Journal of Systems Thinking in Practice
JSTINP, *QUARTERLY*
Homepage: jstinp.um.ac.ir
Online ISSN: 2821-1669, Print ISSN: 2980-9460
Research Article
DOI: 10.22067/JSTINP.2023.84746.1077
JSTINP 2023; Vol. 2, No. 4

Topic Modeling on System Thinking Themes Using Latent Dirichlet Allocation, Non-Negative Matrix Factorization and BER Topic

Rouhollah Bagheri [a]*, Nahid Entezarian [a], Mohammad Hossein Sharifi [a]

[a] *Department of Management, Faculty of Economics and Administrative sciences, Ferdowsi University of Mashhad, Mashhad, Iran.*

How to cite this article

Bagheri, R., entezarian, N., Sharifi, M. H., 2023. Topic Modeling on System Thinking Themes Using Latent Dirichlet Allocation, Non-Negative Matrix Factorization and BERTopic. *Journal of Systems Thinking in Practice, 2*(4), pp.33-56. doi: 10.22067/jstinp.2023.84746.1077.
URL: https://jstinp.um.ac.ir/article_44582.html

ABSTRACT

In recent years, there has been a growing interest in Systems Thinking (ST) as a significant area of research. It has become increasingly crucial to provide a detailed overview of the ST domain and to identify the prevailing research focuses and trends within this realm. This study represents the most comprehensive and pioneering effort, using topic modeling analysis to analyze the landscape of ST research from the past to the present. The primary objective of this study was to identify the current state of research and the predominant areas of focus within articles related to ST. To achieve this research aim, a search was conducted on August 20, 2023, using the Scopus database, yielding 1400 articles. The bibliometric analysis findings of this study indicate a substantial surge in the number of publications in this field, especially since 2016, with a significant majority of these studies originating from the United States. While the research is characterized by its interdisciplinary nature, most publications fall within social science. Employing Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), and Bidirectional Encoder Representations (BER) Topic algorithms for topic modeling analysis, the study classified the articles into ten distinct topics. These topics encompass "comprehending and modeling complex systems," "sustainability in business," "interdisciplinary learning and problem-solving in education," "enhancing healthcare delivery," "system dynamics modeling," "engineering education," "chemistry education," "enhancing patient outcomes," "environmental sustainability," and "improving organizational performance." The most prominent topics that represent common research areas in the field of Systems Thinking include "system dynamics modeling," "enhancing healthcare delivery," "interdisciplinary learning and problem-solving in education," "comprehending and modeling complex systems," "environmental sustainability," and "improving organizational performance". In conclusion, this study is expected to provide valuable guidance for future research in the field of Systems Thinking by aiding in identifying research interests and trends.

Keywords

System Thinking, Topic Modeling, BER Topic, Latent Dirichlet Allocation, Non-Negative Matrix Factorization.
.

Article history

Received: 2023-10-11
Revised: 2023-12-03
Accepted: 2023-12-10
Published (Online): 2023-12-28

Number of Figures: 12 **Number of Tables: 3** **Number of Pages: 24** **Number of References: 61**

**Corresponding author*
Email: rbagheri@um.ac.ir

1. Introduction

Systems thinking is a multidisciplinary approach that emphasizes understanding the interconnectedness and interdependencies within complex systems (Hossain et al., 2020). It has evolved as a cognitive process (Grohs et al., 2018) and an essential framework for addressing complex problems and promoting holistic understanding. Systems thinking has garnered significant attention across various fields, including management, engineering, ecology, and social sciences (De Souza, 2022). The literature on systems thinking encompasses various topics, theories, and applications, reflecting the diverse perspectives and methodologies researchers and practitioners employ in this domain.

In recent years, topic modeling techniques have gained prominence as a method for uncovering latent themes and patterns within textual data. Prominent topic modeling approaches such as Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), and BERTopic have been utilized to extract and analyze themes from textual corpora. While these techniques have been widely applied in diverse domains, their specific application to systems thinking literature remains an area of exploration that holds promise for gaining deeper insights into the field's fundamental concepts and emerging trends.

This research seeks to investigate the application of LDA, NMF, and BERTopic in the context of systems thinking themes, aiming to uncover the salient topics, trends, and challenges within the literature. By employing these topic modeling techniques, this study sheds light on the underlying themes and dynamics that characterize systems thinking discourse, thereby contributing to a more nuanced understanding of this interdisciplinary domain. Therefore, the current research examines the following questions in the existing literature on systems thinking (ST):

1) What salient themes can be extracted from the ST investigations using LDA, NMF, and BERTopic based on textual analysis?
2) Given the evolution of ST literature over time, how do the topics derived from LDA, NMF, and BERTopic reflect the current and emerging trends in systems thinking literature, and how can better insights be gained?
3) What are the evolving trends and challenges in systems thinking research, and how do these trends manifest in the topics generated by LDA, NMF, and BERTopic?

Through the exploration of topic modeling on systems thinking themes, this research aims to address the following key objectives: to identify and analyze the prevalent themes within systems thinking literature, to compare and contrast the outcomes of LDA, NMF, and BERTopic in extracting these themes, and to discern the evolving trends and challenges in

systems thinking scholarship. This study seeks to provide valuable insights that can inform future research, practice, and argument in systems thinking by addressing these objectives. This research uncovers the prevalent themes associated with systems thinking by employing rigorous textual analysis and natural language techniques.

The key contribution of this research is its ability to reveal the current state of the field, research interests, and evolving trends within systems thinking. Applying advanced topic modeling techniques adds depth and nuance to the analysis, allowing for a more comprehensive literature assessment. As a result, the study provides valuable insights for researchers, practitioners, and policymakers, guiding future research directions and facilitating a deeper understanding of the multidisciplinary aspects of systems thinking. Overall, the main contribution of this research is its pioneering use of advanced topic modeling techniques to illuminate the complex and evolving landscape of systems thinking research.

The paper's structure is as follows: Following the Introduction, section 2 provides an introduction to the theoretical background, methods, and models employed. Section 3 details the research methodology and data collection process for topic modeling. Section 4 presents the findings of the descriptive analysis. Section 5 discusses the datasets, describes data pre-processing, and analyses the results. Section 6 encompasses the discussion, conclusion, and suggestions for future work.

2. Theoretical background

2.1. System thinking

Systems thinking emerged around 1960 as a framework for comprehending the dynamic behavior of complex systems, as linear cause-and-effect thinking became inadequate for generating comprehensive solutions (Ratter, 2012). Adams and Keating (2011) assert that grasping the principles of system theory, in conjunction with the thought process developed in systems thinking, is a fundamental step in comprehending complex systems. Systems thinking is a conceptual framework, a body of knowledge and tools developed over the past seventy years to make the full patterns more transparentand help us see how to change them effectively (Senge, 2006).

However, systems thinking is not merely a tool. However, it represents a broader worldview, a way of perceiving the world that evolves as an individual's capacity and willingness to consider it holistically grows. The disciplined use of systems thinking tools and skills (Nagahi et al., 2020) not only supports but can also potentially alter one's worldview, influencing the decisions made

regarding applying these tools (Shin et al., 2022). Building on the extensive history of applying systems thinking in business and its more recent integration in K-12 education (National Research Council, 2012), systems thinking can be defined in practical terms as the capacity to perceive a problem or phenomenon as a system comprised of interacting elements that produce emergent behaviors (Shin et al., 2022) presented the importance of different perspectives, definitions, and taxonomies of systems thinking developed over the years. Figure 1 exhibits the development of influential systems thinking perspectives, methods, and tools. The intent is not to provide an exhaustive list of scholars' works but to indicate the progress and expansion of the systems thinking domain.

