*. The x-axis indicates the relative influence of the feature according to Equation 3-5 and the numerical label indicates the number of outages specific to that primary cause and organization within the dataset from 2005-2018. Positive relative influence values indicate the importance of that feature value to the classification of extended outages whereas negative values indicate the importance to the correct classification of short outages. Among feature values that occurred more than once per year, *Org 1*'s top two feature values contributing to extended outages were *contact by trees* and *deterioration due to age*. Similarly, *Org 4*'s top two feature values that occurred more than once per year were *deterioration due to age* and *equipment failure*, with each occurring more than five times per year, on average. The low numerical values corresponding to some higher-ranked features, indicate that the latter do not occur frequently; however, when they do, there is a greater likelihood of the outage being an extended outage.

Figure 3-9. *Primary cause name* feature PDP for extended outage classification for *Org's 1* and *4*.

Figure 3-10 shows the PDP for *Org 1* and *Org 4* to investigate the feature values for *subcomponent name*. *Org 1's* PDP indicates the critical features as the transmission line structure and hardware as subcomponents where extended outages commonly occur. *Org 4's* PDP indicates more than one extended outage per year, on average, occurred within the transmission line structure and in the ground wires. Figure 3-10 also shows that outages occurring in the insulation system are not often leading to extended outages, as indicated by the relative influence for both *Org's 1* and *4*. There are also several subcomponents where outages are not frequent; however, if they do occur, there is again a higher likelihood of them being an extended outage.

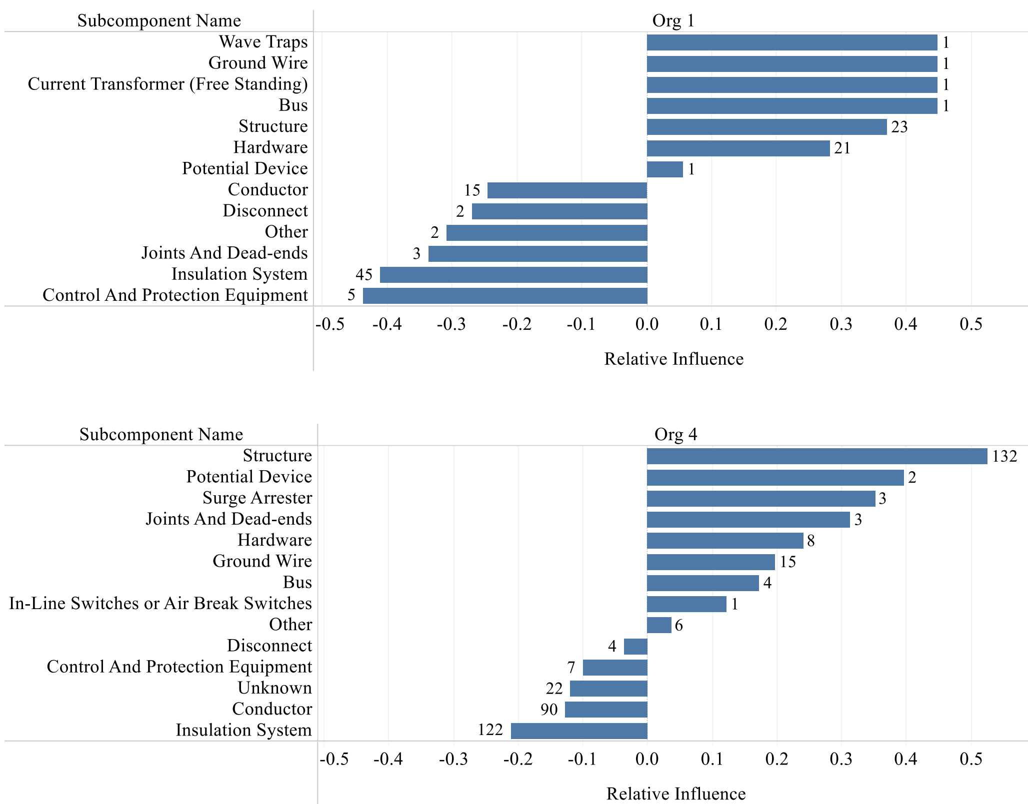


Figure 3-10. *Subcomponent name* feature PDP for extended outages classification for *Org's 1* and *4*.

3.5. INSIGHTS FOR RESILIENCE-GUIDED POWER INFRASTRUCTURE ASSET MANAGEMENT

The demonstration application goal was to show the utility of the classification framework in terms of accurately distinguishing extended and short outages, soon after forced outage occurrence. This was demonstrated/exhibited using five organization-specific datasets from the CEA database. The demonstration application illustrated that a utility company could quickly and accurately predict the rapidity following a forced outage, indicating the applicability of the proposed framework. In addition to the demonstration of the rapidity-classification model, critical features were identified as they strongly influenced the correct classification of the extended outages.

These extended outage-critical features yield valuable information that can enable utilities to plan their intervention and maintenance strategies as part of their asset management plan. For example, Figure 3-9 showed that primary causes of *equipment failure*, *deterioration due to age*, and *contact by trees* were critical to the specific utilities viewed in that figure. *Org 1* should thus review their tree trimming practices and make sure there are not lapses in the coverage or too long a period between trimmings, therefore mitigating the future occurrence of such extended outages. Alternatively, *Org 4* had equipment failure as a key primary cause of extended outages, indicating a need to review their maintenance practices on their transmission line assets. Both *Org 1* and *Org 4* had deterioration due to age as a key primary cause for extended outages, calling for a possible review of their intervention plan for their transmission line assets. Focus on the intervention and maintenance strategies for the transmission line structure subcomponents should occur as this was a key feature in the location of many extended outages for both reviewed organizations. These are some of the asset managerial insights that can be seen from the PDPs in Figure 3-9 and Figure 3-10.

The application of this framework was limited to the available databases that currently collect power infrastructure outages, but it is highly recommended to collect additional feature values for each forced outage to obtain even more managerial insights. This could include collecting the age of the component when the outage occurs, the date and type of the last maintenance event, the geographic location of the asset, and the future impacts of climate change. For example, important features from Figure 3-9 were *deterioration due to age* and *equipment failure* and having more detail on these features would be beneficial to utilities. Therefore, adding geographic features would be beneficial to the organizations that contribute the outage data as they span all over Canada or North America, depending on if using the CEA or NERC TADS databases, respectively. These additional features could provide valuable insights for utilities when they prepare asset management plans and reports for regulatory approval and evaluate the future impacts of climate change on their infrastructure assets.

3.6. CONCLUSION

Recently, there has been an increasing number of forced power outages, necessitating the need for accurate prediction of resilience key performance indicators to improve the ability of a utility to rapidly return to normal operation levels following forced outages. This study presented a framework to classify the rapidity metric of resilience following power infrastructure component forced outage occurrences. The utility of the framework was demonstrated using actual transmission line forced outages data. The rapidity was split into two classes, *1-480 mins* for short outages and *>480 mins* for extended outages. Specific datasets were then created for five outage data-contributing organizations and these datasets were subsequently split into training and testing sets for developing and validating the machine learning classification model, respectively. Bagging, random forest, and artificial neural network models were developed for each organization-specific dataset, with the random forest model having the best performance among all other models. The latter model was then deployed to classify forced outages and the results were displayed annually for the precision and recall metrics. It was found that the random forest model efficiently classified the forced outages for all five organizations, demonstrating the utility of the developed framework. Specific to each organization, the important features for rapidity classification were identified with *primary cause name* and *subcomponent name* being the most influential. These features were further explored using partial dependence analysis to investigate the feature value influence on extended outage prediction. Finally, asset managerial insights were presented based on the influential feature values. The demonstration application illustrated the capability of the described framework in classifying forced outages soon after their occurrence, allowing utilities and other power infrastructure stakeholders to proactively prepare and respond rapidly to outage risks. The application of this framework was limited to the available data currently collected by existing databases, but the authors recommend that additional features related to the resilience metric of resourcefulness (e.g., staff and spare components available) be collected and incorporated into the prediction model to provide a better representation of an organizations resilient response to forced outages.

3.7. ACKNOWLEDGEMENTS

The financial support for the study was provided through the Canadian Nuclear Energy Infrastructure Resilience under Systemic Risks (CaNRisk) – Collaborative Research and Training Experience (CREATE) program of the Natural Science and Engineering Research Council (NSERC) of Canada. The support of the INTERFACE Institute and the INViSiONLab is also acknowledged in the development of this study. The authors are grateful to the Canadian Electricity Association (CEA) for providing the transmission line data used for the framework demonstration application.

3.8. NOTATIONS

t	Number of bootstrapped classifiers
m	Number of features employed at a decision tree split
q	Number of features in the dataset
$f_k(x)$	Relative influence on the log probability values for the input feature x for class k

x	Input feature
$p_k(x)$	Probability of input feature *x* in class *k*
$p_j(x)$	Probability of input feature *x* in class *j*
K	Number of classes in the output feature

3.9. ACRONYMS

CEA	Canadian Electricity Association
FN	False negative
FP	False positive
NERC	North American Electric Reliability Corporation
PDP	Partial dependence plot
SVM	Support vector machine
TADS	Transmission Availability Data System
TN:	True negative
TP:	True positive

3.10. REFERENCES

Abdelfatah, M., El-Shimy, M., and Ismail, H. M. (2013). "Outage data analysis of utility power transformers based on outage reports during 2002-2009." International Journal of Electrical Power and Energy Systems, Elsevier, 47(1), 41–51.

Aggarwal, C. C. (2015). Data Mining: The textbook. Springer International Publishing, New York.

Barker, K., Lambert, J. H., Zobel, C. W., Tapia, A. H., Ramirez-Marquez, J. E., Albert, L., Nicholson, C. D., and Caragea, C. (2017). "Defining resilience analytics for interdependent cyber-physical-social networks." Sustainable and Resilient Infrastructure, Taylor & Francis, 2(2), 59–67.

Bhusal, N., Abdelmalak, M., Kamruzzaman, M., and Benidris, M. (2020). "Power system resilience: Current practices, challenges, and future directions." IEEE Access, 8, 18064–18086.

Bian, J. J., Ekisheva, S., and Slone, A. (2014). "Top risks to transmission outages." IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Bie, Z., Lin, Y., Li, G., and Li, F. (2017). "Battling the Extreme: A Study on the Power System Resilience." Proceedings of the IEEE, 105(7), 1253–1266.

Black, J., Hoffman, A., Hong, T., Roberts, J., and Wang, P. (2018). "Weather Data for Energy Analytics: From Modeling Outages and Reliability Indices to Simulating Distributed Photovoltaic Fleets." IEEE Power and Energy Magazine, 16(3), 43–53.

Borra, S., and Di Ciaccio, A. (2010). "Measuring the Prediction Error. A Comparison of Cross-Validation, Bootstrap and Covariance Penalty Methods." Comput. Stat. Data Anal., Elsevier Science Publishers B. V., NLD, 54(12), 2976–2989.

Bruneau, M., Chang, S. E., Eguchi, R. T., Lee, G. C., O'Rourke, T. D., Reinhorn, A. M., Shinozuka, M., Tierney, K., Wallace, W. A., and Von Winterfeldt, D. (2003). "A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities." Earthquake Spectra, 19(4), 733–752.

Canadian Electricity Association. (2016). Adapting to climate change: state of play and recommendations for the electricity sector in Canada. Ottawa.

CEA. (2018). Instruction Manual for Reporting Component Forced Outages for Transmission Equipment. Ottawa, ON.

Chi, S., Suk, S. J., Kang, Y., and Mulva, S. P. (2012). "Development of a data mining-based analysis framework for multi-attribute construction project information." Advanced Engineering Informatics, Elsevier Ltd, 26(3), 574–581.

Ciapessoni, E., Cirio, D., Pitto, A., Panteli, M., Van Harte, M., and Mak, C. (2019). "Defining Power System Resilience." Electra, October(306), 32–34.

Claesen, M., and De Moor, B. (2015). "Hyperparameter Search in Machine Learning." MIC 2015: The XI Metaheuristics International Conference, arXiv, Agadir, 14-1-14–5.

Davila Delgado, J. M., Oyedele, L., Bilal, M., Ajayi, A., Akanbi, L., and Akinade, O. (2020). "Big Data Analytics System for Costing Power Transmission Projects." Journal of Construction Engineering and Management, 146(1), 1–10.

Delen, D., and Ram, S. (2018). "Research challenges and opportunities in business analytics." Journal of Business Analytics, Taylor & Francis, 1(1), 2–12.

Dokic, T., Pavlovski, M., Gligorijevic, D., Kezunovic, M., and Obradovic, Z. (2019). "Spatially Aware Ensemble-Based Learning to Predict Weather-Related Outages in Transmission." Proceedings of the 52nd Hawaii International Conference on System Sciences, Hawaii International Conference on System Sciences, 1–10.

Ekisheva, S., and Gugel, H. (2015a). "North American transformer outage rates and durations in assessment of transmission system reliability and availability." IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Ekisheva, S., and Gugel, H. (2015b). "North American AC circuit outage rates and durations in assessment of transmission system reliability and availability." IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Ekisheva, S., Lauby, M. G., and Gugel, H. (2016). "North American transformer outages initiated by transmission equipment failures and human error." IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Ekisheva, S., Papic, M., Clemons, M. S., Quest, R., Pakeltis, M. J., and Weisman, K. (2018). "Outage statistics, reliability and availability of DC circuits in North American Bulk Power System." IEEE Power and Energy Society General Meeting, IEEE Computer Society, 1–5.

Eskandarpour, R., Khodaei, A., and Arab, A. (2017). "Improving power grid resilience through predictive outage estimation." 2017 North American Power Symposium (NAPS), IEEE, Morgantown, WV, USA, 1–5.

Feurer, M., and Hutter, F. (2019). "Hyperparameter Optimization." Automated Machine Learning, Springer, Cham, 3–33.

Fujikawa, Y., and Ho, T. (2002). "Cluster-based Algorithms for Filling Missing Values." Pacific-Asia conference on knowledge discovery and data mining, Springer, Berlin, Heidelberg, 549–554.

Gholami, A., Shekari, T., Amirioun, M. H., Aminifar, F., Amini, M. H., and Sargolzaei, A. (2018). "Toward a consensus on the definition and taxonomy of power system resilience." IEEE Access, IEEE, 6, 32035–32053.

