

Thirty-six years of contributions to queueing systems: a content analysis, topic modeling, and graph-based exploration of research published in the QUESTA journal

Aminath Shausan¹ · Aapeli Vuorinen²

Received: 23 April 2023 / Revised: 23 April 2023 / Accepted: 25 April 2023 / Published online: 2 June 2023 © The Author(s) 2023

Abstract

We investigate the 36-year history of research published in the journal *Queueing Systems: Theory and Applications* (QUESTA), to uncover trends over time in the research topics and themes covered as well as in authorship, co-authorship, and institutional affiliation. Our analysis includes three different approaches. First, we conduct a content analysis of titles and abstracts using selected keywords to examine trends in the three themes of *models, methods*, and *concepts* applied in each article. Second, we employ unsupervised topic modeling to identify more hidden topics discussed in the journal. Finally, we analyze the co-authorship graph to identify trends in co-authorship and changes in collaboration practices between authors and their research institutions. Our findings reveal a persistent popularity of studies focused on the basic modeling of queues, queueing networks, and queueing systems. We also confirm an increase in collaboration among authors over time.

Keywords Content analysis · Topic modelling · BERTopic · NLP · QUESTA

1 Introduction

QUESTA, published by Springer, is a well-established journal for research in queueing systems, that is, probabilistic and statistical models of resource sharing. Articles are primarily of a probabilistic nature, and the journal has a focus on publications that are largely mathematical in nature. Publications that develop methods to evaluate

Aminath Shausan a.shausan@uq.edu.au

School of Mathematics and Physics, The University of Queensland, Brisbane, Queensland, Australia

² Department of Industrial Engineering and Operations Research, Columbia University, New York, USA

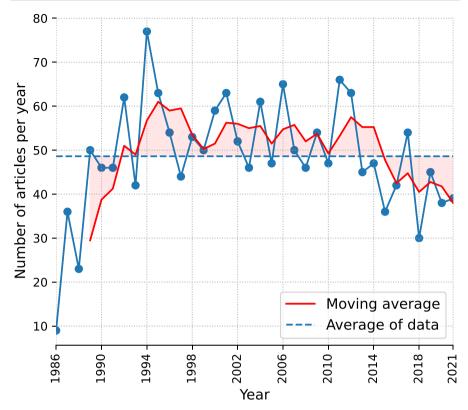


Fig. 1 Evolution of the number of articles published per year and a moving average with a window of 4 years.

the performance of models are generally welcomed by the journal. Practical areas of applications of these studies include computer and communication systems, traffic and transportation, production, storage, and logistics.

From its establishment in June 1986 until the 99th volume in October 2021, there were approximately 1,750 articles published in QUESTA, averaging around 49 articles per year (Fig. 1). The journal began with a low of 9 articles published in 1986, then rapidly increasing until 1994 when 77 articles were published. This was followed by a plateau of a relatively stable 49 to 61 yearly publications, until a gradual decline starting in 2014 that hit 39 articles in 2021.

The special 100th issue of QUESTA was published in February 2022 and includes 100 in-depth notes on views on the past, present, and future of queueing, written by established and upcoming researchers. We hope in our work to complement these excellent articles with a data scientific analysis based overview and some insightful commentary on the quantitative trends that have taken place in the field as well as the community creating the research. Our analysis attempts to answer questions such as: has the focus of the field moved from certain models to others? Has the collection of analytical tools available and being used shifted from some methods to different ones? Have the research topics changed overtime? Has the set of authors or institutions

Models	Methods	Concepts
Single server	Laplace transforms	Insensitivity
Multiserver	Large deviations	Tail asymptotics
Queueing network	Diffusion	Product form
Polling	Fluid limit	Reversibility
Priority queues		Bandwidth
Tandem		Stability
Vacations		

Table 1 List of topics within each of the three themes.

changed or remained the same? Has the way in which people work and collaborate in the field changed over time? We refer to [1-3] for some findings concerning the authors and topics in the journal's first 21 volumes.