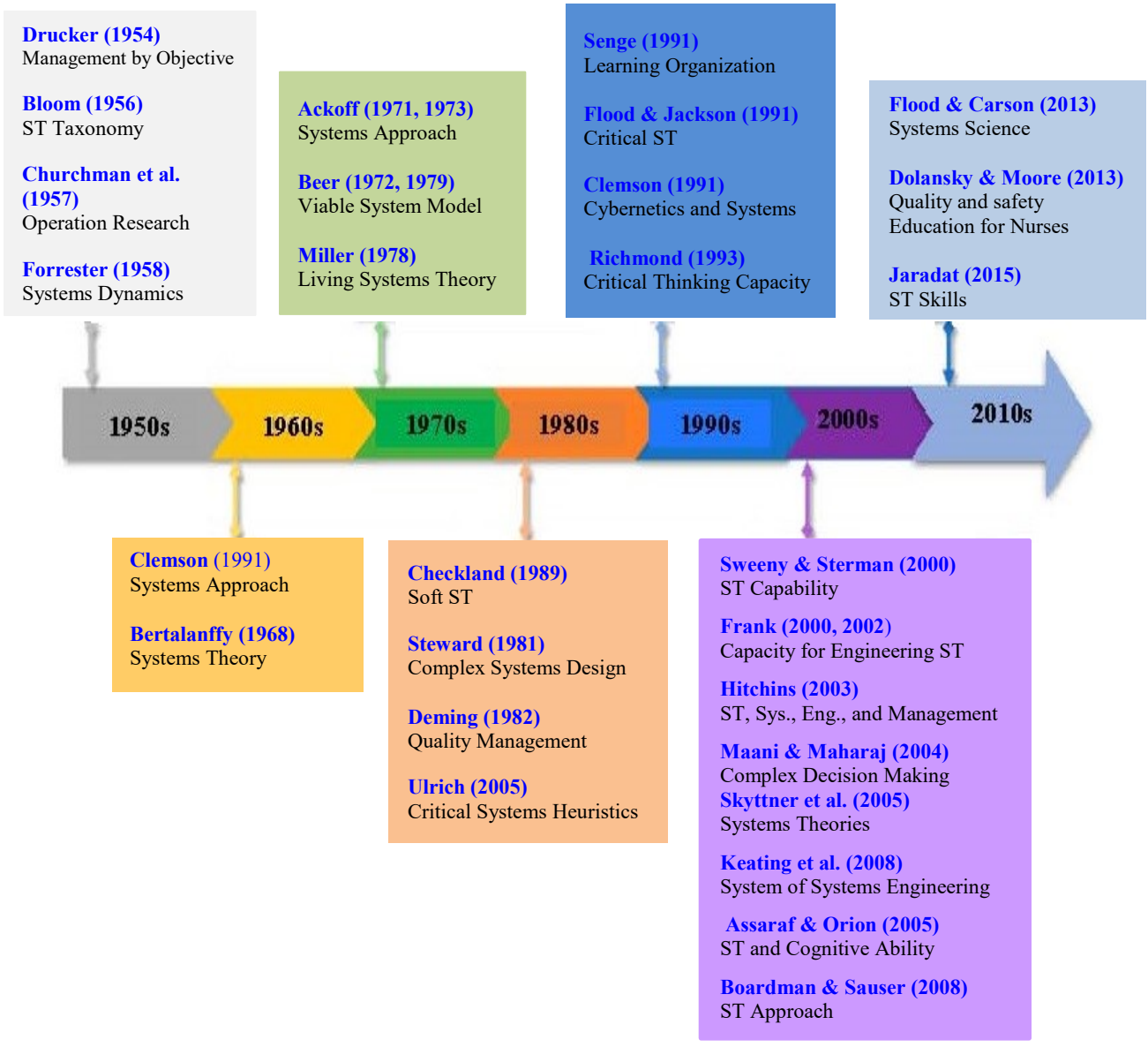


Figure 1. Top impacting systems thinking tools, methods and approaches over time (Source: Hossain et al.,, 2020).

 DOI: 10.22067/jstinp.2023.84746.1077

It is clear from Figure 1 that various perspectives, methods, tools, and approaches to systems thinking have evolved and advanced over the years. The primary focus of most of these influential research works has been on the complex nature of systems and the introduction of methods and approaches related to systems thinking.

Systems thinking is based on the understanding of interrelationships and emergent behaviors. It is often associated with theories, frameworks, and methodologies to understand complex situations and design interventions to address them. Papers published and conferences conducted in traditional systems forums have a bearing on theoretical developments, conceptual advancements, and methodological applications of ST. However, use cases of ST as a cognitive skill that can be applied without traditional systems frameworks and methodologies are rare (Chowdhury, 2023a).

The history of systems thinking can be divided into four distinct stages or waves. The first wave, which emerged in the post-World War II era, focused on the interconnectedness of social and organizational systems and aimed to manage complex problems. This wave, known as complicated systems thinking, was influenced by the belief that social reality could be improved and managed with a functionalist mindset. Methodologies associated with the first wave included Systems Analysis, Systems Engineering, System Dynamics, Organizational Cybernetics, and the Viable System Model. However, this wave faced criticism for its emphasis on prediction and control and its neglect of human agency (Chowdhury, 2023b).

The criticism of the first wave led to a significant paradigm shift and the emergence of the second wave, known as soft systems thinking. This wave, characterized by scholars Checkland (1981), emphasized interpersonal relationships, open dialogue, and learning. Methodologies associated with the second wave included Strategic Assumption Surfacing and Testing, Soft Systems Methodology, Interactive Planning, Interactive Management, and Structured Dialogical Design. However, the second wave also faced criticism for its inability to address power dynamics and hidden influences, leading to a growing fragmentation between hard and soft systems thinking.

To bridge this fragmentation, the third wave, critical systems thinking (CST), emerged with a focus on liberation and emancipation. This wave employed recent developments from complexity theory and emphasized the importance of human interaction and interrelationships. The foundational methodologies of critical systems thinking included Critical Systems Heuristics, Methodological Pluralism, and Systemic Intervention. This wave aimed to address the criticisms of the previous waves and provide a more comprehensive approach to

understanding and addressing complex systems (Chowdhury, 2023b). Criticizing the third wave, Jackson (2019) proposed the System of Systems Methodology (SOSM) as a framework for understanding the strengths and weaknesses of various systems methodologies.

An emerging fourth wave of systems thinking unifies and advances the field by identifying the DSRP as underlying the diversity of the three waves. The fourth wave of systems thinking introduces a new focus on simple rules to address complexity, universality, content agnosticism, and an emphasis on metacognition for deeper understanding and emotional intelligence. These elements enable us to directly confront VUCA (Volatility, Uncertainty, Complexity, and Ambiguity). The fourth wave achieves a newfound balance between the systems and the thinking aspects of systems thinking.

2.2. Topic modeling

Topic modeling is an unsupervised method used to uncover hidden topics within a collection of documents (Alami et al., 2021). It is employed to analyze textual documents and automatically extract their underlying themes (Sharma et al., 2021). Topic modeling can link words with the same context and differentiate between the uses of words with different meanings (Barde & Bainwad, 2017). These models aim to identify structural patterns in texts to extract meaningful information. The topic models create clusters of words representing the main subjects in the given collection, providing an automated way to identify common themes in the presented papers (Gurcan et al., 2021; Principe et al., 2021). In Topic modeling, a range of statistical and probabilistic approaches is used to estimate the probability of a particular sequence of words appearing in a sentence. Topic modeling analyzes large text corpora to form the basis for predicting words. These models are more efficient when compared to other methods because they consider the meaning and semantics of words and sentences, as well as the relationships between words.

Various approaches, including algorithms such as LDA, NMF, and BERTopic, can be used for performing topic modeling. Applying these algorithms will help uncover the hidden topic patterns in the text's metadata. However, it is imperative to acknowledge that each model has strengths and weaknesses, and the findings necessitate in-depth qualitative interpretation (Egger & Yu, 2022).

2.2.1. Latent dirichlet allocation (LDA)

LDA is Latent Dirichlet Allocation, a graphical probabilistic model for analyzing text. It

operates on a bag-of-words representation, treating each document as a compact numerical vector. LDA outlines its underlying generation process: a) It calculates the likelihood that a specific document is responsible for producing a particular topic. b) It determines the likelihood that a topic generates a specific word. This model involves drawing topic distributions from a Dirichlet distribution, where each element (or topic) in the sampled mix is independent of the others. This generative procedure defines a combined probability distribution covering visible documents and the concealed topic structure. LDA then endeavors to deduce these hidden topics based on the observable words (Abdelrazek et al., 2022). Previous works have shown that LDA approaches are beneficial and valuable in topic modeling (Jelodar et al., 2019). This paper investigated scholarly articles related to systems thinking (ST) using the LDA algorithm to discover the research development, current trends, and intellectual structure of systems thinking.

2.2.2. Non-negative matrix factorization (NMP)

Both LDA and NMF employ statistical modeling to uncover thematic patterns within a text collection, as Egger and Yu indicated in their 2022 study. Nevertheless, there is a notable distinction between the two. LDA is categorized as a generative statistical model, while NMF takes a decompositional, non-probabilistic approach rooted in matrix factorization, belonging to linear algebraic algorithms. Non-negative Matrix Factorization (NMF), introduced by Lee and Seung in 2000, is a widely employed tool for analyzing high-dimensional data. It excels at automatically extracting sparse, meaningful features from a set of non-negative data vectors. NMF serves as an unsupervised technique for reducing the dimensionality of non-negative matrices. A key advantage of NMF compared to LDA is its computational efficiency and scalability, as highlighted by Egger and Yu (2022). NMF is more inclined to produce higher-quality topics than LDA with the same experimental settings. Learning with NMF-based schemes is another effective method in short text topic mining in addition to the popular LDA-based paradigms (Chen et al., 2019).