Goldberg, D. E. (1989). Genetic Algorithms in Search, Optimization and Machine Learning. Optimization, and Machine Learning, Addison-Wesley Longman Publishing Co. Inc., Boston, MA.

Gondia, A., Siam, A., El-Dakhakhni, W., and Nassar, A. H. (2020). "Machine Learning Algorithms for Construction Projects Delay Risk Prediction." Journal of Construction Engineering and Management, 146(1), 04019085.

Goyal, A., Aprilia, E., Janssen, G., Kim, Y., Kumar, T., Mueller, R., Phan, D., Raman, A., Schuddebeurs, J., Xiong, J., and Zhang, R. (2016). "Asset health management using predictive and prescriptive analytics for the electric power grid." IBM Journal of Research and Development, 60(1.4), 1–14.

Greenwell, B. M. (2017). "pdp: An R package for constructing partial dependence plots." R Journal, 9(1), 421–436.

Haggag, M., Ezzeldin, M., El-Dakhakhni, W., and Hassini, E. (2020). "Resilient cities critical infrastructure interdependence: a meta-research." Sustainable and Resilient Infrastructure, Taylor & Francis, 1–22.

Haggag, M., Siam, A. S., El-Dakhakhni, W., Coulibaly, P., and Hassini, E. (2021a). "A deep learning model for predicting climate-induced disasters." Natural Hazards, Springer Netherlands, 107(1), 1009–1034.

Haggag, M., Yorsi, A., El-Dakhakhni, W., and Hassini, E. (2021b). "Infrastructure performance prediction under Climate-Induced Disasters using data analytics." International Journal of Disaster Risk Reduction, Elsevier Ltd., 56 (November 2020), 102121.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning. Springer Series in Statistics, Springer New York, New York, NY.

Idaho National Laboratory. (2010). HTGR Resilient Control System Strategy. Idaho Falls.

James, G., Witten, D., Tibshirani, R., and Hastie, T. (2013). An Introduction to Statistical Learning with Applications in R. Springer Texts, Springer Science+Business Media, New York, NY.

Jufri, F. H., Widiputra, V., and Jung, J. (2019). "State-of-the-art review on power grid resilience to extreme weather events: Definitions, frameworks, quantitative assessment methodologies, and enhancement strategies." Applied Energy, 239(January), 1049–1065.

Kohavi, R. (1995). "A study of cross-validation and bootstrap for accuracy estimation and model selection." IJCAI-95, C.S. Mellish, ed., California: Morgan Kaufmann, Montréal, Quebec, 1137–1143.

Kuhn, M. (2019). The caret Package.

Li, D., Deogun, J., Spaulding, W., and Shuart, B. (2005). "Dealing with Missing Data: Algorithms Based on Fuzzy Set and Rough Set Theories." Transactions on Rough Sets IV. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 37–57.

Liaw, A., and Wiener, M. (2002). "Classification and Regression by randomForest." R News, 2(3), 18–22.

Molnar, C. (2021). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. Creative Commons License, Online.

Mukherjee, S., Nateghi, R., and Hastak, M. (2018). "A multi-hazard approach to assess severe weather-induced major power outage risks in the U.S." Reliability Engineering and System Safety, Elsevier Ltd, 175, 283–305.

Nateghi, R. (2018). "Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems." IEEE Access, IEEE, 6, 13478–13489.

Nearchou, A. C. (2004). "The effect of various operators on the genetic search for large scheduling problems." International Journal of Production Economics, 88, 191–203.

North American Electric Reliability Corporation. (2007). Transmission Availability Data System Final Report. Princeton, New Jersey.

Omran, S., and El Houby, E. M. F. (2020). "Prediction of electrical power disturbances using machine learning techniques." Journal of Ambient Intelligence and Humanized Computing, Springer Berlin Heidelberg, 11(7), 2987–3003.

Panteli, M., and Mancarella, P. (2015a). "Influence of extreme weather and climate change on the resilience of power systems: Impacts and possible mitigation strategies." Electric Power Systems Research, Elsevier B.V., 127, 259–270.

Panteli, M., and Mancarella, P. (2015b). "The Grid: Stronger, Bigger, Smarter?" IEEE Power and Energy Magazine, IEEE, 13(May/June), 58–66.

Papic, M., Agarwal, S., Allan, R. N., Billinton, R., Dent, C. J., Ekisheva, S., Gent, D., Jiang, K., Li, W., Mitra, J., Pitto, A., Schneider, A., Singh, C., Vadlamudi, V. V., and Varghese, M. (2017). "Research on Common-Mode and Dependent (CMD) Outage Events in Power Systems: A Review." IEEE Transactions on Power Systems, IEEE, 32(2), 1528–1536.

Papic, M., Agarwal, S., Bian, J., Billinton, R., Dent, C., Dobson, I., Jirutitijaroen, P., Li, W., Menten, T., Mitra, J., Schneider, A., Singh, C., and Vadlamudi, V. V. (2014). "Effects of dependent and common mode outages on the reliability of bulk electric system - Part II: Outage data analysis." IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Papic, M., Clemons, M., Ekisheva, S., Gent, D., Norris, J., Pakeltis, M., Peacock, M., and Till, D. (2018). "Performance analysis of North American AC circuits on common structures using TADS and CEA outage statistics." 2018 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), IEEE, 1–6.

Papic, M., Clemons, M., Ekisheva, S., Langthorn, J., Ly, T., Pakeltis, M., Quest, R., Schaller, J., Till, D., and Weisman, K. (2016). "Transmission Availability Data System (TADS) reporting and data analysis." 2016 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), IEEE, Beijing, 1–7.

Patil, B. M., Joshi, R. C., and Toshniwal, D. (2010). "Missing value imputation based on k-mean clustering with weighted distance." Communications in Computer and Information Science, Springer-Verlag, 600–609.

Polikar, R. (2012). "Ensemble Learning." Ensemble Machine Learning, C. Zhang and Y. Ma, eds., Springer Science+Business Media, New York, 1–34.

Preston, B. L., Backhaus, S. N., Ewers, M., Phillips, J. A., Dagle, J. E., Tarditi, A. G., Looney, J. (Pat), and King, T. J. J. (2016). Resilience of the U.S. Electricity System: A Multi-Hazard Perspective. Springfield, VA.

Raoufi, H., Vahidinasab, V., and Mehran, K. (2020). “Power Systems Resilience Metrics: A Comprehensive Review of Challenges and Outlook.” Sustainability, 12(22), 1–24.

Rohani, A., Taki, M., and Abdollahpour, M. (2018). “A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I).” Renewable Energy, 115(C), 411–422.

Salem, S., Siam, A., El-Dakhakhni, W., and Tait, M. (2020). “Probabilistic resilience-guided infrastructure risk management.” Journal of Management in Engineering, 36(6), 1–15.

Schaller, J. (2012). “Common mode event perspectives from the Canadian electricity association equipment reliability information system.” IEEE Power and Energy Society General Meeting, IEEE, 1–6.

Schaller, J., and Ekisheva, S. (2016). “Leading causes of outages for transmission elements of the North American bulk power system.” IEEE Power and Energy Society General Meeting, IEEE, 1–5.

Scrucca, L. (2013). “GA: A package for genetic algorithms in R.” Journal of Statistical Software, 53(4), 1–37.

Xie, J., Alvarez-Fernandez, I., and Sun, W. (2020). "A Review of Machine Learning Applications in Power System Resilience." 2020 IEEE Power & Energy Society General Meeting (PESGM), IEEE, 1–5.

Yosri, A., Elleathy, Y., Hassini, S., and El-Dakhakhni, W. (2021). "Genetic Algorithm-Markovian Model for Predictive Bridge Asset Management." Journal of Bridge Engineering, 26(8), 04021052.

Zhang, P., Xu, Y., Luo, F., and Dong, Z. Y. (2019). "Power Big Data: New Assets of Electric Power Utilities." Journal of Energy Engineering, 145(3), 04019009.

Zhou, K., Fu, C., and Yang, S. (2016). "Big data driven smart energy management: From big data to big insights." Renewable and Sustainable Energy Reviews, Elsevier Ltd, 56, 215–225.

Zumel, N., and Mount, J. (2020). Practical Data Science with R. Manning Publications Co., Shelter Island.

Chapter 4

INFRASTRUCTURE ASSET MANAGEMENT SYSTEM OPTIMIZED CONFIGURATION: A GENETIC ALGORITHM-COMPLEX NETWORK THEORETIC APPROACH

ABSTRACT

An effective infrastructure asset management (AM) system is crucial for utilities, city managers, governments, and other asset-owning organizations to navigate the numerous challenges associated with operating and managing infrastructure assets. In this paper, the AM system is represented as a network (comprised of nodes and links) that describes the major components necessary for operating an AM system within an organization and the information connections between such components. The ISO 55001 standard specifies the requirements for an AM system and outlines the criticality levels of different AM system components—reflected in the AM system network by the link weights. The main challenges of operating and managing an AM system pertain to information asymmetry between AM system components (e.g., not using the correct information for decision-making) and information overload within AM system components (i.e., too much information flow undermining stakeholders' ability to identify the correct information needed for decision-making). These challenges cause systemic risks (possibility of dependence-induced disruptions) within the AM system network due to the connectedness of AM system components. Systemic risks can be mitigated through

built-in network resilience by adding AM system connections. This network re-configuration is a complex problem with multiple potential solutions depending on the number of connections to be added to the AM system. For this reason, this study employs a genetic algorithm approach to solve for the optimal network configurations considering multiple objective functions based on average centrality values (i.e., betweenness-, closeness-, eigenvector-centrality, vulnerability index, and a weighted combination of the four centralities) for the most critical 15 AM subject areas. The considered objective functions are evaluated for only one to fifteen link additions to limit AM network over-connectedness which could otherwise lead to information overload. Managerial insights are outlined to explain how the optimization results can be deployed to mitigate the systemic risks within an organization's AM system based on the different objective function evaluations.

KEYWORDS: centrality analysis, complex network theory, genetic algorithm, infrastructure asset management, ISO 55001, systemic risk, vulnerability index.

4.1. INTRODUCTION

Organizations that own and manage infrastructure assets face many challenges in operating their businesses, including infrastructure ageing, evolving (usually more strict) regulatory operating conditions, limited renewal financial resources, and losing valuable experience through retirements (Canadian Construction Association et al. 2016; American Society of Civil Engineers 2021). A well-defined asset management (AM) program can offset and minimize the impacts of these organizational challenges (Bertling Tjernberg 2018; Infrastructure Canada 2018; Canadian Infrastructure Report Card 2019; American Society of Civil Engineers 2021). AM is the structured decision-making and execution of plans developed to achieve a balance between asset performance and risk through the optimal allocation of available resources and the procurement of additional resources (Uddin et al. 2013; Ross 2019). Organizations that own, manage, and operate infrastructure assets typically implement AM systems to achieve their organizational strategic plan and objectives (Hodkiewicz 2015). An AM system is a set of interrelated and interacting elements of an organization, whose function is to establish the AM policy and AM objectives, and the processes needed to achieve those objectives (International Organization for Standardization 2014a).

As outlined by the Institute of Asset Management's (IAM) *Asset Management Anatomy* document, a typical AM system is composed of six divisions and 39 subject areas (Institute for Asset Management 2015), as shown in Figure

4-1. Each of the 39 AM subject areas was designed to illustrate the breadth of a certain set of activities within an AM system, the relationships between these activities and the need to integrate them, and the critical role for AM to align with and deliver the strategic plan of an asset-intensive organization (Institute for Asset Management 2015). As such, each AM subject area is a functional component necessary for the implementation and operation of an AM system within an organization that owns and manages infrastructure assets. For a detailed description of each AM subject area, the authors refer to the *Asset Management Anatomy* guide developed by the IAM (Institute for Asset Management 2015). Goforth et al. (2021) introduced the concept explaining how a typical AM system can be viewed as a network of connected components with nodes as AM subject areas from the IAM (2015) and links as information flow between/connecting nodes as defined by the Global Forum on Maintenance and Asset Management (2014).

Figure 4-1. The AM divisions and subject areas as defined by the Institute for Asset Management (2015).

The ISO 55001 standard *Asset management—Management systems—Requirements* specifies requirements for the AM system within the context of an organization (International Organization for Standardization 2014b). This standard was designed to be applied to all asset types and by all types and sizes of organizations (International Organization for Standardization 2014b). ISO 55001 groups the main requirements of an AM system according to the context of the organization, leadership, planning, support, operation, performance evaluation, and improvement (International Organization for Standardization 2014b). There are currently only 27 ISO 55001 certified organizations in North and South America indicating that they meet or exceed the specifications for an AM system as outlined in the standard (International Organization for Standardization 2021).

Organizations use the principles and guidelines described within the ISO 55001 standard to improve their AM system operations (Woodhouse 2014; Hodkiewicz 2015; Konstantakos et al. 2019).

An efficient AM system must include strong information flow between the connected subject areas as miscommunication can lead to system dysfunction. In addition, an effective AM system minimizes the exposure of AM stakeholders to information overload caused by having access to too much information and data (Herrera et al. 2011; Prajogo et al. 2018). An AM system may be influenced by systemic risks (i.e., dependence-induced disruptions) specifically related to information asymmetry caused by the malfunction of one or multiple specific subject areas or the interruption of information flow between multiple subject areas (Goforth et al. 2022). Systemic risks might also be induced by information overload within an AM subject area where there is too much information flow, and an AM stakeholder is overwhelmed and unable to make decisions (Goforth et al. 2022). Examples of information asymmetry in an AM system include stakeholders that use different information to support their AM subject area-specific decision process, not responding to other stakeholders' decisions promptly, and the isolation of AM subject areas due to inadequate information-sharing procedures or protocols (Bergh et al. 2019). These examples of information asymmetry and information overload are critical challenges for the effective operation of an AM system within an organization and can potentially initiate cascading disruptions throughout an AM system (Brunetto et al. 2014; Xerri et al. 2015; Pell et al. 2015; de la Pena et al.

2016; Golightly et al. 2018). However, such challenges and their associated systemic risks can be mitigated by increasing the network resilience to dependence-induced disruptions through re-configuration of the connections between the different AM subject areas (Barabási 2016; Goforth et al. 2022).