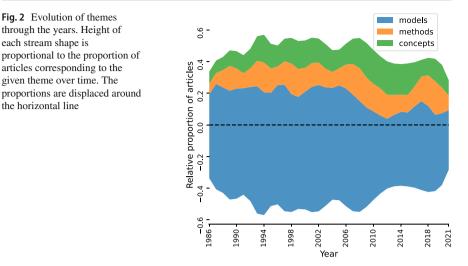
In order to answer these questions, we perform three types of analyses on all articles that have appeared in the journal from 1986 until 2021. We first perform content analysis on the articles by looking at the occurrence of selected keywords in the titles and abstracts of articles, and analyzing the trends therein. Secondly, we perform unsupervised topic modeling to uncover further topics in the corpus without influence from our preconceived notions of topics and themes. Finally, we study authorship statistics and the co-authorship graph to look at changes in the way authors in the field collaborate, as well as changes in the institutions authors of the journal are affiliated with. It is important to note that our findings on themes and topics are derived solely from titles and abstracts, and an analysis of the contents of articles may reveal different trends.

The remainder of this article is structured as follows: In Sect. 2, we discuss our content analysis of the titles and abstracts, followed by topic modeling in Sect. 3. In Sect. 4, we provide details for our findings from the co-authorship graph and the temporal trends in authorship and institutional affiliation. We close with concluding remarks in Sect. 5.

2 Content analysis

In this section, we analyze the content of the journal articles to examine a set of predefined topics and how they have evolved over time. To do so, we curated a list of topics, each associated with a keyword which we grouped into three overarching themes: *models*, *methods*, and *concepts*. Our reasoning is that most articles would examine a *concept* (such as **stability**) within a *model* (such as a **priority queue**), and use a particular *method* (such as **fluid limit**) to do so. Table 1 lists the topics categorized under their respective themes.

With this list of topics, we then tallied the number of times that each keyword occurred in the titles or abstracts of articles and have plotted trends in the frequencies over time. As such, an abstract may contain multiple topics, but we only count each



topic once per article. We have performed a number of preprocessing steps; for further details, see our online code repository [13]. This included relabeling instances of Kendall's notation to words, for instance replacing M/M/1 with single server.

Using the trends of these topics, we wish to determine among other questions if the focus of the journal has changed from certain models to others, or if the analytical tools applied in analyzing the models have changed over the years. In the rest of this section, we discuss our findings about the trends in the themes, topics, and co-occurrence of topics across the themes. A limitation of this modeling approach is that we may have missed some important topics, skewing our results.

2.1 The models theme is the most frequent

As expected, the three overarching themes have appeared regularly throughout the years, with the models theme being the most prevalent, as depicted in Fig. 2. It is not surprising that models dominate the other themes since QUESTA is primarily a modeling journal, and most articles presumably discuss models more frequently than concepts and methodology. The frequency of each theme has been relatively consistent throughout the period, except for a slight dip between 2005 and 2018, during which the models theme dropped significantly in comparison with the other two. This drop in frequency may be due to the fact that in more modern queueing theory of the past 15 to 20 years, the models we looked here may no longer be common, and there may have been an introduction of many other models which we have not considered in the current analysis.

2.2 Single server, multiserver, and stability are the most frequent topics

While analyzing the themes provides us with a very general trend, we now delve deeper into which topics contributed the most to these changes. To accomplish this,

each stream shape is

the horizontal line

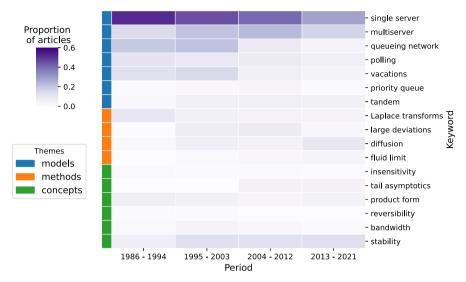


Fig. 3 Trend of topics in the three themes over four periods. The shades of the purple color indicate the proportion of articles for each topic.

we analyze the trends in the topics using a similar method to our analysis of themes. We partition the period studied into four equal sub-periods to reduce variability and compute the proportion of articles belonging to each topic within each period. The heatmap in Fig. 3 displays the trend of topics across each period, highlighting the topic corresponding to each theme. To visualize the evolution of the topics per year as line graphs, we refer to Fig. 13 in the Appendix section.

