Financial hedging trends in megaprojects: A Topical Network Analysis of the infrastructure literature

Research Problem and Purpose

Infrastructure projects, as long-term government contracts, play a pivotal role in providing essential services to society and fostering economic development through the expansion of infrastructure capacity. However, the inherent benefits of these projects are accompanied by significant uncertainties throughout their lifecycle. To address these challenges, governments, constrained by public budgets, increasingly advocate for active private sector participation and finance, resulting in a complex interplay of interests. Consequently, there is an ongoing imperative to effectively manage capital infrastructure projects, primarily from a financial standpoint, encompassing megaprojects and Public-Private Partnerships (PPPs).

Although considerable research has been dedicated to defining and proposing mechanisms for financial risk management in infrastructure, existing literature has not provided a full understanding of the dynamics of risk in the infrastructure management field. Exploratory research has primarily taken two forms: an analysis of financial risk management literature (Owolabi et al., 2020; Ullah et al., 2022; Woodward, 1995; Xiong et al., 2019) and the use of case studies (Farmahini Farahani et al., 2022; Karimi & Piroozfar, 2022; Moorhead et al., 2022). However, these approaches are fragmented and do not depict the complete situation in the field. Also, the current volume of data generated intensifies the time required for information handling. Acknowledging this limitation, data mining has emerged in various domains as an Artificial Intelligence (AI) aid to process the exponentially increasing information.

In recent years, there has been a growing interest in the application of AI in the Construction Management (CM) field. However, there remains a need to adapt AI capabilities to CM for efficient and in-depth analysis processes. AI, emulating human rationality through machine automation, involves Deep Learning (DL) and Machine Learning (ML) as its constituent systems. To harness the capabilities of AI, many fields have witnessed advances in data mining. Specifically, Natural Language Processing (NLP) has emerged as an AI technique that translates human language into data used in ML and DL processes. NLP employs text strings and words, allowing the automation of language through a representation of terms, meanings, and grammatical properties (Lee JeeHee et al., 2019).

As identified by Wu et al. (2022), the application of NLP encompasses approaches to understanding domain knowledge formation, with Topic Representation and Modeling (TM) uncovering hidden data structures in texts, and Network Analysis (NA) facilitating relationship exploration between concepts. TM has found applications in document classification, transfer learning, and qualitative research, providing insights into information clusters and text summarization (Park et al., 2019). TM involves a probabilistic clustering of words based on their semantic coherence, commonly used as an unsupervised tool (Zhu et al., 2022). Despite, being useful when managing vast amounts of data, TM cannot provide a clear understanding of the relations that occur between obtained topics. This limitation highlights the need for an approach that allows the analysis of relations between clusters in CM. NA, a valuable tool for evaluating word relationships in text, reveals subgroups and semantic structures (Park et al., 2019). This approach facilitates the identification of highly

connected nodes and links governing network behavior. NA has been extended in literature review analyses (Castelblanco et al., 2021; Machete & Marques, 2021; Shadiyar et al., 2020), providing relevant insights into the specific needs of CM research and aiding in envisaging relationships between thematic clusters, thus complementing TM in NLP.

This work proposes a literature review on financial risk in the CM infrastructure field using NLP extensions to develop a domain knowledge discovery formation approach (Chen et al., 2009; Sterman, 1985). A semantic analysis is performed by applying TM to case studies. This information serves as input for an NA employing a Separable Temporal Exponential Random Graph Model (STERGM) (Krivitsky et al., 2011), we analyze the stochastic probability of tie formation and dissolution in a network over time. Through a relationship exploration in the domain (Wu et al., 2022), the NA approach aims to provide relevant insights into the specific avenues of CM research. Particularly, NLP tools can aid in visualizing relationships between thematic clusters addressed in this domain review. Thus, this work also demonstrates an application of NLP tools to explore research trends in CM.

In summary, this literature review aims to leverage the benefits of AI tools through NLP by conducting a state-of-the-art review on a relevant theme in CM. From a need in CM, this research analyzes financial mechanisms to effectively manage risks in infrastructure projects by developing a Case-Based Reasoning (CBR) (Aamodt & Plaza, 1994) literature review. This study uses information from journal articles related to CM, specifically megaprojects and PPPs.