2.2.3. Bidirectional encoder representations from transformers (BERT)

BERTopic is an unsupervised clustering-based technique for topic modeling. It utilizes Bidirectional Encoder Representations from Transformers (BERT) to create contextual sentence vector representations. These vector representations capture semantic information, labeling topics based on their contextual meaning. BERTopic operates as a topic modeling

algorithm by following a three-step process. First, it converts each document into an embedding representation using a pre-trained language model. Afterward, it reduces the dimensionality of these embeddings to improve the clustering process. Finally, it extracts topic representations from the document clusters using a customized class-based variation of TF-IDF, as introduced by Grootendorst (2022). The main advantage of BERTopic is that it allocates clusters precisely and gives topic name suggestions based on clusters. Topic modeling based on BERT does not require the number of topics to be specified in advance (Egger & Yu, 2022).

3. Research methodology

The research objective is to utilize NLP and Topic modeling methods in research papers within the systems thinking field. The proposed approach consists of stages for identifying topical interests, depicted in Figure 2. These stages are executed step-by-step, encompassing data preparation, pre-processing, and feature extraction. Subsequently, three distinct topic modeling techniques are introduced: LDA, NMF, and BERTopic. Finally, the model's performance is evaluated using topic evaluation metrics and validating their coherence. The proposed approach is elaborated in the following sections.

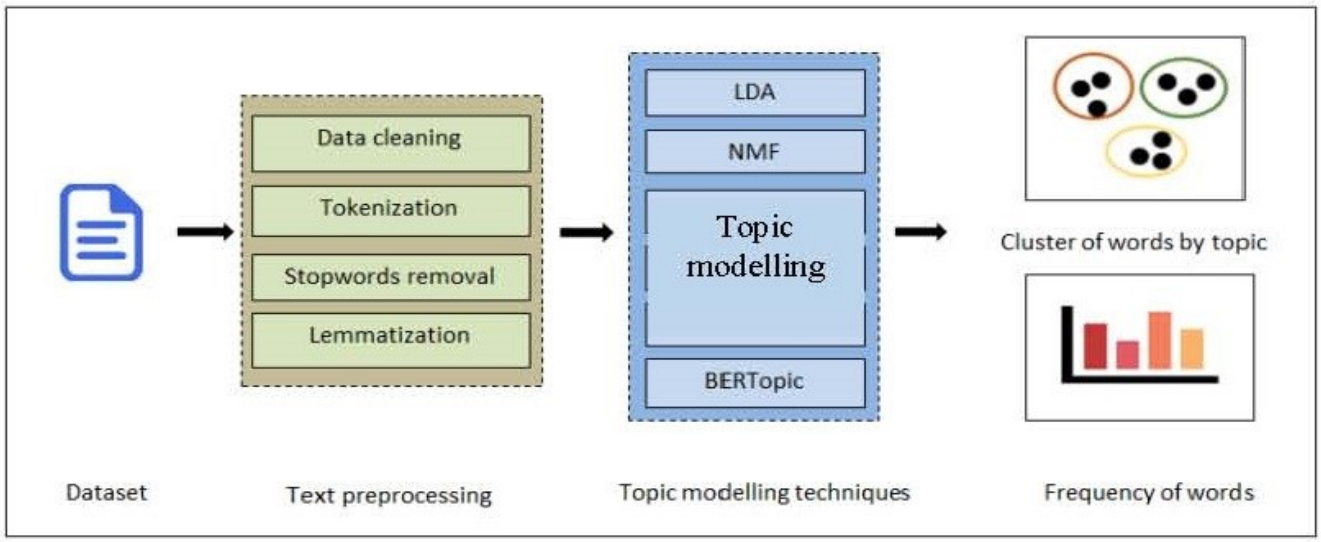


Figure 2. Overview of topic modelling schema using LDA, NMF, and BERTopic (Krishnan, 2023)

3.1. Data collection

About the database, a total of 1,400 documents encompassing articles, conference papers, and book chapters from the Scopus database are obtained. The search criteria were based on identifying the terms 'systems thinking' and 'ST' within the papers' metadata. Specifically, information such as the abstract, keywords, year, source, country/territory, document type,

subject area, open-access status, and the language of the document, which is English, is focused. Articles from the Scopus database are filtered and extracted based on their document type (article), document language (English), and document topic (systems thinking).

The topic modeling algorithm is applied for these documents, including LDA, NMF, and BERTopic. This algorithm was executed using the Python version 3.9.

3.2. Data pre-processing

Before conducting frequency analysis and topic modeling algorithms, the data was ready by performing actions such as eliminating stop words, tokenizing, lemmatizing, and stemming terms. This preparatory phase aims to remove less meaningful terms and prevent the counting of similar words in different verb forms more than once (Blum & Raviv, 2023). Data pre-processing is important in enhancing and purifying the raw data, ensuring that it includes only essential attributes for the topic modeling task (Krishnan, 2023).

As shown in Figure 2, different stages of text pre-processing include Tokenization, Lowercasing, Stop Words removal, stemming, and lemmatization. These text pre-processing steps were widely used for dimensionality reduction (Mursi et al., 2023).

Normalization of the case: This pre-processing technique involves converting all the text of the articles to lowercase or uppercase letters and merging exact words to reduce dimensions.

Stop Words Removal: Stop words are common words in a language often considered unnecessary in text mining programs. These words include pronouns, prepositions, articles, and auxiliary verbs that were removed from the text of the articles. Removing stop words, stemming, and term frequency can significantly affect the classification output (Mursi et al., 2023).

Tokenization: This pre-processing technique involves dividing the text of the articles into discrete and recognizable elements, such as words, phrases, symbols, or complete sentences, to work more effectively on the text. Depending on the domain and language, tokenization can significantly improve classification accuracy.

Stemming: In the stemming process, all suffixes and prefixes are removed from the words, reducing different word forms that convey the same meaning to their root or stem. The quality of a model depends on how well stemming can correctly match different forms of a word to the same root (Mursi et al., 2023).

Lemmatization: In the Lemmatization technique, parts of speech are identified, and inflectional forms are associated with their corresponding lemmas. The lemmatization process involves replacing a given token with its corresponding lemma (Mursi et al., 2023).

4. Descriptive analysis findings

The graph in Figure 3 illustrates the distribution and proportions of publications related to the field of ST over different years. Figure 2 1400 articles are included in this category, comprising 1326 research articles and 74 review articles published across 121 journals. Initially, there was no clear pattern in the earlier years. However, it is noteworthy that there has been a substantial annual growth in the number of articles, particularly after 2016, indicating a consistent linear increase over time.

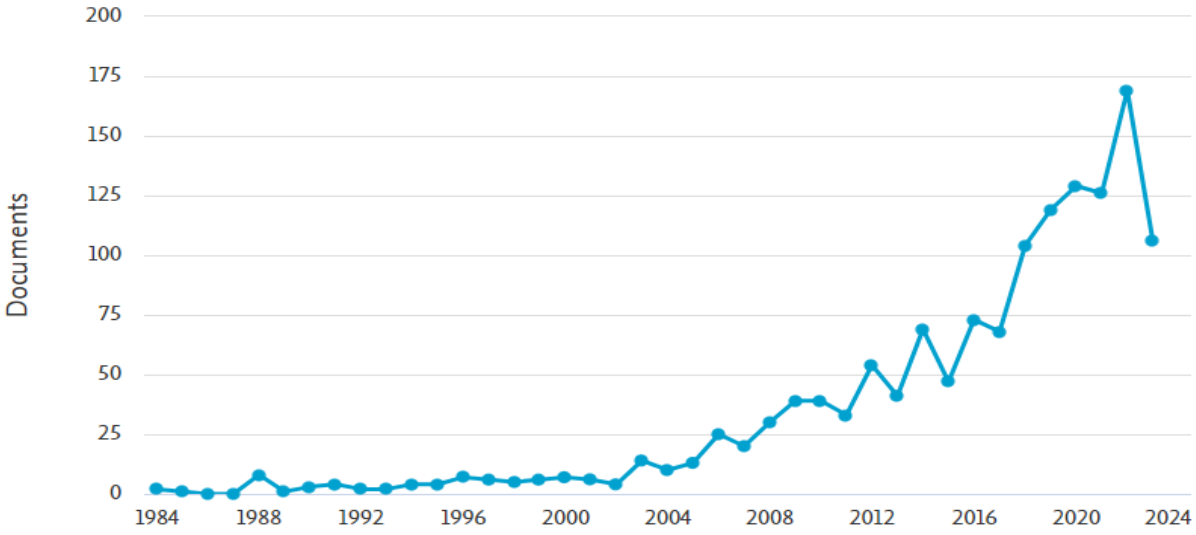


Figure 3. Distribution of articles in the field of ST by years

Considering the sources of publication, it is observed that the top 10 countries with the highest publication rates contribute to 84% of the total publications. In simpler terms, nearly four-fifths of the 1400 publications are generated by these ten countries, with the "United States," "United Kingdom," and "Australia" leading the way in this domain (refer to Figure 4).