As shown in Fig. 3, the **single server** *model* is the most frequent topic in all four periods, followed by the **multiserver** *model* and the **stability** *concept* from 1995 onward. Conversely, the *concepts* of **reversibility** and **insensitivity** are the least frequently occurring topics in all periods. Each topic has some variance in its trend across periods. For example, the **polling** *model* and the **Laplace transforms** *concept* were popular during 1986–1994, but their occurrence decreased afterward. Similarly, the *models* **queueing network** and **vacations** were heavily used until 2003, but then decreased in popularity. While the use of **tandem**, **large deviations**, **diffusion**, and **bandwidth** gained some prominence during 1995–2003, they were used less frequent thereafter. **Tail asymptotics** gained attention from 2004–2012. Interestingly, the **product form** *concept* appeared consistently in all four periods.

In addition to the observed changes in topic trends across periods, significant changes in topic combinations have occurred within each period. During 1986–1994, the most popular topics were the *models* single server, multiserver, queueing network, polling, vacations, and the Laplace transforms *method*. These topics remained popular during 1995–2003, except that the *concept* of stability replaced both polling and Laplace transforms. However, according to this analysis, from 2004 onward, only single server, multiserver, and stability remained popular.

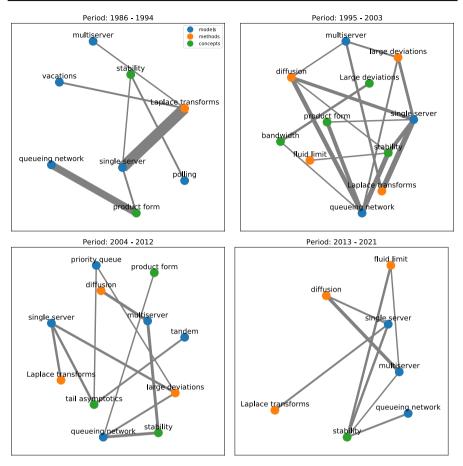


Fig. 4 Visualization of co-occurrence topics. Line thickness is proportional to the number of articles containing the pair within the corresponding period. Node colors indicate the theme and node placement is arbitrary.

2.3 Laplace transforms are commonly applied to study single-server queues

Acknowledging that a typical QUESTA publication may feature a combination of topics from each theme, we now examine the co-occurrence of topics across themes to gain insight into their relationships. We evaluate this by analyzing the frequency of topics pairs that appear across the themes in the abstracts corresponding to the four periods, as shown in Fig. 3. The corresponding graphs formed by the pairs in the periods are displayed in Fig. 4, where we only show pairs that occurred at least three times in that period.

Over time, some pairs either appeared and then disappeared or have always been together. For example, during 1986–1994, the most dominant pairs were **single server-Laplace transform** and **queueing network-product form**. These two pairs continued to be among the most dominant pairs over the next two periods but appeared less

frequently compared to the first period. In the final period, **product form** vanished from pairs, but the **single server-Laplace transform** pair remained. Similarly, the **queueing network-stability** pair, which first appeared during 1995-2003, continued to appear in the later periods, as did the **multiserver-diffusion** pair.

It is interesting to note that no graph has disconnected components, which seems to indicate that there is only one interlinked set of *models*, *methods*, and *concepts* that can be combined together, instead of, for instance, separate clusters of models and methods that go together. However, the graph is not a complete tripartite graph, so clearly not all combinations are popular; in fact, there is only one triangle in the four graphs. More topics appeared in the two graphs corresponding to the period 1995-2012, which coincides with the time frame during which the journal published an average of over 49 articles within a four-year window (Fig. 1, red curve). Note that this causes a bias as we use an absolute cut-off for which links to show. Among the four periods, the most connected is 1995–2003, with the single server and queueing network *models* each present in five pairs. In this graph, single server co-occurs with Laplace transforms, large deviations, and diffusion, which are from the *methods* theme, as well as the *concepts* of **product form** and **stability**. Similarly, **queueing** network simultaneously occurs with the *methods* diffusion and large deviations, and the concepts product form, bandwidth, and stability. The most prominent pairs in each graph usually contain a *model* except for the *method-concept* pair of **fluid limit-stability** in the last period.