Brief Research Methodology and Approach

This literature review employed the TM approach to identify keyword patterns within thematic clusters, establishing meaningful connections between keywords and published works. This innovative approach in CM helps to distill key insights from the vast literature. Having extracted this information, NA was used to explore the dynamics between various topics, also shedding light on emerging trends and predicting theoretical development. The process involves the following steps as integrated in Figure 1 and explained in their corresponding subsections.

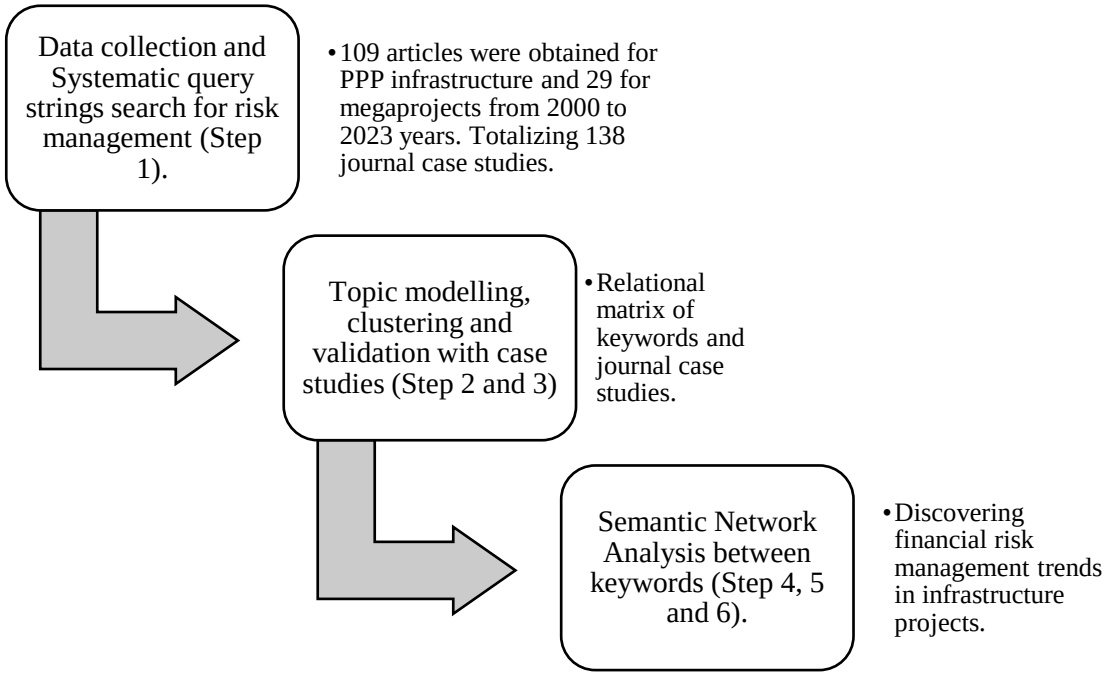

Figure 1. Methodological integration

Step 1: Data Collection and Systematic Search

Case studies were collected from academic journals in two categories of project scale: PPP and megaprojects. A systematic search was conducted on Scopus and Web of Sciences (WoS) databases, covering cases published between 2000 and 2023. The database selection followed similar procedures previously used in the literature (Castelblanco et al., 2021; Cui et al., 2018; Guevara Jose et al., 2020), i.e., employing a strings' search with the keywords as presented in Table 1.

Table 1. Query Strings for case studies in Financial risk management applied to infrastructure

Financial risk management in infrastructure projects	
PPP	**Megaprojects**
(PFI OR "PFI/PPP" OR "PPP/PFI" OR BOT OR "public private partnership*" OR "privatized infrastructure" OR "private infrastructure" OR "public/ private" OR "private finance" OR "build operate transfer" OR "privately financed" OR "build-operate-transfer" OR "public-private" OR "private finance initiative" OR "build/operate/transfer" OR toll OR "toll road" OR P3 OR "concession* toll") AND ("financial hedging" OR "financial coverage" OR "financial structure*" OR "guarantee*" OR "financial option*" OR "financial risk*" OR "financial risk management") AND "case study*"	("megaproject*" OR "mega project*" OR "complex project" OR "major project*" OR "giant project*" OR "large project*" OR "large-scale project") AND ("financial hedging" OR "financial coverage" OR "financial structure*" OR "guarantee*" OR "financial option*" OR "financial risk*" OR "financial risk management") AND "case study*"

As follows, Figure 2 summarizes the conducted procedure of the systematic search.