Improving the resilience of a network to dependence-induced disruptions (i.e., systemic risks) is a complex problem that requires the use of heuristic-based optimization techniques as there are multiple solutions depending on the number of links to be added to the network, the length of those links, and the targeted reduction in systemic risk levels (Barbosa et al. 2018; Bhavathrathan and Patil 2018; Nozhati et al. 2019; Morshedlou et al. 2021; Vishnu et al. 2021). Several studies have been conducted in other fields to enhance a system network's resilience through adding new connections between its different components (Parotsidis et al. 2015; Papagelis 2015; Crescenzi et al. 2016; Ohara et al. 2017). While introducing additional connections within a system may provide faster information transfer (Parotsidis et al. 2016; Medvet and Bartoli 2021), determining an optimal number and configuration of added connections is an onerous task. The connection addition process has thus been formulated as an optimization problem, where previous studies have deployed greedy algorithms (Parotsidis et al. 2015; Crescenzi et al. 2016; Parotsidis et al. 2016; Ohara et al. 2017), path screening techniques (Papagelis 2015), and a heuristic-based genetic algorithm (GA) (Zhao et al. 2018; Pizzuti and Socievole 2018; Paterson and Ombuki-Berman 2020; Medvet and Bartoli 2021), albeit for different system applications (e.g., transportation and social networks) and not for optimizing the AM system to reduce its vulnerability to systemic risks.

Therefore, the objective of the present study is to identify optimal link configurations to be added to the AM system network, such that the expected systemic risks due to information asymmetry and information overload would be significantly mitigated. This study first explains the structure of the AM network and its relation to ISO 55001. Next, a description of the network centrality measures used to determine the most critical AM subject areas in the AM system network is presented. Subsequently, the application procedure of the GA is described for different centrality measure-based objective functions. Results are then presented in terms of optimal link configurations to satisfy the different objective functions and the practical implications of the link additions to a real AM system. Finally, managerial insights are presented to allow for the comparison between the different link configuration scenarios according to the different centrality measures investigated.

4.2. THE ASSET MANAGEMENT NETWORK STRUCTURE

Goforth et al. (2021) conceptualized the AM system as a network of connected nodes to facilitate its analysis. Building on this methodology, a typical AM system, proposed by the IAM, is represented in this study by a weighted, directed network as shown in Figure 4-2. This AM system network consists of 39 nodes, each representing a specific AM subject area. Node labels in Figure 4-2 correspond to the AM subject areas shown in Figure 4-1, where the node color indicates the AM division. This study specifies the criticality of each AM network subject area as the number of ISO 55001 clauses that are related to it as defined by the Institute for Asset Management (2015). These relationships are shown in Figure 4-3, with the total number of connected ISO 55001 clauses (i.e., the AM subject area weight) in the bottom row.

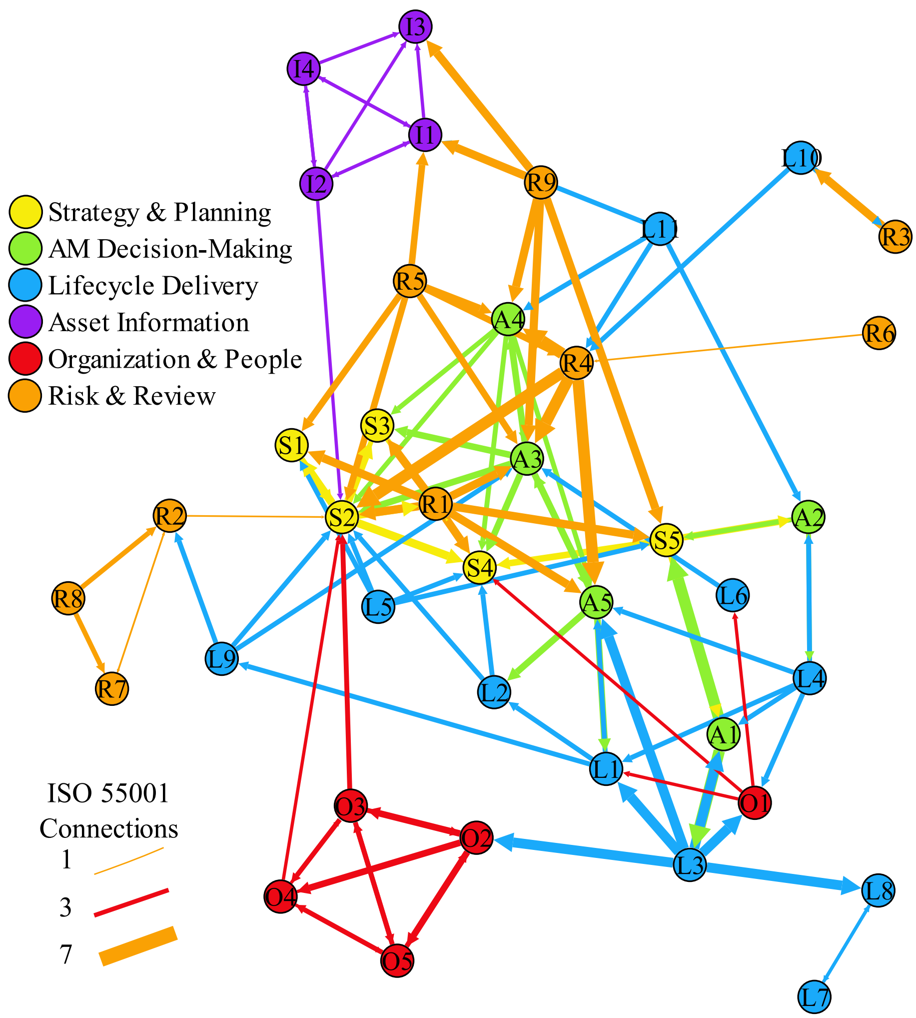


Figure 4-2. The AM network structure based on the AM subject areas presented in Figure 4-1.

ISO 55001 Clause \ AM Subject Area	S1 - Asset Management Policy	S2 - Asset Management Strategy & Objectives	S3 - Demand Analysis	S4 - Strategic Planning	S5 - Asset Management Planning	A1 - Resourcing Strategy	A2 - Shutdown & Outage Strategy	A3 - Capital Investment Decision-Making	A4 - Lifecycle Value Realization	A5 - Operations & Maintenance Decision-Making	L1 - Maintenance Delivery	L2 - Asset Operations	L3 - Resource Management	L4 - Shutdown & Outage Management	L5 - Technical Standards & Legislation	L6 - Asset Creation & Acquisition	L7 - Systems Engineering	L8 - Configuration Management	L9 - Reliability Engineering	L10 - Fault & Incident Response	L11 - Asset Decommissioning & Disposal	I1 - Data & Information Management	I2 - Asset Information Strategy	I3 - Asset Information Systems	I4 - Asset Information Standards	O1 - Procurement & Supply Chain Management	O2 - Competence Management	O3 - Asset Management Leadership	O4 - Organizational Structure	O5 - Organizational Culture	R1 - Stakeholder Engagement	R2 - Asset Performance & Health Monitoring	R3 - Contingency Planning & Resilience Analysis	R4 - Risk Assessment & Management	R5 - Sustainable Development	R6 - Management of Change	R7 - Asset Management System Monitoring	R8 - Management Review, Audit, & Assurance	R9 - Asset Costing & Valuation
4.1 Understanding the organization and its context		X	X																										X	X					X				
4.2 Understanding the needs and expectations of stakeholders								X		X																					X				X				X
4.3 Determining the scope of the AM system																																							
4.4 AM system		X																																					
5.1 Leadership and commitment	X																											X		X				X					
5.2 Policy	X																														X								
5.3 Organizational roles, responsibilities, and authorities		X				X							X														X	X	X										
6.1 Actions to address risks and opportunities for the AM system			X		X	X	X	X	X	X																							X	X					
6.2.1 AM objectives		X				X	X	X	X	X																					X				X				X
6.2.2 Planning to achieve AM objectives			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X												X	X	X				X
7.1 Resources						X							X													X													
7.2 Competence						X							X														X												
7.3 Awareness	X				X																						X												
7.4 Communication																												X		X	X								
7.5 Information requirements															X			X				X	X		X						X								X
7.6 Documented information															X							X	X	X	X														
8.1 Operational planning and control											X	X	X	X		X	X			X	X												X	X					
8.2 Management of change											X	X	X	X		X				X	X													X		X			
8.3 Outsourcing						X																				X	X							X					
9.1 Monitoring, measurement, analysis, and evaluation																																X	X	X			X		X
9.2 Internal audit																																						X	
9.3 Management review																																						X	
10.1 Nonconformity and correction action																			X	X																			
10.2 Preventative action																			X														X						
10.3 Continual improvement	X	X			X																																	X	
Total ISO 55001 Clause Connections	4	5	3	1	4	7	3	4	3	4	3	3	6	3	3	3	2	2	3	3	3	2	2	1	2	2	4	3	2	3	5	1	5	7	4	1	1	3	5

Figure 4-3. ISO 55001 clauses mapped to AM subject areas as defined by the Institute for Asset Management (2015).

Links between node pairs are defined according to the Global Forum on Maintenance and Asset Management (2014) that describe each AM subject area and identifies its connections to others. Such connections are essential to provide information necessary to make decisions within each AM subject area and thus the overall AM system. Links in the AM network are directed and weighted, as information transfer occurs from the source to target nodes, where the link weight is defined as the source node weight. In Figure 4-2, link direction is represented by an arrowhead at the target node, and link thickness reflects the corresponding weight, where thicker links indicate a greater ISO 55001 influence and therefore greater criticality to the overall AM system operation. The AM network shown in Figure 4-2 is represented mathematically through an adjacency matrix $\boldsymbol{A}$, which can be found in Appendix S1, with entries $a_{i,j}$ that reflects the link weight and direction between nodes i and j.

4.3. CENTRALITY MEASURES

The *betweenness centrality* is a metric that identifies the criticality of a specific node as the fraction of shortest paths passing through it (Freeman 1977). The betweenness centrality of a node i (BC_i) is thus calculated as:

$$BC_i = \sum_{j \neq i \neq k} \frac{\rho_{jk}(i)}{\rho_{jk}} \tag{4-1}$$

where ρ_{jk} is the weighted length of all shortest paths connecting nodes j and k and $\rho_{jk}(i)$ is the weighted length of these shortest paths that traverse node i. Regarding the AM network, the BC_i values reflect the importance of the corresponding AM subject area to the operation of the whole AM system and can thus be used to reflect its influence on the information flow throughout the AM system.

The *closeness centrality* is a metric that identifies nodes that are key to rapidly process and relay information to other nodes in the network (Estrada and Knight 2015). The closeness centrality of a node i (CC_i) is evaluated as:

$$CC_i = \frac{\sum_j d(i,j)}{N} \tag{4-2}$$

where $d(i,j)$ is the shortest path length between nodes i and j and N is the total number of nodes in the network. When applied to the AM network, the closeness centrality can be used to identify the AM subject areas that are critical for the rapid processing and transferring of information to other AM subject areas.

The *eigenvector centrality* is a metric that identifies nodes that are highly connected to influential nodes within the network (Thai and Pardalos 2012). The eigenvector centrality of a node i (EC_i) is:

$$EC_i = \frac{1}{\lambda} \sum_j a_{i,j} EC_j \tag{4-3}$$

where λ is the maximum eigenvalue of the adjacency matrix $\boldsymbol{A}$. For the AM network, a greater value of EC_i indicates that the corresponding AM subject area is connected to other highly influential AM subject areas and is thus critical to the information transfer within the AM system.

The *vulnerability index* is related to the size of the network's giant component (i.e., the largest connected set of nodes), and is used to identify the critical nodes that are highly sensitive to disruptions by quantifying the fraction of non-operational nodes when a specific node is triggered to fail (Ezzeldin and El-Dakhakhni 2019). The vulnerability index of node i (VI_i) is given by:

$$VI_i = \frac{N - N'}{N} \tag{4-4}$$

where N' is the total number of operational nodes in the network's giant component after node i was triggered to fail. It should be noted that the value of VI_i is estimated considering the cascading disruption effect within the network as follows: 1) once a node i is triggered to fail, information moving through the network is redistributed based on the shortest paths available; 2) following the redistribution, nodes are considered operational when they can sustain their original information share in addition to those transferred from other nodes; and 3) the redistribution process continues until all nodes in the network are functional and the corresponding N' value is then obtained. Therefore, calculating the values of VI_i necessitates adopting the concept of overflow modelling.

Several studies have adopted the concept of overflow modelling in different fields, where the flow is simulated by the exchange of a single unit between node pairs along the shortest path connecting these nodes (Motter and Lai 2002; Ezzeldin and El-Dakhakhni 2019; Goforth et al. 2020; Alzoor et al. 2021). Within the AM network, this flow is the information or data shared between two AM subject areas. The betweenness centrality has been extensively employed as a metric for how much flow is transmitted through a specific node (Kinney et al. 2005; Kourtellis et al. 2013; Mahyar et al. 2018). As such, the same methodology is adapted in this study, where the BC_i is assumed to be equivalent to the amount of information shared by node i (L_i). In addition, the maximum amount of information (i.e., the capacity) that can be managed by a node i (CAP_i) is assumed to be linearly proportional to its initial load $L_i(0)$, as:

$$CAP_i = \alpha L_i(0) \tag{4-5}$$

where α is a design capacity tolerance that represents the ability of a node to sustain additional information due to any disturbance in the network. The information overload of a node can be represented through Equation 4-5. Therefore, within an AM system, the VI_i represents the susceptibility of an AM subject area to information overload.