A limitation of this kind of manual content analysis is that we have started from a predefined notion of what topics and themes should be, but this may be guided by our biases and flawed perceptions of relevance. To form a more objective view of topics, at the expense of some interpretability, we next turn to topic modeling using an unsupervised approach.

3 Topic modeling analysis

In this section, we utilized the BERTopic framework [6] to uncover latent topics in the corpus of articles. This technique involves three main steps, which we performed automatically using the BERTopic software. We first turn these documents (a concatenation of titles and abstracts) into document embeddings (a high-dimensional vector representation amenable to computation) using Sentence-BERT [12], a modification of the pre-trained BERT language model [4], for sentence embedding. We then perform dimensionality reduction of the embedding vectors using Uniform Manifold Approximation and Projection (UMAP) [9, 10] with cosine distance, then cluster the vectors representing documents using HDBSCAN¹. Finally, we derive topic descriptions by surfacing important words from each class by running TF-IDF² across the classes. This kind of topic modeling is a popular tool applied in inferring quantitative perspectives from corpora of texts [5, 8, 11, 14].

¹ Hierarchical Density-Based Spatial Clustering of Applications with Noise

² Term Frequency-Inverse Document Frequency

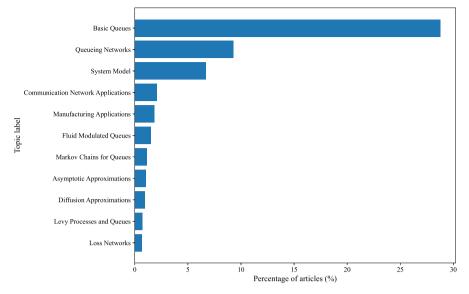


Fig. 5 Percentage of articles belonging to each topic cluster.

Performing this procedure on the corpus of documents produces 11 topics, which we list in Table 2 in order of frequency, along with their 10 most important keywords in order of descending TF-IDF score. We labeled each topic using the important keywords in consultation with the editorial team for this project whom are subject matter experts. Apart from topics 4 and 5, which concern the applications of queues, the rest of the topics relate to the theory of basic queues, queueing systems, and their analytic tools.

Figure 5 shows the percentage of articles belonging to each of the 11 topics. The largest topic in the journal focus is *Basic Queues* (29%), followed by *Queueing Networks* (9%) and *System Models* (7%). Each of the remaining topics contains less than 3% of the published articles.

These topics and their important keywords are surfaced by the unsupervised algorithm automatically without our input. Our only input is interpreting the topics by naming them based on the important keywords. Note, however, that the underlying sentence embeddings procedure uses BERT [4], a transformer language model pretrained in an unsupervised way on a huge corpus of text from the internet and elsewhere. This is the core of what allows the embeddings to model language and place similar documents near each other in the embedding space. However, this also means that the embeddings will have biases and probably will not have a sufficiently nuanced model of the field of queueing systems, likely causing the resultant clustering to be somewhat simplistic and noisy.

3.1 Changes in topic popularity over time

While Fig. 5 provides a general overview of the popularity of topics in the journal, it does not provide insight into trends over time. In order to do so, we group each article

Tonic	Lahel	Most important keywords
Tupte	Lauci	MUSH HIPOTRAIL ACY WOLDS
1	Basic queues	Time, service, system, customer, singleserver, server,
		Distribution, policy, wait, model
2	Queueing networks	Network, queueing network, node, station, wireless, show,
		Stability, model, traffic, condition
3	System model	Process, arrival, markov, time, distribution, stationary,
		System, random, discrete, service
4	Communication network	Packet, connection, traffic, source, loss, flow,
	Applications	Buffer, tcp, bandwidth, control
5	Manufacturing applications	Production, assembly, machine, system, inventory, line,
		Item, time, model, component
9	Fluid modulated queues	Fluid, buffer, markov, model, modulate, content, state,
		Second, finite, source
7	Markov chains for queues	Birth, death, process, rate, quasi, periodic, system, mean
		Time, matrix
8	Asymptotic approximations	Gaussian, fractional, brownian, source, motion,
		Large deviation, traffic, heavy, fluid, input
6	Diffusion approximations	Brownian, motion, diffusion, reflect, semimartingale,
		Convergence, srbm, quadrant, solution, boundary
10	Levy processes and queues	Reflect, sided, process, barrier, reflection, lvy, passage, first,
		Tail, input
11	Loss networks	Loss, erlang, probability, server, bound, system,
		Approximation. traffic. model. classical