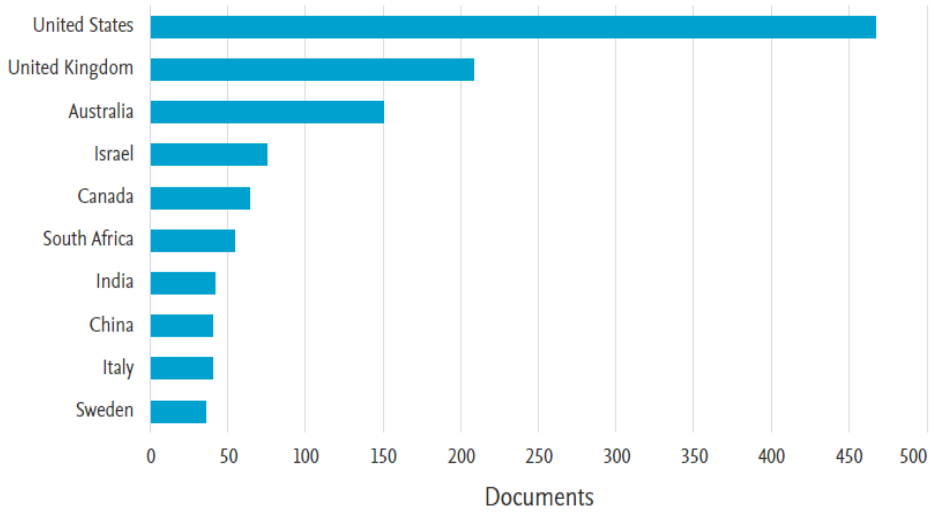


Figure 4. Prominent countries in articles in the field of ST

The findings regarding prominent areas of study in the context of systems thinking (ST) articles are presented in Figure 5, which illustrates the primary topic areas featuring the most articles. The majority of these articles are concentrated in the field of social sciences.

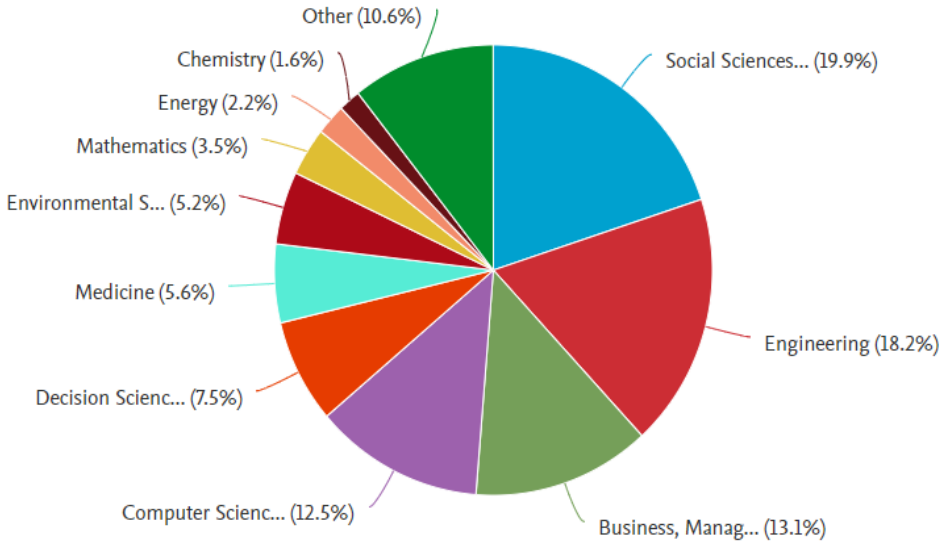


Figure 5. Top most-published subject areas and percent of publications

5. Implementation of topic models

5.1. Results of LDA topic modeling

LDA-based topic models typically require a predetermined selection of the number of topics as input to the model. For evaluating the effectiveness of an LDA topic model, a coherence score estimates how effectively the aggregate set of words can characterize a given topic it comprises (Chen et al., 2019). The LDA model employs the collapsed Bayes sampling method

in Gensim. As shown in Figure 6, our LDA topic model with four topics achieves a significant coherence score exceeding 0.36.

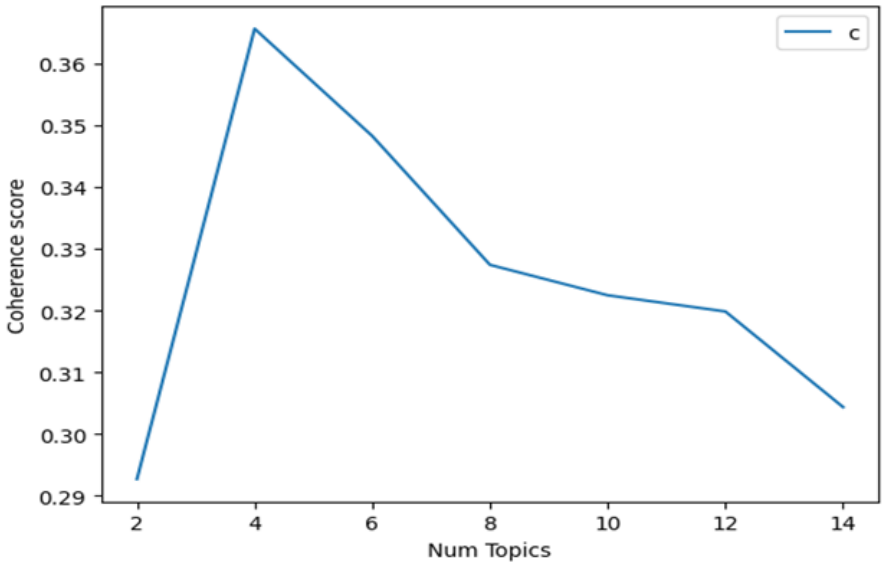


Figure 6. Coherence score of the LDA topic model for numbers of topics.

This section presents findings derived from an analysis of published research on Systems Thinking (ST). The main goal of using topic modeling is to identify the most commonly discussed themes within a body of literature rather than conducting a comprehensive literature review, as highlighted by Gurcan et al. in 2021. Of the web-based interactive visualization tools, pyLDAvis by Genism is the most commonly used tool for depicting the information embedded within a topic model, as noted by Islam (2019). This tool enables the visualization of topic distributions across a given text corpus and helps establish connections between the topics and the associated documents (research papers).

The present study employs an LDA model to ascertain four distinct topics and their distribution across the documents. In Figure 7, each bubble in the left-hand plot represents a specific topic, with the size of the bubble indicating its prevalence. An effective topic model is characterized by having reasonably large, non-overlapping bubbles distributed across the chart rather than clustering in a single area. On the right-hand side, words represent the most significant keywords. Placing the cursor over one of the bubbles updates the words and bars on the right-hand side to show the top 30 salient keywords that form the selected topic, along with their estimated term frequencies, as described by Islam (2019). These words are closely linked to research in the field of Systems Thinking and enable us to identify four distinct topics based on their semantic content, effectively outlining subfields of research in the domain of Systems

Thinking. Additionally, Table 1 presents the keywords generated by the LDA analysis for a selection of 1400 articles focusing on the application of systems thinking.