4.4. CRITICAL SUBJECT AREAS IN THE ASSET MANAGEMENT NETWORK

The centrality measures described previously were evaluated for the AM system network shown in Figure 4-2. Figure 4-4 presents the BC_i, CC_i, EC_i, and VI_i for the top 15 nodes (i.e., AM subject areas) as these were the most critical nodes to AM system operation. Such nodes represent the most critical nodes within the typical AM system proposed by the IAM that can facilitate the systemic risk propagation within the AM system. The AM subject areas that are critical based on all four centrality measures considered include *Strategic Planning* (S4), *Resource Management* (L3), and *Lifecycle Value Realization* (A4). This observation is well aligned with the definition of AM as the structured decision-making and execution of plans (i.e., strategic plan) developed to optimize a balance between asset performance and risk (i.e., lifecycle value realization) using available resources or the procurement of additional resources (i.e., resource management). The AM subject areas identified within the top 15 based on *three* centrality measures include *Asset Management Planning* (S5), *Operations & Maintenance Decision-Making* (A5), *Stakeholder Engagement* (R1), *Maintenance Delivery* (L1), *Shutdown & Outage Strategy* (A2), and *Shutdown & Outage Management* (L4), whereas those identified based on *two* centrality measures include *Asset Management Strategy & Objectives* (S2), *Asset Performance & Health Monitoring* (R2), *Asset Information Strategy* (I2), *Sustainable Development* (R5), *Asset Costing and Evaluation* (R9), *Procurement and Supply Chain Management* (S2), *Technical Standards &*

Legislation (S2), *Capital Investment Decision-Making* (O1), *Risk Assessment & Management* (R4), and *Asset Operations* (L2). It should be noted that the AM subject areas that have greater centralities for multiple measures are very critical for the functionality of the AM system as they can instigate cascade disruptions when they become dysfunctional. Therefore, adding new connections (i.e., links) between the AM subject areas (i.e., nodes) can enhance the resilience of the network to systemic risks (Goforth et al. 2022).

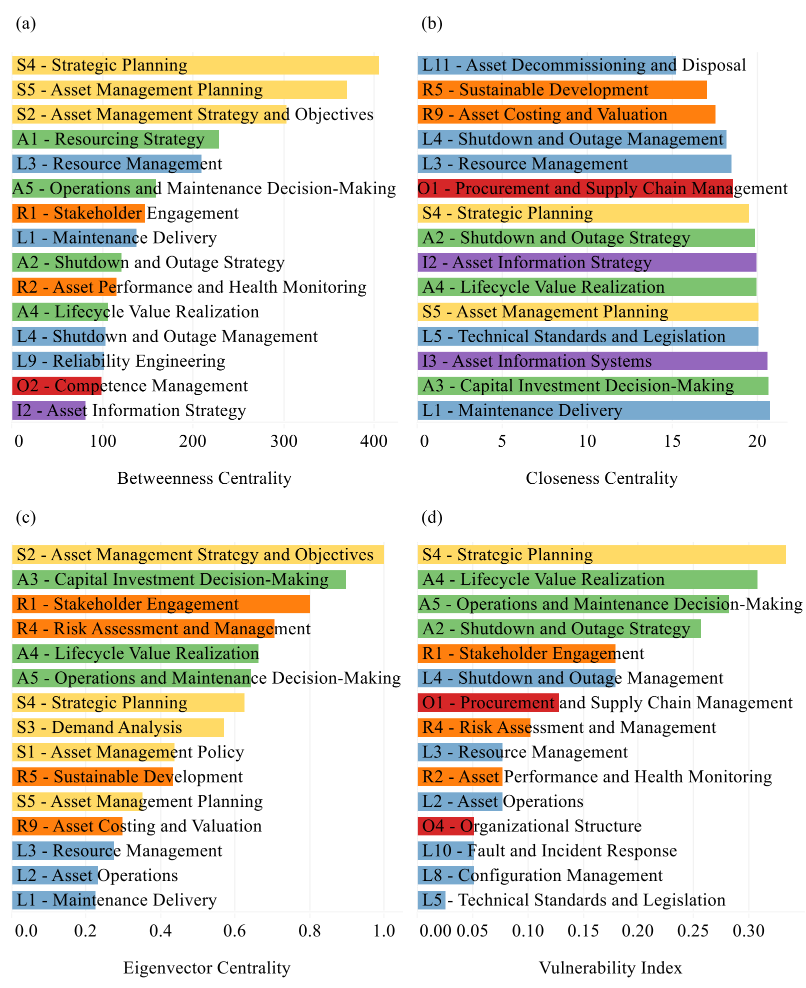


Figure 4-4. The top 15 AM subject areas based on (a) betweenness centrality, (b) closeness centrality, (c) eigenvector centrality, and (d) vulnerability index.

4.5. LINK ADDITION METHODOLOGY

The link addition process is formulated here as an optimization problem with the objective of identifying new connections that can reduce the criticality of highly important AM subject areas and therefore minimize the systemic risk within the AM system. There are 1359 links that do not already exist in the AM system network so there are ${}_{1359}C_r$ possible added link configurations, where r is the specified number of links added to the AM system network. As such, identifying the optimal configuration of added links is challenging and classic linear and nonlinear optimization techniques may trap in local optima. Therefore, GA is employed in the present study to solve the optimization problem, with the objective of minimizing the average centrality of the most critical AM subject areas. Accordingly, the following optimization problem has been formulated:

$$\min_x\left(\frac{a}{N_t}\sum_{i=1}^{N_t}\frac{BC_i(G,x)}{BC_i(G)}+\frac{b}{N_t}\sum_{i=1}^{N_t}\frac{CC_i(G,x)}{CC_i(G)}+\frac{c}{N_t}\sum_{i=1}^{N_t}\frac{EC_i(G,x)}{EC_i(G)}+\frac{d}{N_t}\sum_{i=1}^{N_t}\frac{VI_i(G,x)}{VI_i(G)}\right) \quad (4\text{-}6)$$

subject to:

$$x \notin \mathrm{G} \quad (4\text{-}7)$$

$$a+b+c+d=1 \quad (4\text{-}8)$$

where N_t is the number of critical nodes to be considered, G is the directed graph representing the typical AM system network (shown in Figure 4-2) before adding the new set of links L_{ad} with indices x, and a, b, c, and d are the weighting factors of the betweenness centrality, closeness centrality, eigenvector centrality, and vulnerability index, respectively. The objective function is defined by Equation 4-6 and the constraints are defined by Equations 4-7 and 4-8. The decision variable is the link index vector x which represents the set of newly added links (L_{ad}) to the AM system network to minimize the objective function. Equations 4-6, 4-7, and 4-8 were developed in such a way to ensure that 1) the original AM system nodes are always present in the optimal network configuration and directed links L_{ad} are only added to the original AM system nodes and 2) the AM system still provides the intended functionality of managing the organization's objectives, processes, and assets. It should be noted that the optimization problem represented by Equations 4-6, 4-7, and 4-8 can be solved for either an individual centrality (i.e., betweenness centrality, closeness centrality, eigenvector centrality, or vulnerability index) or a weighted centrality measure. When the former is of interest, the weighting factor corresponding to the considered centrality is 1.0 whereas those corresponding to other centrality measures are zero. When the latter is of interest and a minimized weighted centrality is desired, a, b, c, and d should be chosen according to the desired weighting scheme.

The application of the GA starts with defining the desired number of links to be added (i.e., the size of the set L_{ad}). A population of individuals is subsequently generated randomly, where each individual contains the possible link indices x to be added. Each individual is then assigned a fitness value based on its ability to achieve the objective presented in Equation 4-6. Individuals are evolved continually through a set of reproduction mechanisms, including: 1) elitism, where individuals with high fitness values are replicated in the following generations; 2) crossover, where two individuals are selected based on their fitness and subsequently mixed to produce two offspring; and 3) mutation, through which the entries of a single individual are changed randomly. The reproduction process continues until a termination criterion is achieved. Such criterion may be a maximum number of generations, a certain fitness value, a specific computational time, or a combination of two or more criteria.

In this study, the GA was applied for different link addition configurations with up to 15 links to assess the optimal combination of link additions that yielded the most improved centrality measure. It should be emphasized that adding links increases the information connectivity within the AM system network but can also lead to information overload for AM stakeholders and possibly a breakdown in system functionality (Herrera et al. 2011; Prajogo et al. 2018). Up to 15 added links were chosen as a representative number to illustrate the impact on centrality reduction with respect to links added to the AM system network in order to develop a Pareto Front. As the number of added links increased, a larger population and a higher number of maximum generations were employed to enhance the likelihood of achieving a globally optimal solution. It should be noted that the GA convergence to a globally optimal solution is primarily governed by the population size and the maximum number of generations employed (Yosri et al. 2021). Therefore, the global optimality of a GA solution can be evaluated through: 1) employing initial populations with different sizes for the same maximum number of generations and subsequently evaluating the variability across the obtained solutions; 2) utilizing different values for the maximum number of generations for the same initial population and compare the resulting solutions; or 3) using fixed-sized sets of randomly generated initial populations and assessing the variability in resulting solutions for the same maximum number of generations. The third approach has been employed in this study to evaluate the global optimality of the GA solutions, where the largest population size employed was 5,500 when 15 links were added and the most generations used was 32. Convergence was defined when the evaluated function changed by less than 1×10^{-3} from the previous GA iteration.

4.6. ANALYSIS RESULTS

Figure 4-5 shows the average centrality value of the most critical 15 AM subject areas (i.e., N_t = 15) for the different numbers of added links when the betweenness centrality, closeness centrality, eigenvector centrality, and vulnerability index are considered individually in the objective function defined as Equation 4-6. To identify the true optimum, a balance between the link addition cost and the improvement in the objective function would need to be obtained as described previously. The cost of a link addition would be in terms of resources (e.g., money, people, data, and technology) needed to establish an information connection which would be provided by an organization, but such cost details are proprietary to specific organizations at the time of the development of this study. Therefore, Figure 4-5 presents a Pareto Front with respect to the objective function value and the number of added links, and the following discussion identifies commonalities across different link addition scenarios for each centrality measure used to define the objective function. A full Pareto Analysis would be completed internally by a specific AM organization using link addition costs unique to their organization.

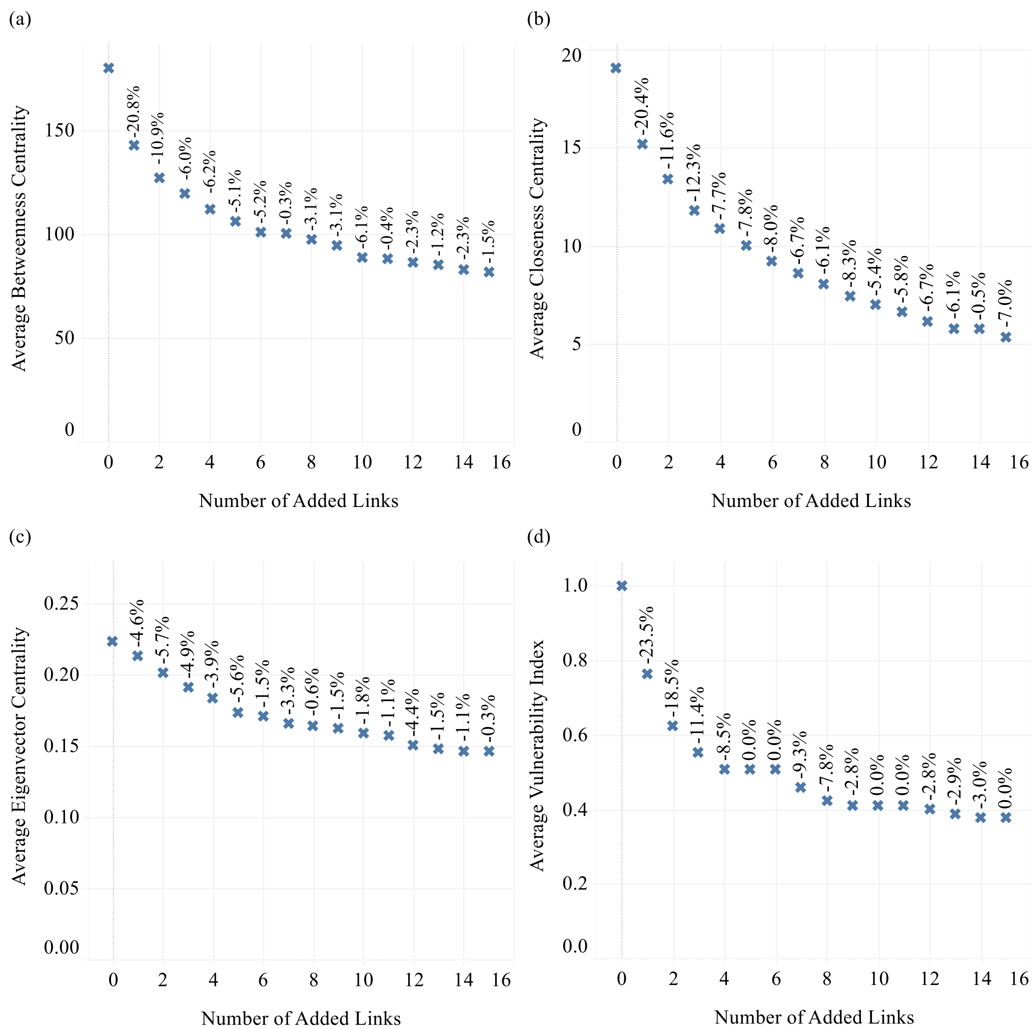


Figure 4-5. Optimization results specific to the number of links added to the AM system network for (a) betweenness centrality, (b) closeness centrality, (c) eigenvector centrality, and (d) vulnerability index. The values above the data points indicate the percent difference from the previous centrality value.

4.6.1. BETWEENNESS CENTRALITY

Figure 4-6 presents the connections that were identified to be added to the typical AM system for up to six added links when minimizing the systemic risk due to critical betweenness centrality AM subject areas is of interest, and Appendix S2-1 provides the labeled connections in a table for one to fifteen link additions. In particular, including just four additional links (i.e., increasing the number of links in the typical AM system from 123, as suggested by the IAM, to 127) would decrease the mean betweenness centrality of the most critical 15 AM subject areas by 38%. The information connection formed between *Asset Management Strategy & Objectives* (S2) and *Resource Management* (L3) is an optimal link addition for one to four link additions and the information connection formed between *Demand Analysis* (S3) and *Management Review, Audit, & Assurance* (R8) is included in the optimal set when adding two, three, and four links. This indicates that the S2→L3 and S3→R8 links are critical in reducing the average betweenness centrality in the highly critical AM subject areas when only up to four link additions are available to implement. Beyond four added links, the S2→L3 and S3→R8 links are replaced by a combination of other links that stem from similar source nodes. The S2→L3 and S3→R8 connections would be feasible for introducing an information or data sharing process in a real AM system. For example, the AM strategy and objectives would specifically provide information on how the resources of the organization should be managed and the analysis of asset demand would provide information related to the review, audit, and assurance processes.