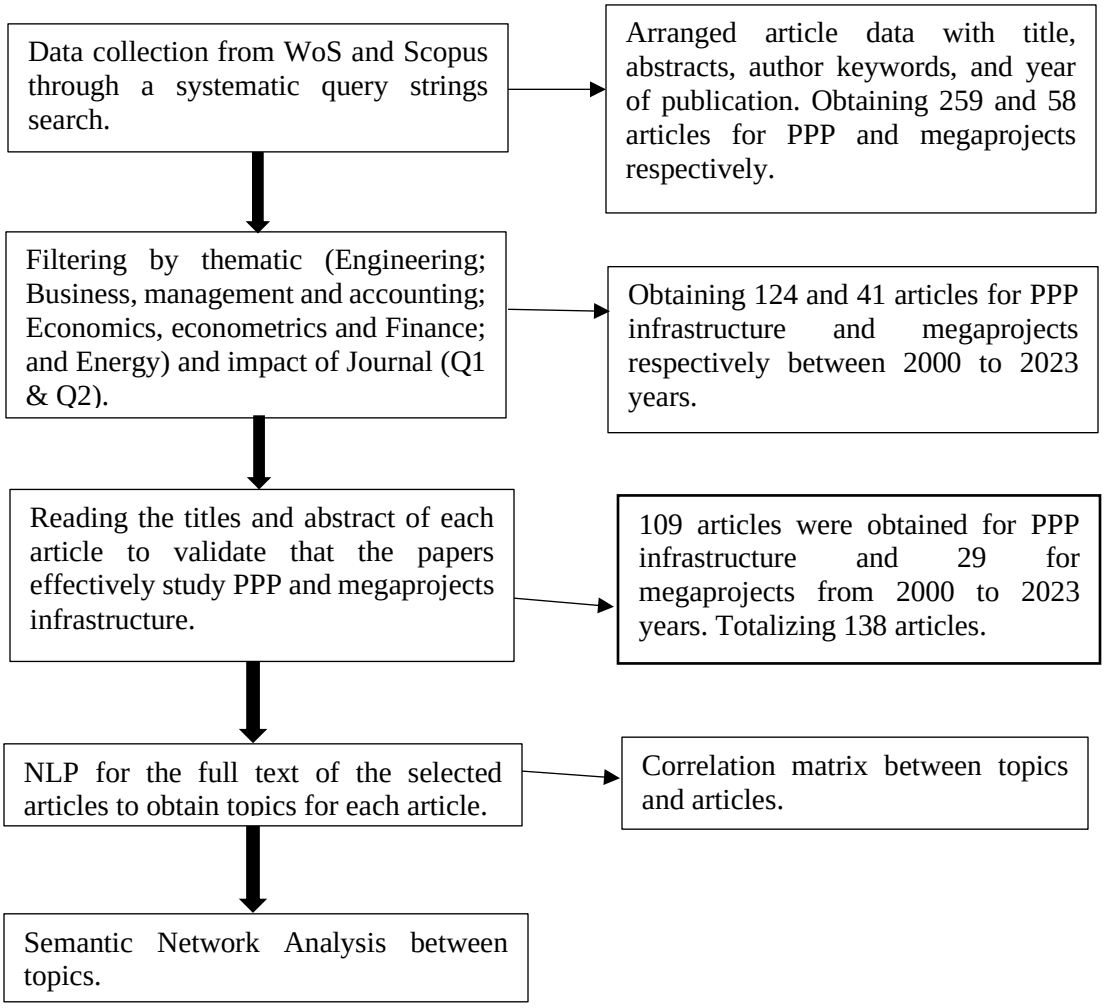

Figure 2. Systematic search procedure

Step 2: Topic Modeling and Clustering

Following the systematic search, two resulting datasets of journal case studies related to financial hedging in infrastructure projects are used. The full text of each case study is used as input data for the topic modeling process. The analysis utilizes a Bidirectional Encoder Representations from Transformers (BERT) approach developed by (Devlin et al., 2019). BERT processes text by establishing semantic relationships between words, thus creating clusters. One of the main features of BERT is the capability to obtain topics from a bag of words of the text. Such capability is a term frequency-inverse document frequency (TF-IDF). This TF is used to identify the differences between documents classified into clusters.