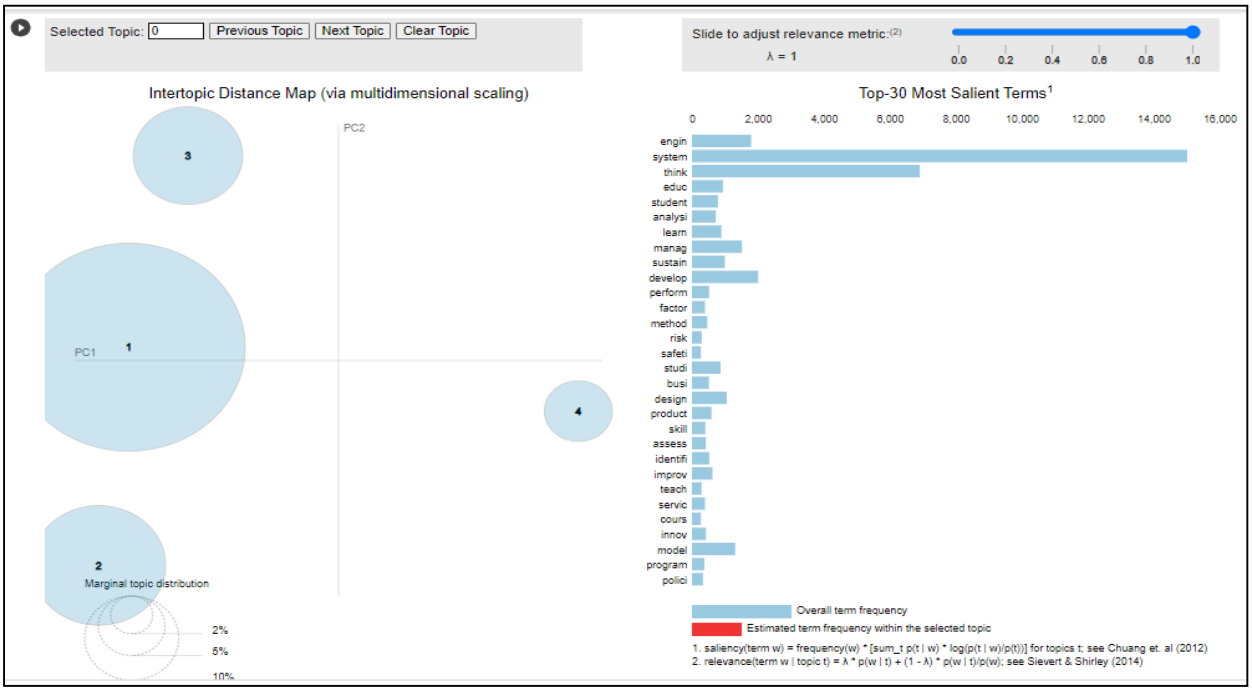


Figure 7. Visual inspection of LDA

Table 1. Sample topics generated by LDA

Topic classification	Keywords	Articles (N=1400), n (%)
Topic 1: understanding and modeling complex systems	System, think, complex, approach, develop, model, understand ,research, critic, practice, theory, paper, inform, process, manag, problem, organ, methodology, design, knowledge, concept, work, analysis, differ, social, , appli, provid, applic, human, dynam	855 (61.1%)
Topic 2: business sustainability	Sustain, manag, system, develop, approach, product, busi, service, innov, polici, industri, model, environment, implement, think, process, improve, oper, effect, chang, casual- loop, impact, public, stakehold, plan, research, dynam, strategi, project, identifi	283 (20.2%)
Topic 3: interdisciplinary learning and problem-solving in education	engine, system, think, educ, student, learn, develop, skill, design, teach, course, program, studi, profession, project, assess, scienc,tool, univers, technolog, abil, Particip, knowledge, sustain, teacher, need, active, experi, train, integr	189 (13.5%)
Topic 4: improve healthcare delivery	System, safeti, risk, analysis, think, factor, accid, measur, nurs, assess, method, studi, identifi, manag, simul, improve, event, paitent, hospital, model, control, report, level, result, survey, associ, medic, standard, construct	73 (5.2%)

Undoubtedly, the most prominent and widely discussed subject is "comprehending and modeling intricate systems". This theme is the focus of approximately 61.1% of the studies. This area of focus delves into the challenges associated with understanding and modeling intricate and ever-changing systems, such as conceptual models, soft systems methodology,

ecological systems, social systems, and economic systems. The second most prevalent theme, which appears in 20.2% of the selected papers, is 'business sustainability". This subject is instrumental in fostering the creation of sustainable products and services, enabling businesses to grasp the intricate interactions among various components of the system and their contributions to environmental issues. It also aids businesses in identifying innovation opportunities and designing approaches for system-level transformation. The subsequent theme, "interdisciplinary learning and problem-solving in education," constitutes 13.5% of the papers. This topic explores how systems thinking can be harnessed to teach critical thinking, instructional leadership, problem-solving, and interdisciplinary learning. Finally, "enhancing healthcare delivery" is the subject of 5.2% of the selected papers. This topic examines how systems thinking can be utilized to enhance healthcare delivery, manage safety, improve patient outcomes, and enhance population health. Figure 8 presents a visual representation of the distribution of these topics over the years and shows the relative changes in their popularity over time, as determined by Latent Dirichlet Allocation (LDA).

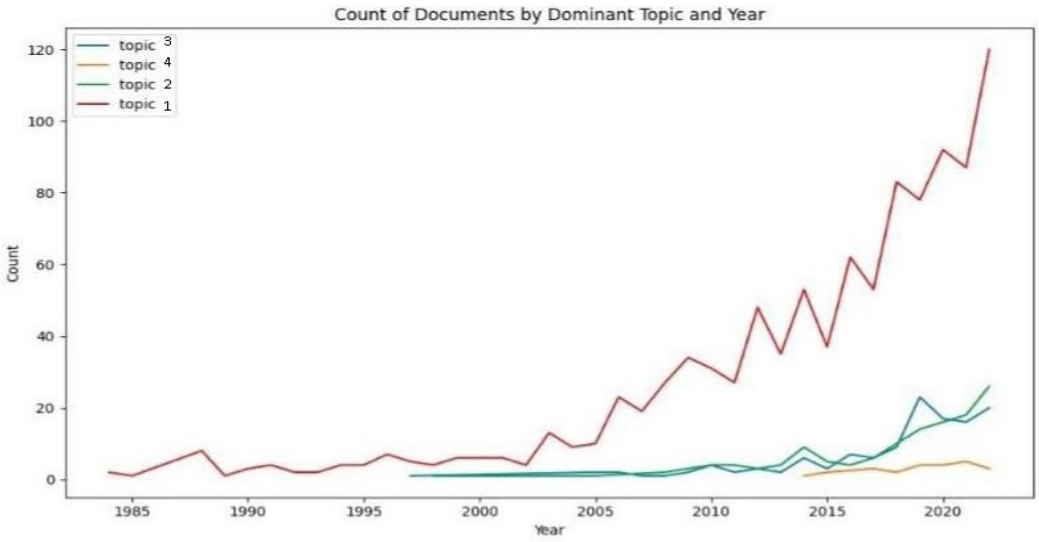


Figure 8. Topic distribution over time with LDA.

5.2. Results of NMF topic modeling

LDA is a widely used approach for identifying and monitoring relevant topics in extensive text collections (Jelodar et al., 2019). Nevertheless, alternative techniques exist, such as BERTopic and NMF, which are also commonly employed. These two approaches have been incorporated to evaluate the robustness of our findings and gain insights from alternative methods. NMF breaks down the term-document matrix into term-feature and feature-document matrices, revealing the words contributing to each topic and the topics found in each document. Unlike LDA, a probabilistic model, NMF employs matrix factorization while enforcing non-negativity

constraints (Blum & Raviv, 2023).

Figure 9 illustrates the relationship between the Optimal Number of Topics and the Coherence Score in NMF. NMF achieves the highest coherence score of 0.36, indicating identifying 10 distinct topics. As a result, NMF outperforms LDA, mainly due to its reliance on TF-IDF weighting instead of raw word frequencies (Egger & Yu, 2022). NMF aims to reduce dimensionality without sacrificing vital information and extract essential features for modeling purposes (Preetham et al., 2022). Table 2 displays the keywords generated by NMF, focusing on 1400 articles that canter around the application of systems thinking.

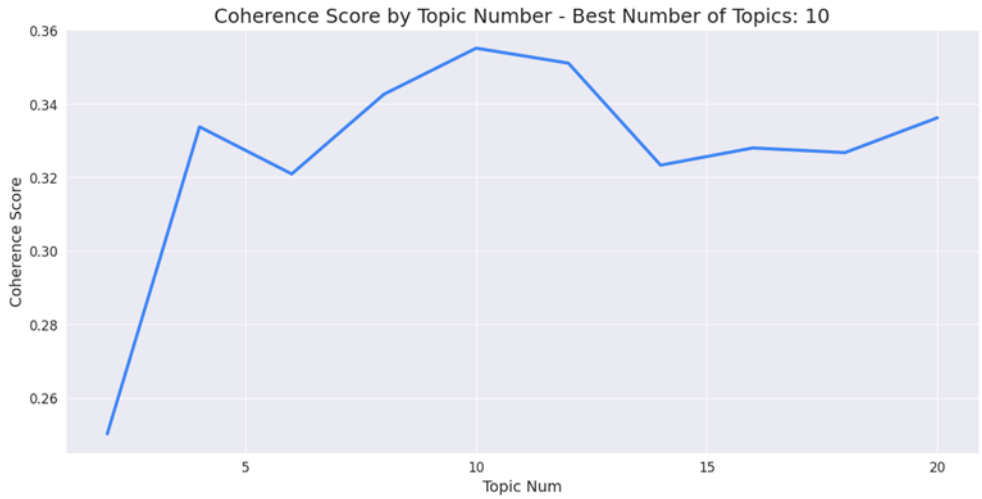


Figure 9. Coherence score of the NMF topic model for numbers of topics.