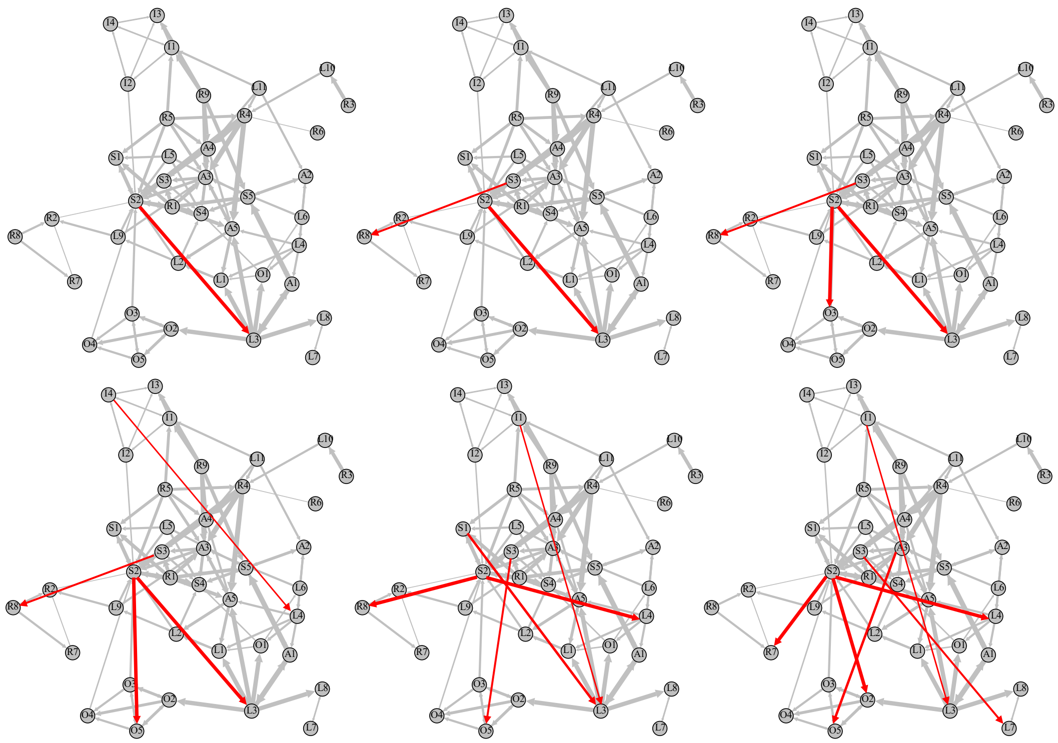


Figure 4-6. Link additions according to the betweenness centrality (labels are defined in Figure 4-1).

4.6.2. Closeness Centrality

Figure 4-7 presents the connections that were identified to be added to the typical AM system for up to six added links when reducing the systemic risk due to nodes with high closeness centrality values is desired, and Appendix S2-2 provides the labeled connections in a table for one to fifteen link additions. The information connection formed between the *Strategic Planning* (S4) and *Asset Decommissioning & Disposal* (L11) areas is optimal among all link addition scenarios from one to six. All optimal link additions identified originate from the *Strategic Planning* (S4) subject area, highlighting the importance of links originating from this node in reducing the average closeness centrality within the AM system (i.e., reducing the average distance of the 15 critical AM subject areas to the other AM subject areas in the AM system network). The S4→L11 link would be feasible within a working AM system if the strategic plan provided additional information to describe the process and the feedback required for asset decommissioning and disposal. Adding the three links identified as optimal within the typical AM system decreases the average closeness centrality of the most critical 15 AM subject areas by 30%, whereas adding the six links identified as optimal provides a 52% reduction. Both scenarios provide a large reduction in the closeness centrality-based objective function without greatly increasing the number of links, and therefore the possibility of information overload, within the AM system network.

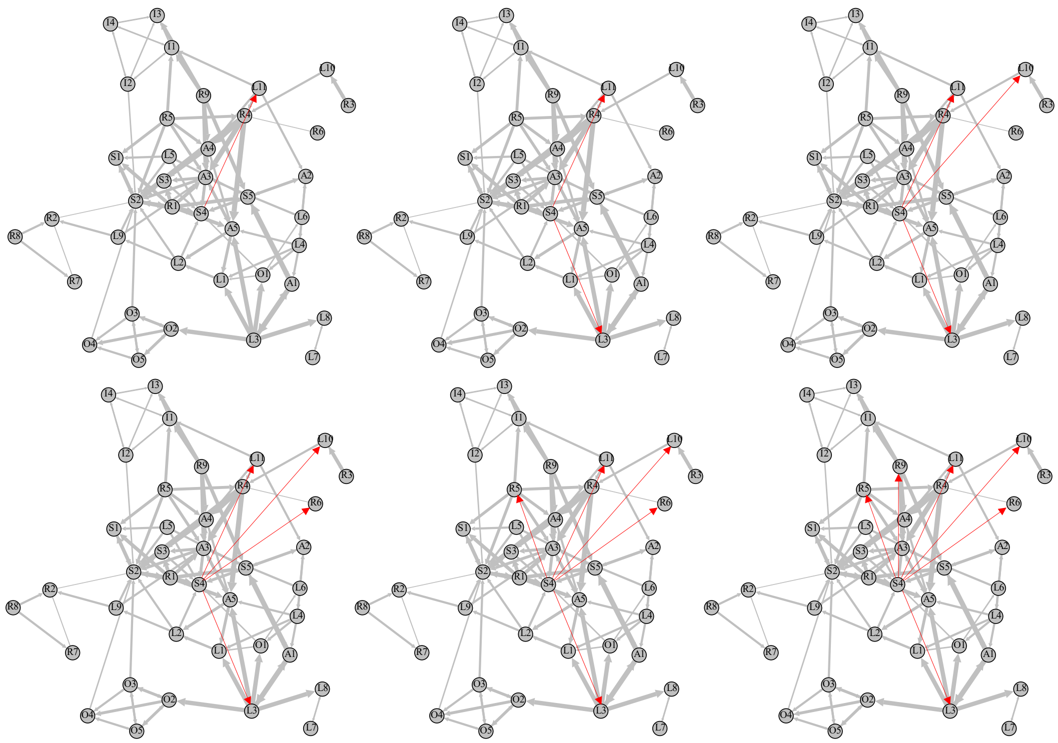


Figure 4-7. Link additions according to the closeness centrality (labels are defined in Figure 4-1).

4.6.3. EIGENVECTOR CENTRALITY

Figure 4-8 presents the connections that were identified to be added to the typical AM system for up to six added links when reducing the systemic risk due to the eigenvector centrality is the focus, and Appendix S2-3 provides the labeled connections in a table for 1 to 15 link additions. An 18% reduction in the systemic risk due to the disruption of AM subject areas with high eigenvector centralities can be obtained by adding just four links to the AM system network. The information connection formed between the *Risk Assessment & Management* (R4) and *Sustainable Development* (R5) areas is optimal for each link addition scenario in

Figure 4-8 and the information connection from *Risk Assessment & Management* (R4) and *Stakeholder Engagement* (R1) areas is optimal among all link addition scenarios from two to six. Such occurrences highlight the criticality of each link in the reduction in eigenvector centrality-based systemic risk (i.e., reducing the reliance of the AM system on information transfer through important AM subject areas connected to other important AM subject areas). Both the R4→R5 and the R4→R1 added links are feasible to be implemented in an AM system if appropriate information or data transfers were available. For example, the risk assessment and management process could provide details on a risk-based plan for sustainable development and for engaging with AM stakeholders. The *Risk Assessment & Management* (R4) subject area is a common source node for many links in Figure 4-8, indicating its importance to the reduction in eigenvector centrality-related systemic risk of the most critical AM subject areas.

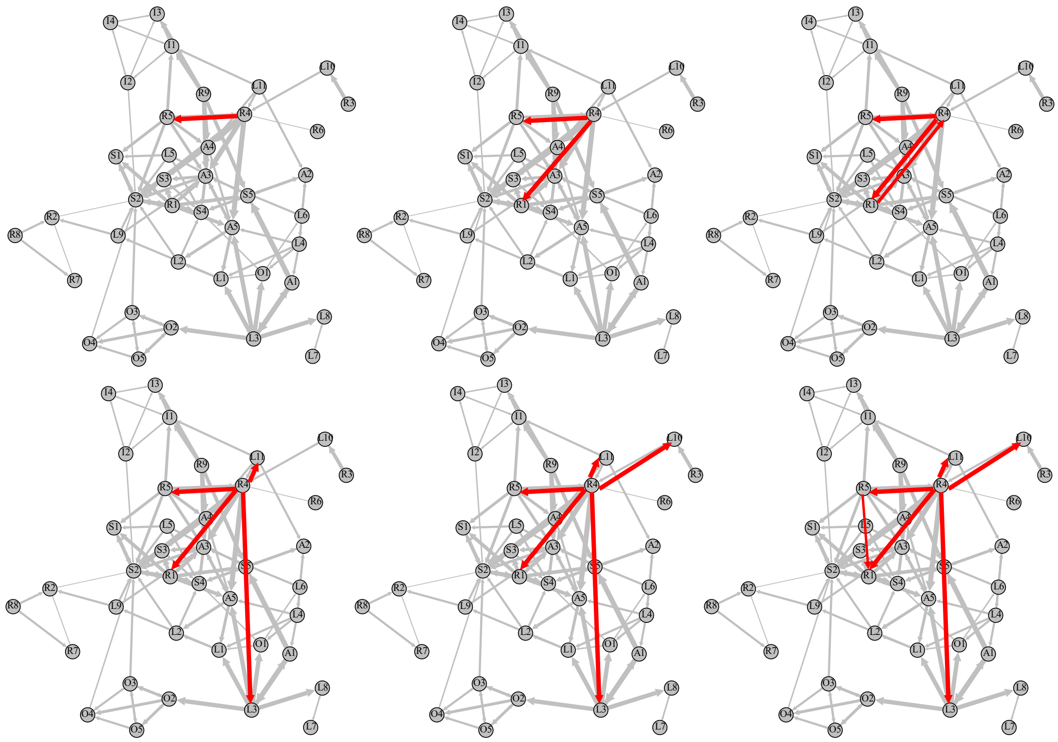


Figure 4-8. Link additions according to the eigenvector centrality (labels are defined in Figure 4-1).

4.6.4. Vulnerability Index

Figure 4-9 presents the connections that were identified to be added to the typical AM system for up to six added links when reducing the systemic risk due to the average vulnerability index of the top 15 subject areas is the only goal, and Appendix S2-4 provides the labeled connections in a table for 1 to 15 link additions. The α value from Equation 4-5 was specified as 0.05, indicating that if a node's information load exceeded its capacity by 5%, the node was considered to be in a state of information overload. 5% was chosen as a representative value to illustrate the application of the previously described methodology and it has been used in other vulnerability index applications (Ezzeldin and El-Dakhakhni 2019; Goforth et al. 2020). The optimal added links presented in Figure 4-9 do not have significant commonalities among each of the six link addition scenarios. There are only three links (i.e., L1→R1, L5→R8, and L8→S4) that are optimal among two consecutive link addition scenarios. This lack of link addition commonality highlights the importance of applying this optimization methodology when looking to reduce the systemic risk related to vulnerability index (i.e., reducing the potential for information overload scenarios within the AM system network), as a small number of specific links do not provide a consistent reduction as was the case with the previous centrality measures. There is still great value in implementing the proposed link addition configurations as adding the identified four optimal links can lead to a 49% decrease in the average vulnerability index of the most critical 15 AM subject areas.

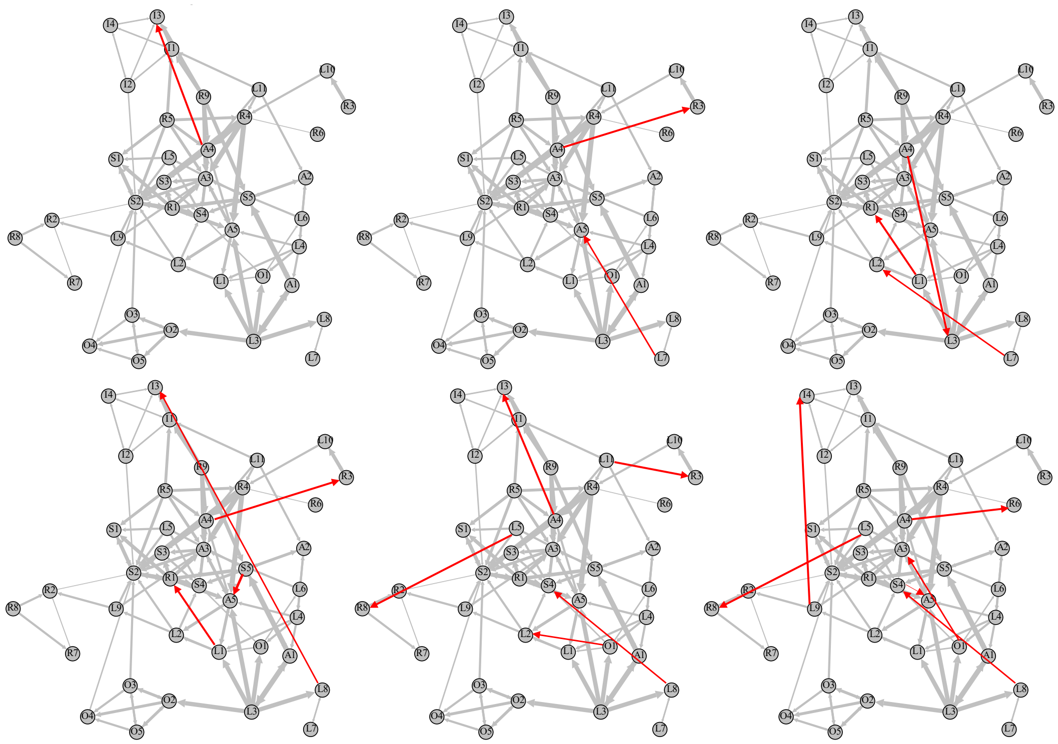


Figure 4-9. Link additions according to the vulnerability index (labels are defined in Figure 4-1).