11

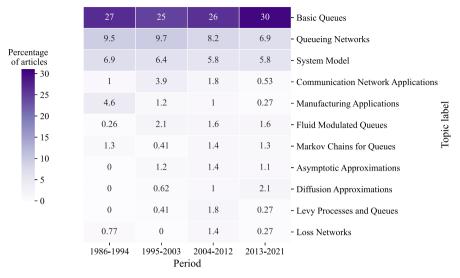


Fig. 6 Evolution of topics over time. Shades of purple color represent the percentage of articles corresponding to each period.

into one of the four periods as in the previous section and plot the percentage of articles in each topic in Fig. 6. As with the overall distribution seen in Fig. 5, *Basic Queues*, followed by *Queueing Networks* and *System Models* have been the most frequent topics in all periods, with *Basic Queues* being the most dominant. While publications on *Communication Network Applications* had a slight increase during 1995–2003, the topic has since faded. *Manufacturing Applications* also started off strong in 1986– 1994, but has since fallen in popularity drastically. The popularity of the remaining topics has been consistently low in all four periods.

3.2 Basic Queues, Asymptotic Approximations, and Levy Processes and Queues are the most common topics within themes

As mentioned previously, we discovered the topics listed in Table 2 without our prior knowledge of their keyword representations. Hence, we now provide a connection between the uncovered topics and the three overarching themes described in Table 1. For this, we searched for topics which were most closely related to each keyword listed in Table 1, using similarity scores generated by the BERTopic algorithm. In Table 3, we list the theme keywords and their most closely matching topic. This process reveals that *Basic Queues, Asymptotic Approximations*, and *Levy Processes and Queues* are the most common topics among *models, methods*, and *concepts* themes, respectively.

Theme	Theme keywords	Торіс
Models	Single server	Basic Queues
	Multiserver	Basic Queues
	Queueing networks	Queueing Networks
	Polling	Fluid Modulated Queues
	Priority queues	System Model
	Tandem	Manufacturing Applications
	Vacations	Basic Queues
Methods	Laplace transforms	Asymptotic Approximations
	Large deviations	Asymptotic Approximations
	Diffusion	Diffusion Approximations
	Fluid limit	Fluid Modulated Queues
Concepts	Insensitivity	Levy Processes and Queues
	Tail asymptotics	Loss Networks
	Product form	Manufacturing Applications
	Reversibility	Levy Processes and Queues
	Bandwidth	Communication Network Applications
	Stability	Queueing Networks

Table 3 The topic most similar to the corresponding theme keyword.

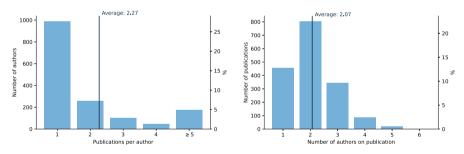


Fig. 7 Histogram of articles published per author (left) and number of authors per article (right).

4 Authorship analysis

We now turn our attention to trends in authorship, co-authorship, and institutional affiliation. The data come from the list of articles received from Springer as well as the DBLP³. In total, there are 1573 unique authors affiliated with 535 research organizations. Figure 7 shows the distribution of articles published per author and authors per article along with averages. Of particular note is the small subset of 175 authors (11.1%) who have five or more publications in the journal. This group has an average 8.9 articles per author, and a significant majority of 1124 (65%) publications have at least one of these authors as a co-author.