Table 2. Sample topics generated by NMF

Topic classification	Keywords	Articles (N=1400), n (%)
Topic 1: system dynamics modeling	model, dynam, causal, loop, diagram, chain, causal loop, suppli chain	181 (12.92%)
Topic 2: engineering education	engineering, cest, capacity, education, engineering think, think engineering, engineering education, capacity engineering	104 (7.42%)
Topic 3: Chemistry education	Chemistry, education, green, chemical, division, green Chemistry, Chemistry education, chemical education	49 (3.5%)
Topic 4: improve healthcare delivery	Health, care, public, public health, healthcare, policy, service, human	130 (9.28%)
Topic 5: interdisciplinary learning and problem-solving in education	Student, education, learn, school, skill, teacher, teach, think skill,	224 (31.36%)
Topic 6: improve patient outcomes	Safety, accident, risk, analysis, factor, nurse, safety culture, safety management	90 (6.42%)
Topic 7: understanding and modeling complex systems	Critical think, cest, methodology, practical, theory, research, social	193 (13.78%)
Topic 8: environmental sustainability	Sustainable, Sustainable develop, develop, environment, social, economy, integration, education Sustainable	132 (9.42%)
Topic 9: business sustainability	design, design thinking, product, process, design process, service, innovation, technology	112 (8%)
Topic 10: improve organizational performance	Management, project, knowledge, knowledge management, project management, business, organization, quality	184 (13.14%)

Upon visualizing the data, it was observed that several instances of topic overlap affect the performance of the NMF model, as documented by Preetham et al. in 2022. The intersecting themes were combined into one topic to overcome this problem. Specifically, topics 1 and 7 were merged, 2, 3, and 5 were combined, 4 and 6 were unified, and 8 and 9 were fused. As a result of this reduction, the total topic count decreased from 10 to 5, making it quite similar to the LDA model. The comparison found similarities in how both models categorized these topics. For instance, topics 1 in LDA and 1 and 7 in NMF shared common terms, representing research papers focused on "understanding and modeling complex systems".

Similarly, topics 2 in LDA and 8 and 9 in NMF pertained to documents related to "business sustainability." Topic 3 in LDA and topics 2, 3, and 5 in NMF exhibited similarities, indicating an emphasis on educational topics. Lastly, topics 4 in LDA and 4 and 6 in NMF delved into research papers concerning "improving healthcare delivery". Figure 10 illustrates the distribution of these topics over the years, demonstrating how their popularity has changed relative to each other.

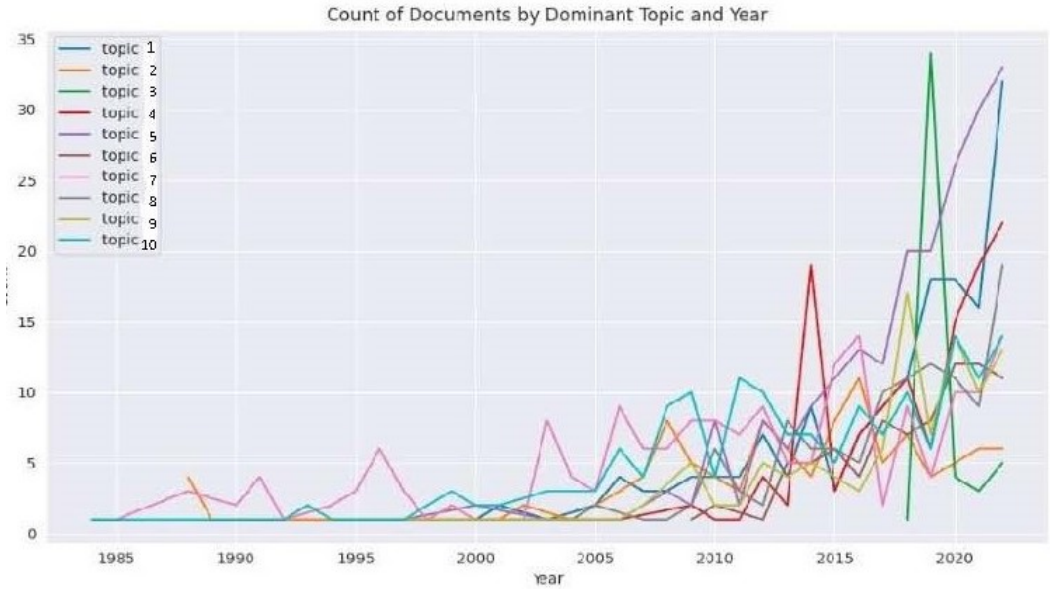


Figure 10. Topic distribution over time with NMF.

Consequently, the key topics in systems thinking research are as follows: "system dynamics modeling" (topic 1), "improve healthcare delivery" (topic 4), "interdisciplinary learning and problem-solving in education" (topic 5), "understanding and modeling complex systems" (topic 7), "environmental sustainability" (topic 8), and "improve organizational performance" (topic 10), which are commonly recognized as significant subjects in this field.

5.3. Results of BER topic modeling

BERTopic utilizes a class-based term frequency-inverse document frequency (c-TF-IDF) algorithm to evaluate term significance within a cluster and create term representations (Sánchez-Franco & Rey-Moreno, 2022). In simpler terms, it assigns higher values to terms more representative of their respective topics. Unlike LDA, BERTopic offers continuous modeling instead of discrete topics (Alcoforado et al., 2022). This model's stochastic nature leads to varied outcomes with repeated runs.

Once the model is calculated, researchers can extract the most significant topics. It is important to note that Topic 0, with a count of -1, consistently represents outliers and should be disregarded. Researchers can also search for keywords and receive the most relevant topics based on similarity scores. They can further delve into individual topics based on their keywords. To facilitate a more comprehensive analysis of the potentially extensive list of topics, BERTopic provides an interactive map of intertopic distances for examining individual topics (Grootendorst, 2022).

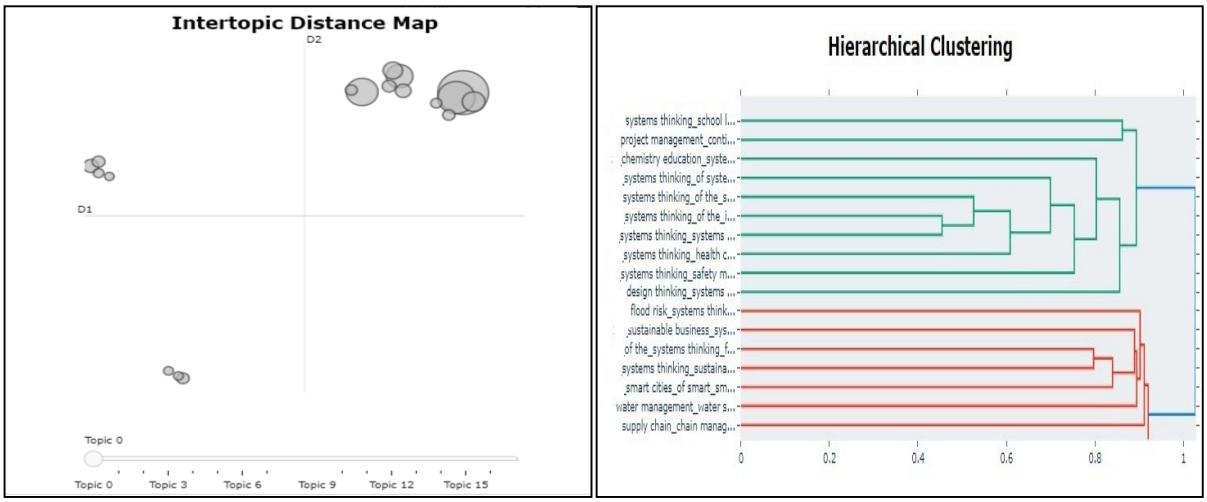
BERTopic enables the hierarchical reduction of topics, especially when closely related topics are evident in the intertropical distance map (Figure 11a). To achieve this reduction, hierarchical clustering is performed using the cosine distance matrix between topic embeddings. Similar topics are grouped (Figure 11b). For instance, Topic 13 (Six Sigma) and Topic 16 (continuous improvement) are clustered together due to their proximity. Similarly, Topics 3 and 9 (sustainable development), Topic 2 and Topic 23 (systems engineering), and Topics 12 and Topic 14 (education) are treated as part of the same cluster. This visualization aids researchers in understanding how the algorithm organizes topics.

After reviewing the proposed topic structure, researchers can interactively determine the appropriate number of topics in an interactive manner. However, it is worth noting that determining an optimal number of topics can be challenging due to topic overlap and the blending of two to three different aspects. For example, this study selected 8 topics from the 24 extracted topics (Table 3) to demonstrate how topics can be reduced (Figure 12).