4.6.5. WEIGHTED COMBINATION

While the previously described applications considered minimizing each of betweenness centrality, closeness centrality, eigenvector centrality, and vulnerability index separately, minimizing a weighted combination of such metrics is also crucial as each centrality measure evaluates a different aspect of the network's exposure to systemic risks. As such, the values of *a*, *b*, *c*, *d* in Equation 4-7 were each assumed as 0.25 and the resulting objective function evaluations are shown in Figure 4-10 for link additions from 1 to 15. Figure 4-11 presents the connections that were identified to be added to the typical AM system for up to six added links and Appendix S2-5 provides the labeled connections in a table for 1 to 15 link additions. The information connection between the *Strategic Planning* (S4) and *Resource Management* (L3) areas is common among most link addition scenarios, highlighting its importance in reducing the combined systemic risk according to the weighted objective function (19% reduction). The importance of the S4→L3 link is also consistent with the criticality of each AM subject area, as both S4 and L3 are critical for all centrality measures as determined from Figure 4-4. Adding just four links can provide a 40% reduction in the weighted objective function value from the original AM system network. This large reduction in the objective function evaluation further highlights the impact of a small number of added links in reducing systemic risk within the AM system network.

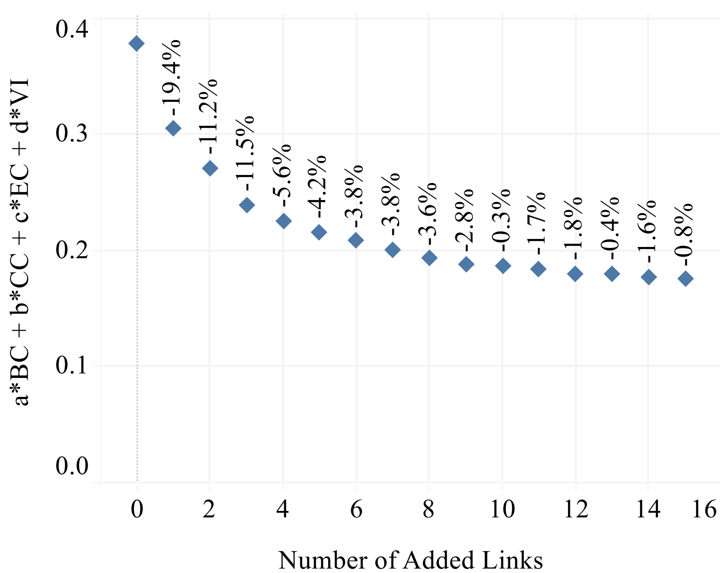


Figure 4-10. An equal weighted combination of betweenness centrality, closeness centrality, eigenvector centrality, and vulnerability index according to Equation 4-6 evaluated for different link additions. The values above the data points indicate the percent difference from the previous centrality value.

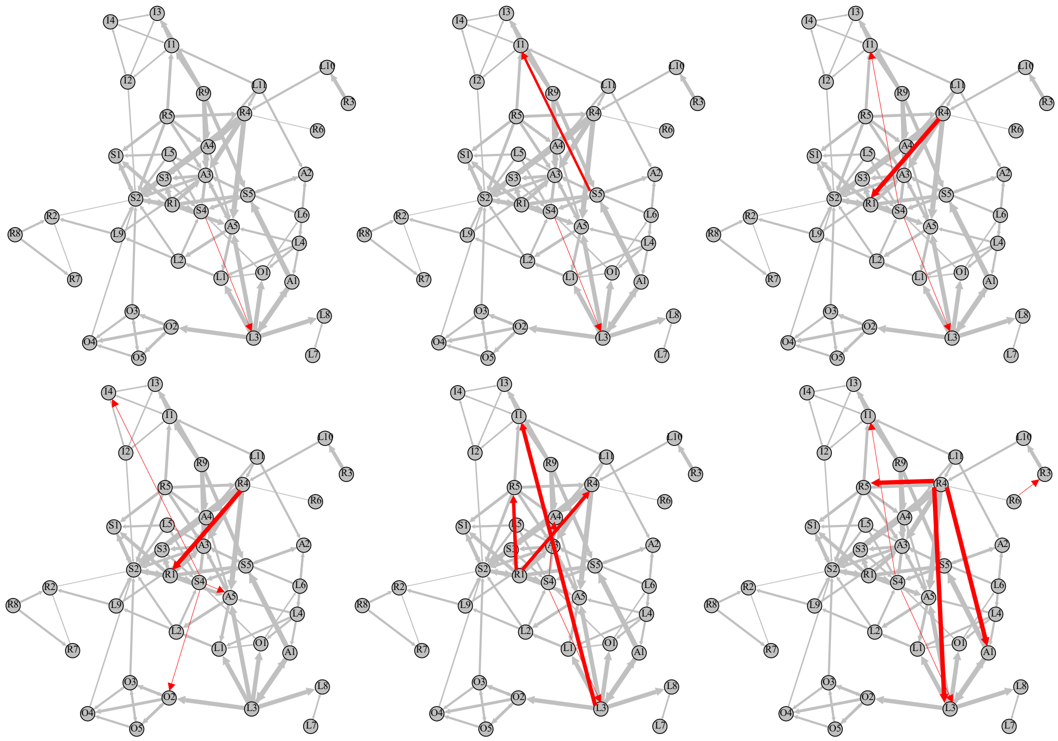


Figure 4-11. Link additions based on an equal weighted combination of betweenness centrality, closeness centrality, eigenvector centrality, and vulnerability index (labels are defined in Figure 4-1).

4.7. Managerial Insights

In each of the cases above, it was found that adding only a small number of links (less than 5) provided large reductions in each of the evaluated objective functions (18% (*EC*) to 49% (*VI*)). This result is valuable for AM organizations as they could reduce their systemic risk exposure with minimal added links and therefore minimize the exposure of their AM stakeholders to information overload. However, there is little commonality in added links among each of the individual centrality-based objective functions. This highlights the importance for organizations and managers, within such organizations, to evaluate the relative importance of each centrality-based systemic risk to their organization and run the optimization methodology accordingly.

The added links associated with the weighted combination, outlined in Figure 4-11, were similar to the optimal links for the closeness and eigenvector metrics. Specifically, the *Strategic Planning* (S4) to *Resource Management* (L3) link was important for the reduction in closeness centrality and it was also important to the weighted combination reduction. Additionally, the links that originated from the *Risk Assessment & Management* (R4) subject area were critical for the reduction in the weighted combination as was similar in the eigenvector centrality reduction. This demonstrates how evaluating a weighted combination of the described metrics allows individual organizations to optimize their own AM system network for their desired systemic risk reduction focus. This facilitates implementation of this optimization methodology in AM organizations across all infrastructure classes.

4.8. CONCLUSION

Implementing a strong AM system is critical for organizations that own, manage, and operate infrastructure assets to ensure that their assets can provide the greatest lifecycle value and the best service to their customers or users. A well-defined asset management system can help organizations alleviate the challenges induced by infrastructure ageing, strict regulatory operating conditions, limited financial means, and losing valuable experience due to retirements. This study evaluated the AM system, developed by the IAM, as a network of connected components (i.e., nodes and links). Nodes within the network represent the AM subject areas, whereas links simulate the information transfer between node pairs. The ISO 55001 standard specifies the best practice for organizations in their implementation and operation of AM systems. Based on this, the AM system network link weights were defined in this study as the number of ISO 55001 clauses that pertain to the source node. The main challenges of AM system operation are caused by information asymmetry between AM system components (e.g., components not using consistent information for decision-making) and information overload within AM system components. These main challenges initiate systemic risks (i.e., dependence-induced disruptions) within the AM system network which can lead to cascading disruptions throughout the AM system network. To address this, this study proposed a connection addition methodology that can be adopted to reduce the systemic risks within a typical AM system.

The proposed methodology employs a GA optimization to identify the optimal connection configuration necessary for reducing the systemic risk. Betweenness centrality, closeness centrality, eigenvector centrality, and vulnerability index were used to evaluate the criticality of the different AM subject areas for facilitating the information transfer within the AM system. Identifying such AM subject areas is essential as their disruption can initiate a systemic risk situation within the AM system and may lead to a complete malfunction. A GA was deployed to reduce the systemic risk within the AM system through minimizing the four centrality measures considered. The objective function was evaluated to minimize the betweenness centrality, closeness centrality, eigenvector centrality, and vulnerability index individually, and optimal link configurations were presented visually for one to six added links. To further extend the application of the proposed link addition methodology, another objective function was defined as a weighted combination of the four centrality measures, and the optimal added link configurations were again presented visually for one to six added links. Among all evaluations, it was found that adding a small number of links (less than 5) provided a large reduction in the objective function values (18% (*EC*) to 49% (*VI*)).

Managerial insights were also presented highlighting that there were very few added link commonalities between the different evaluated objective functions, thus necessitating an organization to clarify their systemic risk reduction goals and highlighting the value of using a weighted combination of different metrics. Overall, organizations could use the results from this study to reduce the exposure of their AM system to systemic risks due to the potential cascading disruption of highly connected and critical AM subject areas. The connection addition methodology proposed in this study enhances the information transfer throughout the AM system while also ensuring that AM stakeholders are not overloaded with too much information. Future work would be to apply this connection addition methodology at different stages of AM deployment within organizations and with link addition costs, allowing a full Pareto analysis to minimize the centrality values while also minimizing the cost of adding new links to the AM system network.

4.9. ACKNOWLEDGEMENTS

The financial support for the study was provided through the Canadian Nuclear Energy Infrastructure Resilience under Systemic Risk (CaNRisk) – Collaborative Research and Training Experience (CREATE) program of the Natural Science and Engineering Research Council (NSERC) of Canada. The *INTERFACE* Institute and the INViSiONLab support in the development of this study is also acknowledged.

4.10. NOTATIONS

BC_i	Betweenness centrality of node *i*
ρ_{jk}	Weighted length of all shortest paths connecting nodes *j* and *k*
$\rho_{jk}(i)$	Weighted length of all shortest paths connecting nodes *j* and *k* that traverse node *i*
CC_i	Closeness centrality of node *i*
$d(i,j)$	Shortest path length between nodes *i* and *j*
N	The total number of nodes in the network
EC_i	Eigenvector centrality of node *i*
$\boldsymbol{A}$	Adjacency matrix with elements $a_{i,j}$
λ	The maximum eigenvalue of the adjacency matrix $\boldsymbol{A}$
VI_i	Vulnerability index of node *i*
N'	The total number of operational nodes in the network's giant component after node *i* was triggered to fail
L_i	Information load on node *i*
CAP_i	The maximum amount of information that can be managed by node *i*

α	Tolerance for exceeding information load capacity
r	The specified number of links added to the AM system network
N_t	The number of critical nodes to be considered in the objective function
G	The directed graph representing the typical AM system network
a	Weighting factor for betweenness centrality
b	Weighting factor for closeness centrality
c	Weighting factor for eigenvector centrality
d	Weighting factor for vulnerability index
x	Indices for newly added links
L_{ad}	The new set of added links

4.11. ACRONYMS

AM	Asset management
GA	Genetic algorithm
IAM	Institute for Asset Management
ISO	International Organization for Standardization

4.12. Supplemental Materials

Appendix S1 – Weighted adjacency matrix.

	S1 - Asset Management Policy	S2 - Asset Management Strategy & Objectives	S3 - Demand Analysis	S4 - Strategic Planning	S5 - Asset Management Planning	A1 - Resourcing Strategy	A2 - Shutdown & Outage Strategy	A3 - Capital Investment Decision-Making	A4 - Lifecycle Value Realization	A5 - Operations & Maintenance Decision-Making	L1 - Maintenance Delivery	L2 - Asset Operations	L3 - Resource Management	L4 - Shutdown & Outage Management	L5 - Technical Standards & Legislation	L6 - Asset Creation & Acquisition	L7 - Systems Engineering	L8 - Configuration Management	L9 - Reliability Engineering	L10 - Fault & Incident Response	L11 - Asset Decommissioning & Disposal	I1 - Data & Information Management	I2 - Asset Information Strategy	I3 - Asset Information Systems	I4 - Asset Information Standards	O1 - Procurement & Supply Chain Management	O2 - Competence Management	O3 - Asset Management Leadership	O4 - Organizational Structure	O5 - Organizational Culture	R1 - Stakeholder Engagement	R2 - Asset Performance & Health Monitoring	R3 - Contingency Planning & Resilience Analysis	R4 - Risk Assessment & Management	R5 - Sustainable Development	R6 - Management of Change	R7 - Asset Management System Monitoring	R8 - Management Review, Audit, & Assurance	R9 - Asset Costing & Valuation
S1 - Asset Management Policy		4																																					
S2 - Asset Management Strategy & Objectives	5		5	5																											5								
S3 - Demand Analysis		3		3																																			
S4 - Strategic Planning		1	1		1																										1								
S5 - Asset Management Planning				4		4	4																																
A1 - Resourcing Strategy					7								7																										
A2 - Shutdown & Outage Strategy					3									3																									
A3 - Capital Investment Decision-Making		4	4	4					4	4																													
A4 - Lifecycle Value Realization		3	3	3				3		3																													
A5 - Operations & Maintenance Decision-Making								4			4	4																											
L1 - Maintenance Delivery										3		3							3																				
L2 - Asset Operations		3		3																																			
L3 - Resource Management						6				6	6							6								6	6												
L4 - Shutdown & Outage Management						3	3			3	3															3													
L5 - Technical Standards & Legislation	3	3		3	3																																		
L6 - Asset Creation & Acquisition								3																															
L7 - Systems Engineering																		2																					
L8 - Configuration Management																	2																						
L9 - Reliability Engineering		3						3																								3							
L10 - Fault & Incident Response																																	3	3					
L11 - Asset Decommissioning & Disposal							3		3													3												3					
I1 - Data & Information Management																							2	2	2														
I2 - Asset Information Strategy		2																				2		2	2														
I3 - Asset Information Systems																						1	1		1														
I4 - Asset Information Standards																						2	2	2															
O1 - Procurement & Supply Chain Management				2							2					2																							
O2 - Competence Management																												4	4	4									
O3 - Asset Management Leadership		3																									3		3	3									
O4 - Organizational Structure		2																									2	2		2									
O5 - Organizational Culture																											3	3	3										
R1 - Stakeholder Engagement	5	5	5	5	5			5		5																													
R2 - Asset Performance & Health Monitoring		1																																			1		
R3 - Contingency Planning & Resilience Analysis																				5																			
R4 - Risk Assessment & Management		7						7	7	7																													
R5 - Sustainable Development	4	4						4	4													4												4					
R6 - Management of Change																																		1					
R7 - Asset Management System Monitoring																																1						1	
R8 - Management Review, Audit, & Assurance																																3					3		
R9 - Asset Costing & Valuation					5			5	5													5		5															

Appendix S2-1. Link additions according to the betweenness centrality (labels are defined in Figure 4-1).