³ https://dblp.uni-trier.de

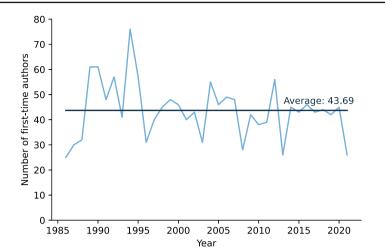


Fig. 8 Number of first time authors per year.

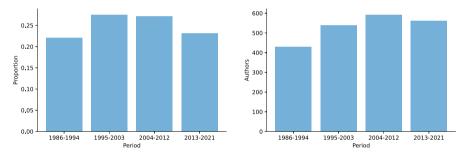


Fig. 9 Proportion of articles per period (left) and number of unique authors per period (right).

The number of first-time authors in QUESTA is shown in Fig. 8. After a slow start, the number of new authors per year increased rapidly and remained high at around 60 per year for the first decade. After that, the number dropped to approximately 40 per year and has remained stable since.

4.1 There is an increase in the trends of the number of collaborations and unique research institutions

To study the temporal trends in authorship, we partitioned the articles into four equalsized time periods, similar to earlier. Figure 9 shows the proportion of articles and the number of unique authors in each period. Examining the two distributions in Fig. 10, we observe that the average number of authors per article has increased in consecutive periods. The same trend can be seen for the average number of unique research institutions per article.

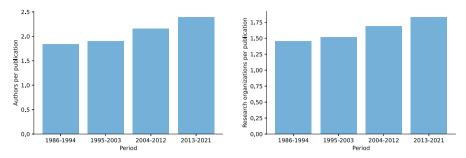


Fig. 10 Average number of authors and unique research institutions per publication.

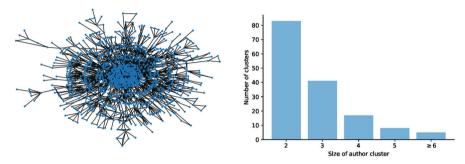


Fig. 11 The full co-authorship graph, showing the main cluster and additional clusters (left). Node placement is computed using a force-directed algorithm. Histogram of cluster size excluding largest cluster (right).

4.2 Most authors are connected to each other

In order to understand co-authorship behavior and trends in it, we constructed the co-authorship graph. This is an undirected graph G(V, E) where the set of nodes, V, corresponds to the set of authors who have ever published in QUESTA, and the edge set, E, have $(i, j) \in E$ if author i and j both appear as co-authors of some article in the corpus. Overall, the graph contains one large connected component (or "cluster") of 1024 authors (70%), with Fig. 11 (right) showing the distribution of the number of authors in all other clusters. As seen in Fig. 11 (left), which illustrates the co-authorship graph, it is hard to see much structure from merely the graph itself. The graph analysis was performed using Python and NetworkX [7].

In addition to this full co-authorship graph, we additionally split the time interval into the aforementioned four sub-periods and constructed four smaller co-authorship graphs which contained only articles from that period. This allowed us to study the changes in co-authorship over time. Figure 12 shows the proportion of authors within the five largest clusters over time, as well as the proportion of authors in the single largest cluster.

It is clear that there has been a significant increase in various types of collaboration among authors in the community. This is evident in the increase in the relative size of the clusters with fewer groups of authors working independently of the main group. Similarly, the number of authors per article has increased, as well as cross-institutional

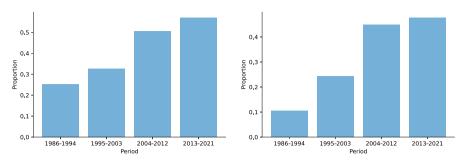


Fig. 12 Proportion of authors in top 5 clusters (connected components) (left) and that of just the largest one (right).

collaboration with later articles having representation from more unique institutions. There may be many reasons for this, one obvious explanation is the increase in ease of collaboration due to the Internet, email, chat applications, and lately videoconferencing, as well as increase in international travel and international conferences.

5 Conclusions

We have presented an analysis of 36 years of research published in the QUESTA journal, utilizing three complementary quantitative methods to examine trends in themes, topics, authorship, and institutional affiliation. We hope that our findings offer valuable insights into the evolution of queueing theory research, highlighting the dominant themes and topics that have been explored in the field over time through the lens of the QUESTA journal.