Table 3. Topics generated by BERTopic

Topic	Representation	Articles (N=1400), n (%)
0 (-1)	systems thinking, of the, in the, of systems	332 (23.71%)
1	systems thinking, health care, public health, ...	202 (14.42%)
2	Systems thinking, systems engineering, engine...	129 (9.21%)

Topic	Representation	Articles (N=1400), n (%)
3	Systems thinking, of the, sustainable developm...	114 (8.14%)
4	Systems thinking, safety management, risk mana...	76 (5.42%)
5	chemistry education, green chemistry, systems ...	56 (4%)
6	Critical systems, systems thinking, of critica...	40 (2.85%)
7	systems thinking, in the, of the, system thinking	36 (2.57%)
8	systems thinking, water management, in the, ...	35 (2.5%)
9	Sustainable development, systems thinking, bio...	34 (2.42%)
10	systems thinking , quality management, ...	32 (2.28%)
11	systems thinking , social entrepreneur, ...	31 (2.21%)
12	systems thinking , higher education, ...	30 (2.14%)
13	Six sigma, systems thinking, asset management...	29 (2.07%)
14	Systems thinking, school leadership, school p...	26 (1.85%)
15	Information systems, systems thinking, soft s...	25 (1.78%)
16	project management, continuous improvement, ...	24 (1.71%)
17	Systems thinking, family farmers, food...	22 (1.57%)
18	systems thinking , cyber security, traffic man, ...	21 (1.5%)
19	electric vehicles, energy consumption, sustain, ...	20 (1.42%)
20	systems thinking , neurological position, ...	20 (1.42%)
21	Supply chain, chain management, supply chains...	19 (1.35%)
22	design thinking, design systems, systems thin...	18 (1.28%)
23	systems engineering, systems thinking, ...	18 (1.28%)
24	knowledge management, knowledge assets, criti.	11 (0.78%)



(a) (b)

Figure 11. (a) BERTopic interactive intertopic distance map; (**b**) Hierarchical clustering of topics

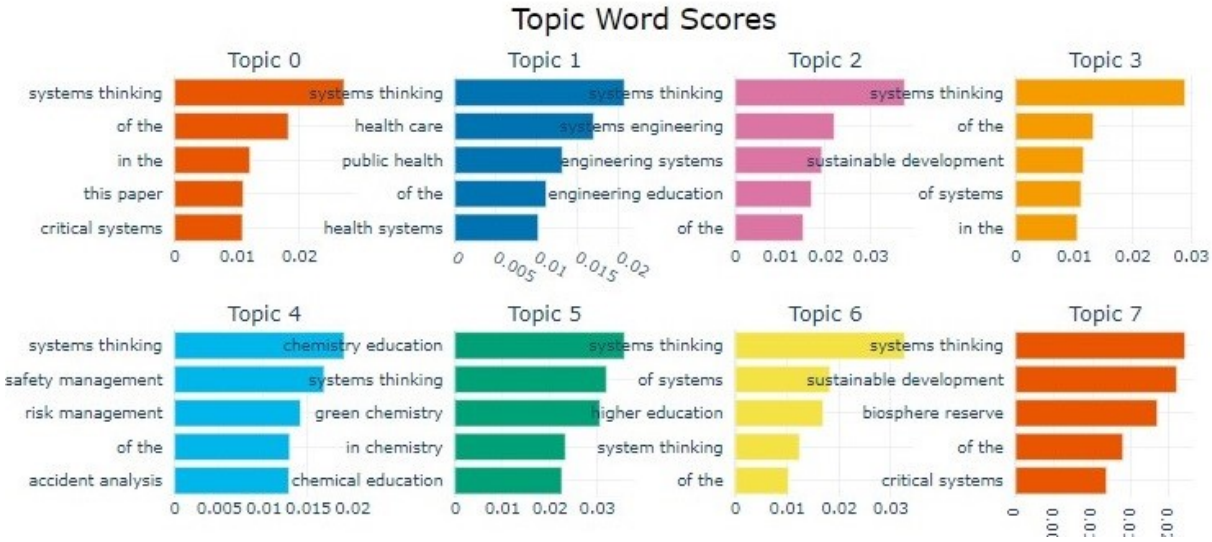


Figure 12. Extracted topics with BERTopic

Discussion and conclusions

This section presents the study's implications, the field's current state, and future inferences based on the results and discussion. The study aimed to examine articles related to system thinking (ST) in the Scopus database from 1984 to 2023. The bibliometric characteristics of the field were analyzed using descriptive analysis. The results showed increased interest in ST over the years, with a linear increase in articles published annually, particularly after 2016. Social science subject areas were the most prevalent in the interdisciplinary field, and the United States, the United Kingdom, and Australia were the leading countries in publications.

The study also focused on the applications of systems thinking. It used keyword analysis and topic modeling techniques such as LDA, NMF, and BERTopic to identify trends in different

domains and identified the most extensively covered areas in the literature The voluminous topics were “understanding and modeling complex systems”, “business sustainability” and “interdisciplinary learning and problem-solving in education”, "improving healthcare delivery", "system dynamics modeling", "engineering education", "chemistry education", "improve patient outcomes", "environmental sustainability" and "improve organizational performance. These topics have seen an increase in publications over time and are expected to continue to be studied in the future.

The study on topic modeling of system thinking themes has provided valuable insights into the multidimensional nature of system thinking. The study effectively discovered and categorized the key themes and concepts supporting system thinking through topic modeling approaches. The study's most innovative aspect is the inclusion of topic modeling analysis, which provided a summarized overview of the literature by grouping articles into logical topics characterized by key relevant terms.

The primary contribution of this study lies in developing a comprehensive understanding of the trends and themes of research in the field of systems thinking. The study's use of multiple topic modeling techniques, such as LDA, NMF, and BERTopic, can inspire further research on applying these techniques in other fields.

However, the study's exclusive focus on journal articles may be a limitation; future research could include different documents. Similar studies conducted at intervals over the years could be analyzed.

The comparison of evolving trends and deeper semantic analyses could explore specific subtopics within the ST domain.

Overall, the study provides valuable insights into the ST field's current state and future trends. The research findings are valuable to researchers and practitioners in the field, providing insights into the field's current state, research interests, and evolving trends. Also, the study's findings have practical implications for practitioners and researchers in diverse fields, highlighting the relevance of system thinking in addressing complex and dynamic challenges. The identified themes can serve as a foundation for further research, as well as for policy development and practical applications aimed at fostering a systems-oriented approach to problem-solving and decision-making.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

Abdelrazek, A., Medhat, W., Gawish, E., and Hassan, A., 2022. November. Topic Modeling on Arabic Language Dataset: Comparative Study. *In International Conference on Model and Data Engineering* ,pp. 61-71. Springer, Cham. https://doi.org/10.1007/978-3-031-23119-3_5.

Ackoff, R. L., 1971. Towards a system of systems concepts. *Management science*, *17*(11), pp. 661-671. https://doi.org/10.1287/mnsc.17.11.661

Ackoff, R. L., 1973. Science in the systems age: Beyond IE, OR, and MS. *Operations Research*, *21*(3), pp. 661-671. https://doi.org/10.1287/opre.21.3.661

Adams, K. M., and Keating, C. B., 2011. Overview of the systems of systems engineering methodology. *International Journal of Systm of Systems Engineering*, 2(2-3), p. 112-119. https://doi.org/10.1504/IJSSE.2011.040549

Alami, N., Meknassi, M., En-nahnahi, N., El Adlouni, Y., and Ammor, O., 2021. Unsupervised neural networks for automatic Arabic text summarization using document clustering and topic modeling. *Expert Systems with Applications*, *172*, p. 114652. https://doi.org/10.1016/j.eswa.2021.114652

Alcoforado, A., Ferraz, T. P., Gerber, R., Bustos, E., Oliveira, A. S., Veloso, B. M. ... and Costa, A. H. R., 2022, March. ZeroBERTo: Leveraging zero-shot text classification by topic modeling. In *International Conference on Computational Processing of the Portuguese Language*. pp. 125-136. Springer, Cham. https://doi.org/10.1007/978-3-030-98305-5_12

Assaraf, O. B. Z., and Orion, N., 2005. Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *42*(5), pp. 518-560. https://doi.org/10.1002/tea.20061

Boardman, J., & Sauser, B., 2008. *Systems thinking: Coping with 21st century problems*. CRC Press.

Barde, B. V., and Bainwad, A. M., 2017. An overview of topic modeling methods and tools. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)* .pp. 745-750. IEEE. https://doi.org/10.1109/ICCONS.2017.8250563

Beer, S., 1972. Brain of the Firm. Harmondsworth: *Allan Lane*.

Beer, S., 1979. The Hart of Enterprise. Chichester: *John Wiley.*

Bertalanffy, L. V., 1968. *General system theory: Foundations, development, applications*. G. Braziller. http://repository.vnu.edu.vn/handle/VNU_123/90608.

Bloom, B. S., 1956. Taxonomy of. *Educational Objectives*. doi: 10.1007/978-1-4419-1428-6_141.

Blum, A., and Raviv, A., 2023. The effects of the financial crisis and Basel III on banks' risk disclosure: A textual analysis. *Finance Research Letters*, *53*, p. 103680. https://doi.org/10.1016/j.frl.2023.103680

Checkland, P., 1981. Systems thinking, systems practice: includes a 30-year retrospective. *Journal-Operational Research Society*, 51(5), pp.647-647.