Link Additions	
1	S2,L3
2	S2,L3 S3,R8
3	S2,L3 S2,O3 S3,R8
4	I4,L4 S2,L3 S2,O5 S3,R8
5	I1,L3 S1,L3 S2,L4 S2,R8 S3,O5
6	A3,O5 I1,L3 S2,L4 S2,O2 S2,R7 S3,L7
7	I3,L4 I4,O5 R7,L4 S2,L3 S2,L8 S2,O3 S2,R8
8	A3,R8 I3,O5 L2,O5 O2,L9 S2,L3 S2,L8 S2,O3 S3,A2
9	A2,O3 I1,O2 I3,11 S1,L9 S2,L4 S2,L8 S2,O1 S2,O3 S2,R8
10	A2,O3 I1,O4 I2,L9 I4,L2 L2,O1 O2,L3 R7,L4 S1,L3 S2,R8 S3,O5
11	I3,L3 L6,L4 O5,L9 R3,L5 R8,A2 S1,O3 S2,L3 S2,L6 S2,L7 S2,R8 S3,R7
12	A4,O2 I1,O3 I2,R8 I3,L3 L2,O4 O5,L4 R8,A2 S1,R7 S2,L3 S2,L6 S2,L7 S3,O3
13	A5,R8 I4,L3 L6,O5 O2,10 O3,L7 O5,L9 R4,O4 R6,O4 R7,O3 S1,O1 S2,L7 S2,R8 S3,O5
14	A3,O3 I1,R8 I2,R1 I4,O3 L1,O2 O4,L3 O5,L4 O5,L9 R2,L3 S1,R7 S2,A2 S2,O1 S2,O2 S3,L8
15	A5,R8 I1,R7 I3,L3 I3,O5 L1,O3 L6,L1 O2,L1 O5,A2 R8,L6 S1,L3 S1,R2 S2,L4 S2,L8 S2,O1 S3,O2

Appendix S2-2. Link additions according to the closeness centrality (labels are defined in Figure 4-1).

Link Additions	
1	S4,L11
2	S4,L3 S4,L11
3	S4,L3 S4,L10 S4,L11
4	S4,L3 S4,L10 S4,L11 S4,R6
5	S4,L3 S4,L10 S4,L11 S4,R5 S4,R6
6	S3,R9 S4,L3 S4,L10 S4,L11 S4,R5 S4,R6
7	L4,11 S2,R9 S4,L3 S4,L10 S4,L11 S4,R5 S4,R6
8	I3,O1 I3,R6 I4,L5 R2,R9 S4,L10 S4,L11 S4,O4 S4,R5
9	I2,L5 I3,R5 I4,L10 R7,L11 S3,R9 S4,L3 S4,O3 S4,R6 S4,R7
10	A5,L10 I2,L5 I3,R5 I3,R6 R7,L8 R7,L11 R7,O1 R7,O3 S3,R9 S4,R7
11	I2,L10 I4,O1 O1,R9 R2,L3 R2,L5 R7,O1 R8,R5 S4,L11 S4,O3 S4,R2 S4,R6
12	A2,L5 I1,L7 I1,L10 I2,R7 I3,O1 L7,R6 R7,R9 S4,I3 S4,L3 S4,L11 S4,O3 S4,R5
13	I3,L5 I3,L7 I3,L11 I3,O4 I3,R2 I3,R6 I4,O1 O4,L4 R6,I2 S2,R9 S4,I3 S4,L10 S4,R5
14	I3,L5 I4,L3 I4,O1 L1,R9 L11,R6 R2,L4 R2,L5 R2,R5 R7,I3 R7,L7 R7,L10 S4,L11 S4,O4 S4,R7
15	A5,L11 I3,R7 I3,S4 L1,R9 L2,L5 R2,I3 R6,L1 R7,L3 R7,L7 R7,R5 R7,R6 S4,L4 S4,L10 S4,O5 S4,R7

Appendix S2-3. Link additions according to the eigenvector centrality (labels are defined in Figure 4-1).

Link Additions	
1	R4,R5
2	R4,R1 R4,R5
3	R1,R4 R4,R1 R4,R5
4	R4,L3 R4,L11 R4,R1 R4,R5
5	R4,L3 R4,L10 R4,L11 R4,R1 R4,R5
6	R4,L3 R4,L10 R4,L11 R4,R1 R4,R5 R5,R1
7	R4,A1 R4,L3 R4,L10 R4,L11 R4,R1 R4,R5 R4,R9
8	R4,L3 R4,L10 R4,L11 R4,R1 R4,R3 R4,R5 R4,R9 R9,R1
9	L11,R5 R4,A1 R4,L3 R4,L10 R4,L11 R4,O2 R4,R1 R4,R5 R4,R9
10	R4,A1 R4,L3 R4,L10 R4,L11 R4,R1 R4,R3 R4,R5 R4,R9 R4,S5 R5,S3
11	R4,L3 R4,L10 R4,L11 R4,O3 R4,R1 R4,R3 R4,R5 R4,R9 R5,L3 R5,L9 R5,L11
12	R4,L1 R4,L3 R4,L9 R4,L10 R4,L11 R4,R1 R4,R3 R4,R5 R4,R6 R4,R9 R9,R1 R9,R4
13	R1,R9 R4,A1 R4,I1 R4,L3 R4,L10 R4,L11 R4,R1 R4,R3 R4,R5 R4,R9 R9,R1 R9,R4 R9,R5
14	R4,A1 R4,L3 R4,L10 R4,L11 R4,R1 R4,R5 R4,R9 R4,S5 R5,L10 R5,L11 R5,R1 R5,R9 R9,R1 R9,R4
15	I2,S1 R4,L3 R4,L4 R4,L9 R4,L10 R4,L11 R4,R1 R4,R3 R4,R5 R4,R9 R5,L5 R5,L10 R5,R9 R9,R1 R9,R4

Appendix S2-4. Link additions according to the vulnerability index (labels are defined in Figure 4-1).

Link Additions	
1	A4,I3
2	A4,R3 L7,A5
3	A4,I3 L1,R1 L7,L2
4	A4,R3 L1,R1 L8,I3 S5,A5
5	A4,I3 L5,R8 L8,S4 L11,R3 O1,L2
6	A4,R6 L5,R8 L8,S4 L9,I4 O1,A3 S4,A5
7	A4,R5 L1,S2 L4,S5 O1,L2 R3,L7 R9,R1 S5,A5
8	A4,R6 A5,I3 L1,S2 L3,S5 L11,R7 R3,L7 R9,R1 S5,A5
9	A1,O1 A4,R3 A5,S4 I2,S4 L6,I3 L7,A5 O1,S2 R5,R7 R9,R7
10	A2,A5 A4,L10 L6,A1 L6,I3 L8,A5 L11,L7 R1,R9 R5,R7 R9,R1 S2,S5
11	A4,S1 L3,A4 L4,A3 L7,R3 L11,R7 O4,I3 R6,S5 R7,S2 S4,A4 S5,A5 S5,L8
12	A3,R6 A4,I1 L1,S2 L5,R8 L7,R1 O1,L2 O4,S4 R1,R9 R5,R3 R9,R1 R9,S2 S4,A5
13	A1,A3 A1,O1 A4,R5 L1,S2 L3,S2 L5,R8 L11,L7 L11,R3 R4,R3 R9,R1 R9,R6 S4,A5 S5,A5
14	A4,R3 L2,S3 L5,L7 L8,S4 O1,S2 R1,L1 R1,R9 R6,R3 R8,I1 R9,I2 R9,R1 R9,R5 S4,A5 S5,A5
15	A1,O1 L7,A3 L11,I3 O1,I3 R1,R3 R1,R9 R4,S5 R9,R1 R9,R5 R9,R6 R9,R7 R9,S2 S4,A4 S4,L1 S4,O2

Appendix S2-5. Link additions based on a weighted combination of betweenness centrality, closeness centrality, eigenvector centrality, and vulnerability index (labels are defined in Figure 4-1).

Link Additions	
1	S4,L3
2	S4,L3 S5,I1
3	R4,R1 S4,I1 S4,L3
4	R4,R1 S4,A5 S4,I4 S4,O2
5	L3,I1 R1,R4 R1,R5 S4,A4 S4,L3
6	R4,A1 R4,L3 R4,R5 R6,R3 S4,I1 S4,L3
7	L10,S4 R1,R5 R2,S4 R4,R1 R4,R5 S4,A5 S4,L7
8	R1,R3 R4,R1 R4,R5 R5,R1 S4,A4 S4,I1 S4,L3 S4,L9
9	A2,L8 R1,R4 R4,R1 R4,R5 S4,A4 S4,I1 S4,I4 S4,L1 S4,O2
10	A4,R7 L5,A1 R4,A1 R4,R1 R4,R5 R4,S4 S4,A4 S4,I4 S4,L8 S4,O2
11	R1,R9 R4,R1 R4,R5 R4,S5 S4,A4 S4,I1 S4,L5 S4,L7 S4,O2 S4,R2 S4,R7
12	R1,R9 R4,A1 R4,R1 R4,R5 R4,R9 R5,I3 R9,R1 S4,A4 S4,I2 S4,L1 S4,L3 S4,L8
13	A4,R9 L10,S4 L11,S2 R1,R9 R4,R1 R4,R5 R4,R9 R8,S4 R9,R1 S4,I3 S4,L1 S4,O2 S4,R5
14	A2,L3 R1,R9 R4,L4 R4,R1 R4,R9 R4,S3 R4,S4 R9,R1 S4,A3 S4,I1 S4,L1 S4,L8 S4,L9 S4,O5
15	A2,S4 L1,A3 O3,S4 R1,R9 R4,L3 R4,R1 R4,R5 R4,R9 R9,R1 R9,R6 S4,I3 S4,L1 S4,L3 S4,L8 S4,R2

4.13. REFERENCES

Alzoor, F. S., Ezzeldin, M., Mohamed, M., and El-Dakhakhni, W. (2021). "Prioritizing Bridge Rehabilitation Plans through Systemic Risk-Guided Classifications." Journal of Bridge Engineering, 26(7), 04021038.

American Society of Civil Engineers. (2021). 2021 Report Card for America's Infrastructure.

Barabási, A.-L. (2016). Network Science. Cambridge University Press, Cambridge, UK.

Barbosa, F., De Sousa, A., and Agra, A. (2018). "Topology design of transparent optical networks resilient to multiple node failures." Proceedings of 2018 10th International Workshop on Resilient Networks Design and Modeling, IEEE, 1–8.

Bergh, D. D., Ketchen, D. J., Orlandi, I., Heugens, P. P. M. A. R., and Boyd, B. K. (2019). "Information Asymmetry in Management Research: Past Accomplishments and Future Opportunities." Journal of Management, 45(1), 122–158.

Bertling Tjernberg, L. (2018). Infrastructure Asset Management with Power System Applications. Taylor & Francis Group, Boca Raton.

Bhavathrathan, B. K., and Patil, G. R. (2018). "Algorithm to Compute Urban Road Network Resilience." Transportation Research Record, 2672(48), 104–115.

Brunetto, Y., Xerri, M., and Nelson, S. (2014). "Building a Proactive, Engagement Culture in Asset Management Organizations." Journal of Management in Engineering, 30(4), 04014014.

Canadian Infrastructure Report Card. (2016). Informing the Future: The Canadian Infrastructure Report Card.

Canadian Infrastructure Report Card. (2019). Monitoring the State of Canada's Core Public Infrastructure: The Canadian Infrastructure Report Card 2019.

Crescenzi, P., D'Angelo, G., Severini, L., and Velaj, Y. (2016). "Greedily improving our own closeness centrality in a network." ACM Transactions on Knowledge Discovery from Data, 11(1), 1–32.

Estrada, E., and Knight, P. (2015). A first course in network theory. Oxford University Press.

Ezzeldin, M., and El-Dakhakhni, W. E. (2019). "Robustness of Ontario power network under systemic risks." Sustainable and Resilient Infrastructure, Taylor & Francis, 1–20.

Freeman, L. C. (1977). "A Set of Measures of Centrality Based on Betweenness." Sociometry, 40(1), 35–41.

Global Forum on Maintenance and Asset Management. (2014). The Asset Management Landscape. Global Forum on Maintenance and Asset Management.

Goforth, E., El-Dakhakhni, W., and Wiebe, L. (2022). "Network Analytics for Infrastructure Asset Management Systemic Risk Assessment." Journal of Infrastructure Systems, In press.

Goforth, E., Ezzeldin, M., El-Dakhakhni, W., Wiebe, L., and Mohamed, M. (2020). "Network-of-Networks Framework for Multimodal Hazmat Transportation Risk Mitigation: Application to Used Nuclear Fuel in Canada." Journal of Hazardous, Toxic, and Radioactive Waste, 24(3).

Golightly, D., Kefalidou, G., and Sharples, S. (2018). "A cross-sector analysis of human and organisational factors in the deployment of data-driven predictive maintenance." Information Systems and e-Business Management, Springer Berlin Heidelberg, 16(3), 627–648.

Herrera, F., Chan, G., Legault, M., Kassim, R. M., and Sharma, V. (2011). The digital workplace: Think, share, do. Deloitte & Touche LLP.

Hodkiewicz, M. R. (2015). "The development of ISO 55000 series standards." Lecture Notes in Mechanical Engineering, 19, 427–438.

Infrastructure Canada. (2018). Investing in Canada: Canada's long-term infrastructure plan. Ottawa.

Institute for Asset Management. (2015). Asset Management – an Anatomy. Bristol, UK.

International Organization for Standardization. (2014a). ISO 55000 — Asset management — Overview, principles and terminology. Geneva.