As presented in Figure 3 the phases performed by BERT consist of clustering through embeddings; construction of topics; and a summary of the obtained topics via their visualization.

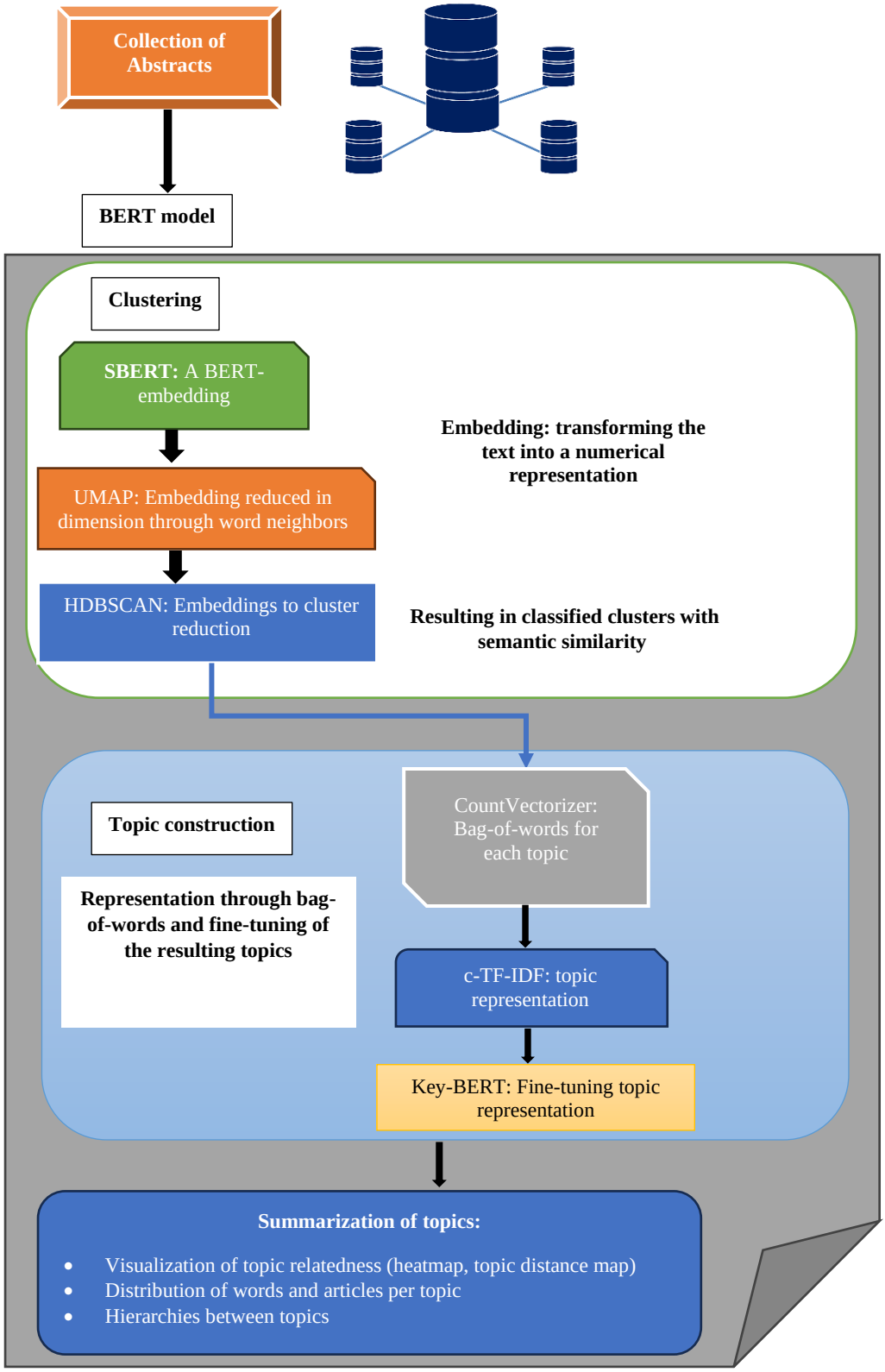

Figure 3. Workflow of BERT

Step 3: Validation of the Topic Modeling

Various statistical tests can validate the model, specifically, this study focuses on topic coherence and topic diversity (Grootendorst, 2022), thus, aligning with the Optimizing and Comparing Topic models is Simple! (OCTIS) framework proposed in the literature (Terragni et al., 2021), Topic coherence measures the relationship between top representative words within a topic, while topic diversity assesses the variation in top words compared to others. In Figure 4, a schematic representation of the framework steps is presented. The combination of fine-tuning method parameters allows optimization by enhancing the clustering quality.

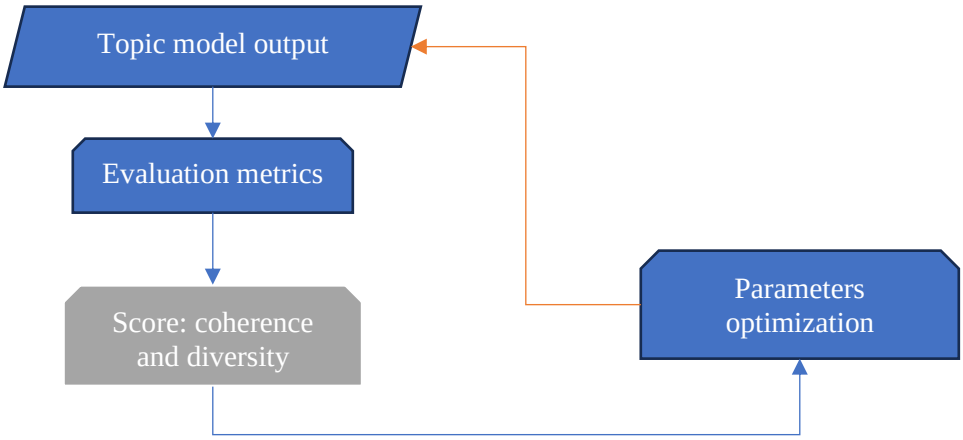

Figure 4. OCTIS framework

After validating the results, the data was processed in a correlation matrix between topics and their related articles.

Step 4: Network Analysis for Discovery Trends

NA is employed to model the interactions between topics. These topics are interrelated through weighted ties regarding the correlation of each topic with each journal case study. Their relations change over time. The aim of NA in this study is to gain insights into the formation and dissolution of ties between topics over time. For this purpose, a preparation of the data and the application of an STERGM enables an analysis of the stochastic probability of tie formation and dissolution in a network over time.

Step 5: Validation of the NA