Checkland, P. B., 1989. Soft systems methodology. *Human systems management*, *8*(4), pp. 273-289.

Chen, Y., Zhang, H., Liu, R., Ye, Z., and Lin, J., 2019. Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*, *163*, pp.1-13. https://doi.org/10.1016/j.knosys.2018.08.011

Chowdhury, R., 2023a. Methodological flexibility in systems thinking: Musings from the standpoint of a systems consultant. *Systemic Practice and Action Research*, *36*(1), pp. 59-86. https://doi.org/10.1007/s11213-022-09597-w

Chowdhury, R., 2023b. Holistic flexibility for deploying systems thinking as a cognitive skill. *Systemic Practice and Action Research*, pp.1-23. https://doi.org/10.1007/s11213-022-09626-8

Churchman, C. W., Ackoff, R. L., and Arnoff, E. L., 1957. Introduction to operations research.

Clemson, B., 1991. *Cybernetics: A new management tool* (Vol. 4). CRC Press.

Deming, W. E., 1982. Quality, productivity, and competitive position. *(No Title)*. URL: https://cir.nii.ac.jp/crid/1130000797039430528

De Souza, D. E., 2022. A critical realist approach to systems thinking in evaluation. *Evaluation*, *28*(1), pp. 72-90. https://doi.org/10.1177/13563890211064639

Dolansky, M. A., and Moore, S. M., 2013. Quality and Safety Education for Nurses (QSEN): The Key is Systems Thinking. *Online Journal of Issues in Nursing*, *18*(3). URL: https://pubmed.ncbi.nlm.nih.gov/26812094/

Drucker, P.F., 1954. The Practice of Management. New York: *Harper & Row*.

Egger, R., and Yu, J., 2022. A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. Frontiers in sociology, 7, p. 886498. https://doi.org/10.3389/fsoc.2022.886498

Frank, M., 2000. Engineering systems thinking and systems thinking. INCOSE Journal of Systems Engineering, vol. 3, no. 3, pp. 163-168. https://doi.org/10.1002/1520-6858

Frank, M., 2002. Characteristics of engineering systems thinking – A 3-D approach for curriculum content. IEEE Transaction on System, Man, and Cybernetics, vol. 32, no. 3, Part C, pp. 203-214. https://doi.org/10.1109/TSMCC.2002.804450

Flood, R. L., and Jackson, M. C., 1991. *Critical systems thinking*. Chichester: Wiley.

Flood, R. L., and Carson, E. R., 2013. *Dealing with complexity: an introduction to the theory and application of systems science*. Springer Science & Business Media.

Forrester, J. W., 1958. Industrial dynamics: a major breakthrough for decision makers. *Harvard business review*, *36*(4), pp. 37-66.

Grohs, J. R., Kirk, G. R., Soledad, M. M., and Knight, D. B., 2018. Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems. *Thinking Skills and Creativity*, *28*, pp. 110-130. https://doi.org/10.1016/j.tsc.2018.03.003

Grootendorst, M. 2022., BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *ArXiv preprint arXiv*: p. 2203.05794. https://doi.org/10.48550/arXiv.2203.05794

Gurcan, F., Cagiltay, N. E., and Cagiltay, K., 2021. Mapping Human-Computer Interaction Research Themes and Trends from Its Existence to Today: A Topic Modeling-Based Review of past 60 Years. *International Journal of Human-Computer Interaction*, 37(3), pp.1–14, https://doi.org/10.1080/10447318.2020.1819668

Hossain, N. U. I., Dayarathna, V. L., Nagahi, M., and Jaradat, R., 2020. Systems thinking: A review and bibliometric analysis. *Systems*, *8*(3), p.23. https://doi.org/10.3390/systems8030023

Hitchins, D. K. 2003. *Advanced systems thinking, engineering, and management*. Artech House.

Islam, T., 2019. Yoga-veganism: Correlation mining of twitter health data. *ArXiv preprint arXiv: p.1906.07668*. https://doi.org/10.48550/arXiv.1906.07668

Jackson, M. C. 2019. Critical systems thinking and the management of complexity. *John Wiley & Sons.*

Jaradat, R. M., 2015. Complex system governance requires systems thinking-how to find systems thinkers. *International Journal of System of Systems Engineering*, *6*(1-2), pp.53-70.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., and Zhao, L., 2019. Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, *78*, pp.15169-15211. https://doi.org/10.1007/s11042-018-6894-4

Keating, C. B., 2008. Emergence in system of systems. *System of systems engineering: Innovations for the 21st century*, pp. 169-190. https://doi.org/10.1002/9780470403501.ch7

Krishnan, A., 2023. Exploring the Power of Topic Modeling Techniques in Analyzing Customer Reviews: A Comparative Analysis. *ArXiv preprint arXiv, p. 2308.11520*. https://doi.org/10.48550/arXiv.2308.11520

Lee, D., and Seung, H. S., 2000. Algorithms for non-negative matrix factorization. *Advances in neural information processing systems*, 13.

Maani, K. E., and Maharaj, V., 2004. Links between systems thinking and complex decision making. *System Dynamics Review: The Journal of the System Dynamics Society*, *20*(1), pp. 21-48. https://doi.org/10.1002/sdr.281

Miller, J. G., 1978. Living Systems,‖ McGraw Hill. *New York*.

Mursi, J. K. A., Subramaniam, P. R., and Govender, I. 2023. Exploring the Influence of Pre-Processing Techniques in Obtaining Labelled Data from Twitter Data. In *2023 IEEE AFRICON* .pp. 1-6. IEEE. https://doi.org/10.1109/AFRICON55910.2023.10293408

Nagahi, M., Hossain, N. U. I., Jaradat, R., Goerger, S. R., Abutabenjeh, S., and Kerr, C., 2020. August. Do the practitioners' level of systems-thinking skills differ across sector types? *In 2020 IEEE International Systems Conference (SysCon)* .pp. 1-5. IEEE. https://doi.org/10.1109/SysCon47679.2020.9275917

National Research Council., 2012. A framework for K-12 science education: Practices, crosscutting concepts, and core ideas. *National Academies Press*. URL: https://nap.nationalacademies.org/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts

Preetham, S., Reddy, B. R., Reddy, D. S. T., and Gupta, D., 2022, November. Comparative Analysis of Research Papers Categorization using LDA and NMF Approaches. In *2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon)* .pp. 1-7. IEEE. https://doi.org/10.1109/NKCon56289.2022.10127059

Principe, V. A., de Souza Vale, R. G., de Castro, J. B. P., Carvano, L. M., Henriques, R. A. P., de Almeida e Sousa Lobo, V., and de Alkmim Moreira Nunes, R., 2021. A computational literature review of football performance analysis through probabilistic topic modeling. *Artificial Intelligence Review*, pp.1-21. https://doi.org/10.1007/s10462-021-09998-8

Ratter, B. M., 2012. Complexity and Emergence. Human-nature interactions in the Anthropocene: Potentials of social-ecological systems analysis, p. 90.

Richmond, B., 1993. Systems thinking: critical thinking skills for the 1990s and beyond. *System dynamics review*, *9*(2), pp. 113-133. https://doi.org/10.1002/sdr.4260090203

Sánchez-Franco, M. J., and Rey-Moreno, M., 2022. Do travelers' reviews depend on the destination? An analysis in coastal and urban peer-to-peer lodgings. Psychology & marketing, 39(2), pp. 441-459. https://doi.org/10.1002/mar.21608

Senge, P. M., 2006. The fifth discipline: The art and practice of the learning organization. Broadway Business.

Senge, P., 1991. The learning organization made plain. *Training & Development*, *45*(10), pp. 37-44.

Sharma, A., Rana, N. P., and Nunkoo, R., 2021. Fifty years of information management research: A conceptual structure analysis using structural topic modeling. *International Journal of Information Management*, *58*, p. 102316. https://doi.org/10.1016/j.ijinfomgt.2021.102316

Shin, N., Bowers, J., Roderick, S., McIntyre, C., Stephens, A. L., Eidin, E. ... and Damelin, D., 2022. A framework for supporting systems thinking and computational thinking through constructing models. *Instructional Science*, *50*(6), pp. 933-960. https://doi.org/10.1007/s11251-022-09590-9.

Skyttner, L., 2005. *General systems theory: Problems, perspectives, practice*. World scientific.

Steward, D. V., 1981. The design structure system: A method for managing the design of complex systems. *IEEE transactions on Engineering Management*, (3), pp. 71-74.

Sweeny, L. B., and Sterman, J. D., 2000, August. Bathtub dynamics: Preliminary results of a systems thinking inventory. In *International System Dynamics Conference, Bergen, Norway*. pp.2-67

Ulrich, W., 2005. A brief introduction to critical systems heuristics (CSH). *ECOSENSUS project site*.