International Organization for Standardization. (2014b). "ISO 55001 Asset management — Management systems — Requirements." International Organization for Standardization, Geneva.

International Organization for Standardization. (2021). “Known Certified Organizations.” ISO/TC 251 Asset Management, <https://committee.iso.org/sites/tc251/social-links/resources/known-certified-organizations.html> (Oct. 28, 2021).

Kinney, R., Crucitti, P., Albert, R., and Latora, V. (2005). “Modeling cascading failures in the North American power grid.” European Physical Journal B, 46(1), 101–107.

Konstantakos, P. C., Chountalas, P. T., and Magoutas, A. I. (2019). “The contemporary landscape of asset management systems.” Quality - Access to Success, 20(169), 10–17.

Kourtellis, N., Alahakoon, T., Simha, R., Iamnitchi, A., and Tripathi, R. (2013). “Identifying high betweenness centrality nodes in large social networks.” Social Network Analysis and Mining, 3(4), 899–914.

de la Pena, C., Gonzalez Fernandez, D., and Rodriguez Gonzalez, J. (2016). “How analytics can improve asset management in electric power networks.” McKinsey & Company Electrical Power & Natural Gas, <https://www.mckinsey.com/industries/electric-power-and-natural-gas/our-insights/> (Jun. 14, 2018).

Mahyar, H., Hasheminezhad, R., Ghalebi, E., Nazemian, A., Grosu, R., Movaghar, A., and Rabiee, H. R. (2018). “Identifying central nodes for information flow in social networks using compressive sensing.” Social Network Analysis and Mining, Springer Vienna, 8(1), 1–24.

Medvet, E., and Bartoli, A. (2021). “Evolutionary Optimization of Graphs with GraphEA.” AIxIA 2020 -- Advances in Artificial Intelligence, M. Baldoni and S. Bandini, eds., Springer International Publishing, Cham, 83–98.

Morshedlou, N., Barker, K., González, A. D., and Ermagun, A. (2021). “A heuristic approach to an interdependent restoration planning and crew routing problem.” Computers and Industrial Engineering, 161(August).

Motter, A. E., and Lai, Y. C. (2002). “Cascade-based attacks on complex networks.” Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 66(6), 4.

Nozhati, S., Sarkale, Y., Ellingwood, B., K.P. Chong, E., and Mahmoud, H. (2019). “Near-optimal planning using approximate dynamic programming to enhance post-hazard community resilience management.” Reliability Engineering and System Safety, 181, 116–126.

Ohara, K., Saito, K., Kimura, M., and Motoda, H. (2017). “Maximizing network performance based on group centrality by creating most effective k-links.” International Conference on Data Science and Advanced Analytics, IEEE, Tokyo, 561–570.

Papagelis, M. (2015). “Refining social graph connectivity via shortcut edge addition.” ACM Transactions on Knowledge Discovery from Data, 10(2), 1–35.

Parotsidis, N., Pitoura, E., and Tsaparas, P. (2015). “Selecting shortcuts for a smaller world.” SIAM International Conference on Data Mining 2015, Vancouver, 28–36.

Parotsidis, N., Pitoura, E., and Tsaparas, P. (2016). “Centrality-aware link recommendations.” Proceedings of the 9th ACM International Conference on Web Search and Data Mining, San Francisco, 503–512.

Paterson, J., and Ombuki-Berman, B. (2020). "A Hybrid Approach to Network Robustness Optimization using Edge Rewiring and Edge Addition." IEEE International Conference on Systems, Man, and Cybernetics, IEEE, 4051–4057.

Pell, R., Svoboda, R., Eagar, R., Ondko, P., and Kirschnick, F. (2015). Effective Infrastructure Asset Management- A holistic approach to transformation. Arthur D. Little - Prism, Hong Kong.

Pizzuti, C., and Socievole, A. (2018). "A genetic algorithm for improving robustness of complex networks." IEEE 30th International Conference on Tools with Artificial Intelligence, IEEE, Volos, Greece, 514–521.

Prajogo, D., Toy, J., Bhattacharya, A., Oke, A., and Cheng, T. C. E. (2018). "The relationships between information management, process management and operational performance: Internal and external contexts." International Journal of Production Economics, Elsevier Ltd, 199, 95–103.

Ross, R. (2019). Reliability Analysis for Asset Management of Electric Power Grids. John Wiley & Sons Ltd., Singapore.

Thai, M. T., and Pardalos, P. M. (2012). Handbook of Optimization in Complex Networks: Communication and Social Networks. (M. T. Thai, P. M. Pardalos, and Laboratory of Algorithms and Technologies for Networks Analysis, eds.), Springer Science+Business Media, New York.

Uddin, W., Hudson, W. R., and Haas, R. (2013). Public Infrastructure Asset Management. McGraw-Hill Education, New York.

Vishnu, C. R., Das, S. P., Sridharan, R., Ram Kumar, P. N., and Narahari, N. S. (2021). "Development of a reliable and flexible supply chain network design model: a genetic algorithm based approach." International Journal of Production Research, Taylor & Francis, 59(20), 6185–6209.

Woodhouse, J. (2014). "Standards in asset management: PAS 55 to ISO 55000." Infrastructure Asset Management, 1(3), 57–59.

Xerri, M. J., Nelson, S., and Brunetto, Y. (2015). "Importance of Workplace Relationships and Attitudes toward Organizational Change in Engineering Asset-Management Organizations." Journal of Management in Engineering, 31(5), 04014074.

Yosri, A., Elleathy, Y., Hassini, S., and El-Dakhakhni, W. (2021). "Genetic Algorithm-Markovian Model for Predictive Bridge Asset Management." Journal of Bridge Engineering, 26(8), 04021052.

Zhao, L., Li, B., Chen, H., and Yao, Y. (2018). "An assembly sequence optimization oriented small world networks genetic algorithm and case study." Assembly Automation, 38(4), 387–397.

Chapter 5

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

5.1. SUMMARY

Implementing and operating an asset management (AM) system within an organization is critical to address the many challenges associated with owning, managing, and operating infrastructure assets. An AM system was modeled as a network of connected nodes and links. Modeling the AM system network enabled the identification of systemic risks (i.e., dependence induced failures) within the AM system specifically associated with information asymmetry and information overload. Once identified, the systemic risks were mitigated by applying descriptive analytics to display only the necessary information to AM stakeholders, ensuring that they would not become overloaded with unnecessary information. Additionally, predictive analytics was employed to forecast a resilience key performance indicator (KPI) as a method to enable AM stakeholders to make decisions using consistent forecasted information. Finally, systemic risks within the AM system network were mitigated by employing prescriptive analytics to develop optimal configurations of the AM network links through minimizing the criticality of dependence-induced failures of important AM system components. Critical AM system components were defined using the ISO 55001 standard and the optional link configurations were evaluated using a genetic algorithm approach. The described descriptive and prescriptive analytics strategies mitigated the AM system exposure to systemic risks related to information asymmetry and information overload and the predictive analytics strategy enhanced the resilience response of a power utility to forced outage occurrences.

5.2. CONCLUSIONS AND CONTRIBUTIONS

The AM system network model presented in this thesis has expanded the understanding of the criticality of specific components necessary for operation of an AM system within organizations that own, manage, and operate infrastructure assets. The research presented in this thesis assessed systemic risks induced by information asymmetry and information overload and presented mitigation strategies specific to such risks using descriptive and prescriptive analytics. In addition, this research presented a predictive analytics approach to enhance the resilience of a power utility to forced outage occurrences. Within this context, the following conclusions are described as they relate to the application of descriptive- and prescriptive analytics to mitigate the systemic risks caused by information asymmetry and information overload and predictive analytics to enhance the resilience.

5.2.1. DESCRIPTIVE ANALYTICS STRATEGY FOR SYSTEMIC RISK MITIGATION

Previous research identified that the main challenges with implementing and operating an AM system within an organization included information asymmetry due to AM stakeholders using different or not consistent information to guide decision-making, and information overload where AM stakeholders were overwhelmed with data which caused challenges in identifying the correct data needed for decision-making. The work presented in this thesis described these challenges as causes for systemic risks (i.e., dependence-induced failures) within an AM system. The AM system was modeled as a network of connected components to identify the AM subject areas (i.e., AM system components) that were most exposed to such systemic risks. Different centrality measures (i.e., betweenness-, closeness-, degree-, eigenvector-centrality, and vulnerability index) were employed to assess the varying criticality of each AM subject area and its exposure to systemic risks. The AM subject areas *Strategic Planning* and *Lifecycle Value Realization* were found to be among the most critical AM subject areas exposed to systemic risk according to each of the five previously identified centrality measures. The AM subject area *Resource Management* was found to be critical among four of the described centrality measures also indicating its critical exposure to systemic risks. These AM subject areas are particularly susceptible to initiating a cascading failure scenario throughout the AM system network. Therefore, the thesis also describes methods for mitigating the exposure to systemic risks which could cause cascading impacts throughout the AM system.

Following the identification of critical systemic risk-exposed AM subject areas, Chapter 2 outlined the descriptive analytics approach that could be used to reduce the exposure of AM stakeholders to information overload. Tailored dashboard visualizations were outlined that presented KPIs specific to each of the critical AM subject areas. By presenting only the information necessary to make decisions related to that AM subject area, an AM stakeholder would not be overwhelmed with access to all available information related to the AM system. These dashboards focus the attention of AM stakeholders and ensure that each AM stakeholder that makes decisions related to an AM subject area uses consistent information. Applying descriptive analytics as described ensures consistent information is used for decision-making across AM subject areas, minimizing the exposure of AM stakeholders to information asymmetry where different or not current information is used for decision-making. Therefore, by deploying descriptive analytics targeted to each critical AM subject area, the potential for systemic risks related to information asymmetry and information overload are reduced.

5.2.2. PREDICTIVE ANALYTICS STRATEGY FOR RESILIENCE ENHANCEMENT

A predictive analytics strategy was developed to forecast resilience KPI metrics ensuring that there were consistent future predictions employed by all AM stakeholders across an organization. Practically, the resilience KPI metric of rapidity was used as the duration of forced outages for power infrastructure asset components. A methodology was presented to allow an organization that owns, operates, and manages power infrastructure assets to predict a specified KPI metric using a consistent data source. The study demonstrated that the described methodology was successful when using transmission line forced outage data to classify the rapidity of an outage soon after its occurrence. Having an accurate estimation of the rapidity of an outage enables a utility to rapidly respond to a forced outage with appropriate resources, therefore enhancing the resilience. By employing a structured methodology for KPI prediction, it ensures that the forecasted values are consistent across an organization's AM subject areas, therefore minimizing the potential of AM stakeholders using different future KPI metric values and in turn reducing the potential for information asymmetry. In addition, the predictive analytics methodology, presented in Chapter 3, only employs data that is relevant to developing the machine learning model for the KPI feature value to be forecasted, ensuring information overload does not occur.

5.2.3. PRESCRIPTIVE ANALYTICS STRATEGY FOR SYSTEMIC RISK MITIGATION

Finally, prescriptive analytics strategies were employed to develop an optimized configuration of the AM system network through the addition of links to reduce the exposure of systemic risk-critical AM subject areas. ISO 55001 was used to indicate the functionality of each AM subject area to the performance of the AM system. Using ISO 55001 allowed for the criticality of the AM system functionality to be evaluated while optimal network configurations were obtained to reduce the exposure of the AM system to systemic risks. Links were added between AM subject areas to improve the resilience of the AM network to cascading failure impacts due to information asymmetry and information overload. The added links provided additional paths for information to flow throughout the AM system network, therefore reducing the impact to the AM system if a critical AM subject area were to become dysfunctional due to information asymmetry or information overload.

5.3. RECOMMENDATIONS FOR FUTURE RESEARCH

This section outlines additional research that could be conducted to build on the work presented in this thesis.

1) Identify systemic risks for organizations at different stages of their AM journey. Not all organizations that own, manage, and operate infrastructure assets have a fully developed AM system as described in this thesis. Therefore, future research could explore organizations at different stages of AM system development and provide an optimal path for building out their AM system functionality. Such different AM implementation stages would require interviews with organizations to grade a scale to which they feel their organization is effective in each of the 39 AM subject areas and what connections exist between them. If organizations do not have all 39 AM subject areas developed, then this research would propose an optimal plan for improving their AM system by adding the remaining AM subject areas to enable organizations that do not have a fully developed AM system to

have a standardized plan which they could follow to improve their AM system. Having a standardized plan for AM system development would enable organizations among all infrastructure industries to be consistent in their AM practices, therefore enabling a regulator of AM to evaluate the effectiveness of AM system performance more efficiently across all infrastructure industries.

2) The AM system network should be integrated with infrastructure networks, enabling a multi-layered network approach investigating the impact of decisions within the AM system on the performance of the assets. In addition, different owners, levels of government, regulators, and political decisions should be incorporated into such an integrated approach. This could be accomplished using an agent-based modeling approach where decisions within the AM system could be simulated on the infrastructure network. Osman (2012) introduced this concept by modeling a road infrastructure network and including maintenance decisions. The proposed model would include all aspects of an AM system as shown in this thesis and incorporate their effects on the performance of the infrastructure network directly. This would enhance the capability of predictive modeling for investigating the effects of decisions on infrastructure performance in the future.

3) Incorporate the other resilience metrics, resourcefulness, robustness, and redundancy, into AM decision-making. Resource strategy and resource

management were found to be critical components of an AM system and incorporating the concept of resourcefulness as presented in Chapter 3 into the resource strategy and management could benefit infrastructure asset owners in bouncing back following disruptive events. This could include optimizing the number of spare resources available for each asset component class. This is especially important given the supply chain issues that plagued the world in 2021 and the increase in disruptive weather events causing catastrophic damage. Having the resources both in terms of available people and infrastructure components to restore service is crucial to maintain critical infrastructure functionality. Incorporating a balance between each of the resilience metrics would enable AM decision-making to adapt to the changing environment (e.g., climate, supply chain) in which infrastructure assets operate within.

5.4. References

Osman, H. (2012). “Agent-based simulation of urban infrastructure asset management activities.” *Automation in Construction*, Elsevier B.V., 28, 45–57.

5.5. Acronyms

AM	Asset Management
KPI	Key Performance Indicator
ISO	International Organization for Standardization