To assess the formation and dissolution of ties in a STERGM, Krivitsky & Handcock (2014) recommend using Goodness of Fit as a validation metric. This metric refers compares the stochastic simulation of the network against the existent data using a confidence interval for the fitness assessment.

As a complement, the Jaccard coefficient (JC) evaluates the influence of variables to explain tie dissolution or formation in the network (Guerra et al., 2020). Therefore, JC is interpreted as a relation between tie formation and dissolution.

Step 6: Analysis of the theory trends

To generate an analysis of the theory trends in the addressed body of knowledge, a discussion on the insights provided by the STERGM was addressed. The analysis relied on a discussion along the materialization of the node formation and dissolution hypotheses based on the behavioral analysis in theory-formation research (Chen et al., 2009; Sterman, 1985). This discussion is based on the knowledge that the authors note in the relational metrics and behavior of the topic formation, and dissolution over time.

Key Findings

From this literature review, we unraveled the interlocks between topics in the study and the application of financial mechanisms for effective CM risk in the infrastructure domain. We reached a deeper understanding of theory trends in the field by discovering the research avenues in the field. NA allowed us to explain the theory formation in financial hedging practices for CM infrastructure. Also, Using NLP, we have optimized a process not possible with common approaches. In particular, we found that financial hedging mechanisms have been adapted from financial and economic trends to manage financial risks and uncertainty in PPP and megaprojects contracts (e.g., real options, equilibrium/bargaining process and transaction costs theoretical approaches).