Through our exploration of themes and topics, we discovered that research has primarily focused on modeling basic queues and queueing networks, while research targeted towards the application aspect of queues has been relatively scarce. Furthermore, our investigation of co-authorship revealed that the number of authors, publishing rates, and first-time author rates have remained constant over time, while collaborations between authors and the number of research institutions involved has steadily increased.

Acknowledgements We thank QUESTA journal editors Michel Mandjes and Yoni Nazarathy for their helpful discussions which supported this project. We note that we are generally not queueing theory experts and relied on some input from Michel Mandjes and Yoni Nazarathy who work in the area of the journal. We also thank the anonymous referee for his valuable comments which helped to improve the content of this work.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted

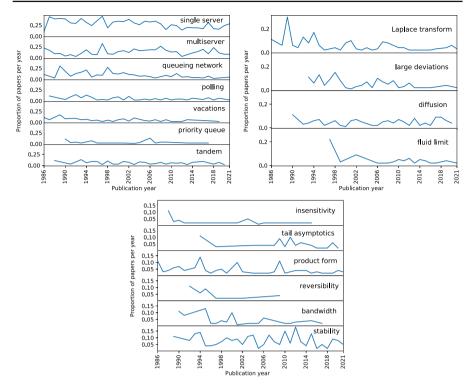


Fig. 13 Evolution of keyword in the *models* (top left), *methods* (top right) and *concepts* (bottom). Each trend line shows the proportion of articles per year.

by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Appendix

In this section, we provide an alternative visualization of the evolution of the theme keywords by depicting the trends as line graphs without partitioning the time frame into sub-periods (Fig. 13).

Although the conclusion from this visualization remains the same as that from Fig. 3, we can infer further details about the trends of each keyword, especially the ones that are less popular ones. For example, the **fluid limit** *method* began around the late 1990s and was initially very popular, but then declined from 2002 onward. Interestingly, the **reversibility** *concept* had the shortest span, starting from the mid-1990s and lasting until 2009.

References

1. Boxma, O.J., Núñez Queija, R.: Author index, Vol. 1-21, (1997)

- 2. Boxma, O.J., Núñez Queija, R.: Preface to author and subject indexes, Vol. 1-21, (1997)
- 3. Boxma, O.J., Núñez Q.R.: Subject index, Vol. 1-21, (1997)
- Devlin, J., Chang, M.-W., Lee, K., Kristina, T.: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint, BERT (2018)
- Egger, R., Yu, J.: A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify twitter posts. Front. Sociol. 7 (2022)
- Grootendorst, M.: BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv preprint, (2022)
- Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using networkx. In: Gaël V., Travis V., Jarrod M., (eds.) Proceedings of the 7th Python in Science Conference, pp. 11 – 15, Pasadena, CA USA, (2008)
- Jiang, H., Qiang, M., Lin, P.: A topic modeling based bibliometric exploration of hydropower research. Renew. Sustain. Energy Rev. 57, 226–237 (2016)
- McInnes, L., Healy, J., Melville, J.: Uniform manifold approximation and projection for dimension reduction. arXiv preprint, UMAP (2018)
- McInnes, L., Healy, J., Saul, N., Grossberger, L.: UMAP: uniform manifold approximation and projection. J. Open Sour. Softw. 3(29), 861 (2018)
- Porturas, T., Taylor, R.A.: Forty years of emergency medicine research: uncovering research themes and trends through topic modeling. Am. J. Emerg. Med. 45, 213–220 (2021)
- Reimers, N., Gurevych, I.: Sentence-BERT: sentence embeddings using siamese BERT-Networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, pp. 3982–3992. Association for Computational Linguistics, 11 (2019)
- Vuorinen, A., Shausan, A.: Insights from the past ninety-nine volumes of QUESTA [Source Code], 2022. https://github.com/aapeliv/questa
- Yau, C.-K., Porter, A., Newman, N., Suominen, A.: Clustering scientific documents with topic modeling. Scientometrics 100(3), 767–786 (2014)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.