For its part, government guarantees and subsidies are the most common mechanisms used by the public party in a twofold manner. Firstly, to attract private investment during the procurement phase. Secondly, as an exercisable guarantee to maintain the project viability and financial sustainability. Therefore, we develop insights into the theoretic trend of concepts related to these financial risk management mechanisms in PPP and megaprojects. As follows, Table 2 presents the main topics fostering knowledge in financial risk management, the concepts are ordered from the most influential including a brief definition of each one.

Table 2. Topics fostering knowledge in financial risk management

Most influential topics to foster knowledge in financial risk management
Finance-related factors: Mechanisms of hedging against risk management like government guarantees and subsidies. Also, it entails instruments to finance projects like bonds and betterment levies.
Contract management: Contract relational measures and practice to manage stakeholders.
Road infrastructure: This kind of project is the most commonly applied in infrastructure development and financial risk management.
Procurement-related factors: Procedures to mitigate financial risk during the procurement phase.
Contract performance: The measure of service during operation and construction.
Governance: The relational procedure of the government, as owner, on stakeholders' management.
Qualitative analysis: as a method of analysis for researching.
Quantitative analysis: as a method of analysis for researching.
Costs: related to the quantification and optimization of the budget.
Concessionaire: related to the private party's actions to manage financial risk.

Despite the benefits of implementing NLP, some limitations were found. The bag of words used to calibrate the NLP model contains concepts of several knowledge disciplines, however, TM might be enhanced if used a NLP model calibrated with only concepts of CM infrastructure. This assumption should be tested by coherence and diversity metrics. Regarding the test of NLP models, there is no consensus between scholars about the best suitable method of NLP accuracy, the most spread practice relies on the test validated by the developer of the model, this situation limits the comparison between methods because a lack in equivalent measures.

Implications

This study presented an integrated understanding of the mechanisms in financial risk CM through a robust methodology and a deep application of AI. Specifically, with NLP we extracted and analyzed large language data to theory trend exploration and discovery. In this sense, the contributions of this study are the insights on financial hedging mechanisms applied to CM infrastructure projects by conducting a knowledge discovery literature review. Additionally, we demonstrate the application of NLP tools to efficient data mining of topics in the literature, complemented with an exploration of the relationships between risk management topics in infrastructure through NA.

Overall, this study is helpful to scholars in the field through a structured procedure that takes advantage of AI capabilities to understand current practices and the formation of trending theories in the CM in the infrastructure body of knowledge. In this article we focus on PPP and megaprojects because of the high-risk features that poses these kinds of projects, however an extension of this work can combine NLP to vast data mining and NA to unravel trends in CM hot-topics.

References

Aamodt, A., & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications*, *7*(1), 39–59. https://doi.org/10.3233/AIC-1994-7104

Castelblanco, G., Guevara, J., Mesa, H., & Sanchez, A. (2021). Semantic Network Analysis of Literature on Public-Private Partnerships. *Journal of Construction Engineering and Management*, *147*(5), 04021033. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002041

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, *3*(3), 191–209. https://doi.org/10.1016/j.joi.2009.03.004

Cui, C., Liu, Y., Hope, A., & Wang, J. (2018). Review of studies on the public–private partnerships (PPP) for infrastructure projects. *International Journal of Project Management*, *36*(5), 773–794. https://doi.org/10.1016/j.ijproman.2018.03.004

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*.

Farmahini Farahani, A., Didehkhani, H., Khalili-Damghani, K., Sarfaraz, A. H., & Hajirezaie, M. (2022). A framework for interactive risk assessment in projects: Case study of oil and gas megaprojects in presence of sanctions. *Journal of Modelling in Management*. https://doi.org/10.1108/JM2-07-2020-0180

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv Preprint arXiv:2203.05794*.

Guerra, A. M., Montes, F., Useche, A. F., Jaramillo, A. M., González, S. A., Meisel, J. D., Obando, C., Cardozo, V., Hunter, R. F., & Sarmiento, O. L. (2020). Effects of a Physical Activity Program Potentiated with ICTs on the Formation and Dissolution of Friendship Networks of Children in a Middle-Income Country. *International Journal of Environmental Research and Public Health*, *17*(16). https://doi.org/10.3390/ijerph17165796

Guevara Jose, Garvin Michael J., & Ghaffarzadegan Navid. (2020). The Forest and the Trees: A Systems Map of Governance Interdependencies in the Shaping Phase of Road Public–Private Partnerships. *Journal of Management in Engineering*, *36*(1), 04019031. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000726

Karimi, S., & Piroozfar, P. (2022). Exploring Causes of Delays in National Road and Highway Projects in Developing Construction Economy. *Journal of Engineering, Project, and Production Management*. https://doi.org/10.32738/JEPPM-2022-0013

Krivitsky, P. N., Handcock, M. S., & Morris, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology*, *8*(4), 319–339. https://doi.org/10.1016/j.stamet.2011.01.005

Lee JeeHee, Yi June-Seong, & Son JeongWook. (2019). Development of Automatic-Extraction Model of Poisonous Clauses in International Construction Contracts Using Rule-Based NLP. *Journal of Computing in Civil Engineering*, *33*(3), 04019003. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000807

Machete, I., & Marques, R. (2021). Financing the Water and Sanitation Sectors: A Hybrid Literature Review. *Infrastructures*, *6*(1). https://doi.org/10.3390/infrastructures6010009

Moorhead, M., Armitage, L., & Skitmore, M. (2022). Risk management processes used in determining project feasibility in the property development process early stages by Australia/New Zealand property developers. *Journal of Property Investment and Finance*. https://doi.org/10.1108/JPIF-08-2021-0071

Owolabi, H. A., Oyedele, L. O., Alaka, H. A., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2020). Critical Success Factors for Ensuring Bankable Completion Risk in PFI/PPP Megaprojects. *Journal of Management in Engineering*. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000717

Park, S., Chung, D., & Park, H. W. (2019). Analytical framework for evaluating digital diplomacy using network analysis and topic modeling: Comparing South Korea and Japan. *Information Processing & Management*, *56*(4), 1468–1483. https://doi.org/10.1016/j.ipm.2018.10.021

Shadiyar, A., Ban, H.-J., & Kim, H.-S. (2020). Extracting Key Drivers of Air Passenger's Experience and Satisfaction through Online Review Analysis. *Sustainability*, *12*(21). https://doi.org/10.3390/su12219188

Sterman, J. D. (1985). The growth of knowledge: Testing a theory of scientific revolutions with a formal model. *Technological Forecasting and Social Change*, *28*(2), 93–122. https://doi.org/10.1016/0040-1625(85)90009-5

Terragni, S., Fersini, E., Galuzzi, B. G., Tropeano, P., & Candelieri, A. (2021). OCTIS: Comparing and Optimizing Topic models is Simple! *Proceedings of the 16th Conference of the European Chapter of the Association for Computational*

Linguistics: System Demonstrations, 263–270. https://doi.org/10.18653/v1/2021.eacl-demos.31

Ullah, S., Mufti, N. A., Saleem, M. Q., Hussain, A., Lodhi, R. N., & Asad, R. (2022). Identification of factors affecting risk appetite of organizations in selection of mega construction projects. *Buildings*. https://doi.org/10.3390/buildings12010002

Woodward, D. G. (1995). Use of sensitivity analysis in build-own-operate-transfer project evaluation. *International Journal of Project Management*. https://doi.org/10.1016/0263-7863(95)00016-J

Wu, C., Li, X., Guo, Y., Wang, J., Ren, Z., Wang, M., & Yang, Z. (2022). Natural language processing for smart construction: Current status and future directions. *Automation in Construction*, *134*, 104059. https://doi.org/10.1016/j.autcon.2021.104059

Xiong, B., Skitmore, M., Xia, P., Ballesteros-Pérez, P., Ye, K., & Zhang, X. (2019). Impact of Corporate Credit Scoring on Construction Contractors in China. *Journal of Construction Engineering and Management*. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001631

Zhu, X., Pasch, T. J., Ahajjam, M. A., & Bergstrom, A. (2022). Environmental Monitoring for Arctic Resiliency and Sustainability: An Integrated Approach with Topic Modeling and Network Analysis. *Sustainability*, *14*(24). https://doi.org/10.3390/su142416493