

From Words to Networks: A Bibliometric Exploration of Co-authorship, Co-citation, and Keyword Trends

Ms. Chandrakala V

Assistant Professor, Department of Commerce, St. Francis College, Bengaluru, Karnataka

Dr. Bharathi Ravishankar

Associate Professor, International School of Management Excellence, Bangalore

Ramya R

Guest Faculty, Department of Commerce, R. C. College of Commerce and Management, Bengaluru, Karnataka

Dr. Basavaraja M. T.

Librarian, ISBR Business School, Bengaluru, Karnataka

Abstract

Purpose: This study aims to map the research landscape of sentiment analysis by examining authorship patterns, intellectual structure, and emerging thematic trends.

Design/Methodology/Approach: A bibliometric analysis was conducted using 2,697 publications retrieved from the Scopus database for the period 2004–2025. Data were analysed using co-authorship, co-citation, and keyword co-occurrence techniques, and visualised through bibliometric mapping tools.

Findings: The results indicate that China and India lead in research productivity, while the United States dominates in citation impact. Influential authors such as Bing Liu and Erik Cambria shape the intellectual structure of the field. Keyword analysis highlights the dominance of sentiment analysis, social media, and data mining, with a clear shift toward machine learning and deep learning approaches. Co-citation analysis reveals three major research clusters: traditional methods, lexicon-based approaches, and AI-driven techniques.

Conclusion: The study demonstrates the rapid growth and increasing interdisciplinary of sentiment analysis, highlighting its evolving methodologies and expanding applications across diverse domains.

Keywords: Network Analysis, Sentiment Analysis, Co-authorship, Co-citation, Keyword Trends, Bibliometrics, VOSviewer.

1. Introduction

In the era of big data and artificial intelligence, sentiment analysis has become a powerful tool for understanding human emotions and opinions through textual data. Also known as opinion mining, sentiment analysis applies natural language processing (NLP), machine learning (ML), and computational linguistics to extract subjective information from various sources, including social media, product reviews, and news articles (Liu, 2012). As organisations, researchers, and policymakers increasingly rely on sentiment analysis to gauge public opinion and predict trends, the field has witnessed exponential growth in academic publications. This surge in research necessitates a bibliometric analysis to systematically map the intellectual structure, influential works, and evolving research trends in sentiment analysis (Aria & Cuccurullo, 2017). Bibliometric analysis is a quantitative method used to analyse academic literature through citation patterns, co-authorship networks, and keyword co-occurrences. It enables researchers to assess the impact of influential authors, institutions, and countries while identifying emerging themes in a field (Donthu et al., 2021). In the context of sentiment analysis, bibliometric techniques can uncover how interdisciplinary collaborations shape the field, which publications and scholars have the most impact, and how research trends have evolved (Cambria et al., 2017). Prior studies have highlighted the dominance of machine learning and deep learning techniques in sentiment analysis, with a growing emphasis on social media analytics, emotion detection, and multilingual sentiment analysis (Pang & Lee, 2008; Medhat et al., 2014).

Despite the increasing volume of sentiment analysis research, a comprehensive bibliometric study that explores its co-authorship networks, co-citation structures, and keyword trends remains limited. Understanding these aspects is crucial in identifying leading contributors, major theoretical frameworks, and the thematic progression of sentiment analysis. By analysing research collaborations, we can evaluate the global academic landscape and the strength of partnerships between institutions and countries (Van Eck & Waltman, 2014). Similarly, examining co-citation networks reveals the foundational works that have shaped sentiment analysis methodologies, from lexicon-based approaches to advanced deep learning models (Liu, 2020). Additionally, keyword analysis helps identify shifting research focus, such as the transition from traditional NLP methods to transformer-based models like BERT and GPT (Devlin et al., 2019).

This study aims to provide a comprehensive bibliometric analysis of sentiment analysis literature, focusing on co-authorship, co-citation patterns, and keyword evolution. By leveraging bibliometric tools, we will uncover the research dynamics, influential studies, and emerging trends shaping sentiment analysis. The findings will offer valuable insights into how the field has evolved, which scholars and institutions drive innovation, and where future research is headed. Ultimately, this analysis will contribute to a better understanding of the scholarly landscape of sentiment analysis and guide future research directions.

2. Objectives

- To analyse the authorship and collaboration networks in sentiment analysis research.
- To assess the intellectual structure of sentiment analysis research through citation and co-citation analysis.
- Visualise and interpret scholarly networks of sentiment analysis literature using a bibliometric mapping tool.
- To explore emerging research trends and thematic areas in sentiment analysis literature.

3. Scope and Methodology

This study adopts a quantitative bibliometric approach to map and analyse the research landscape of sentiment analysis over the period from 2004 to 2025. The scope of the study is limited to scholarly publications indexed in the Scopus database, ensuring the inclusion of high-quality and peer-reviewed literature. Data were retrieved using a title-based search for the term “Sentiment Analysis,” which provides a focused and relevant dataset. A total of 2,394 records were obtained, including bibliographic details such as authors, affiliations, keywords, and citation information (Zhang, Wang, & Liu, 2018). The retrieved data were exported in CSV (.csv) format and subjected to a rigorous data cleaning process to ensure accuracy and consistency. This involved the removal of duplicate records, elimination of incomplete entries, and standardisation of author names and institutional affiliations. Such preprocessing was essential to improve the reliability of subsequent bibliometric analyses and to avoid inconsistencies in network mapping.

The study encompasses key dimensions of bibliometric analysis, including authorship patterns, institutional and country contributions, citation structures, and keyword trends. In line with the objectives, co-authorship analysis was conducted to examine collaboration networks among authors, institutions, and countries. Citation and co-citation analyses were performed to identify influential publications, authors, and sources, thereby uncovering the intellectual structure of sentiment analysis research. Additionally, keyword co-occurrence analysis was used to explore major research themes and emerging trends within the field. To facilitate deeper insights, bibliometric mapping and network visualisation techniques were employed using specialised tools such as VOSviewer. These tools enable the graphical representation of relationships among different bibliographic elements, helping to identify clusters, collaboration patterns, and thematic developments in the research domain.

4. Results and Analysis

Table 1. Co-authorship of the top ten authors

Sl.no	Author	Documents	Citations	Total link strength
1.	Bhattacharyya, Pushpak	16	720	23

2.	Ekbal, Asif	15	591	20
3.	Cambria, Erik	12	1700	11
4.	Yu, Jianfei	8	325	9
5.	Gutiérrez, Yoan	8	46	9
6.	Poria, Soujanya	7	1283	12
7.	Balahur, Alexandra	7	415	0
8.	Martínez-cámara, Eugenio	7	323	13
9.	Mamidi, Radhika	7	38	0
10.	Nakov, Preslav	6	1126	15

Table 1 provides insights into the co-authorship network of the top ten authors. Pushpak Bhattacharyya has the highest number of publications (16 documents) and a significant total link strength (23), indicating strong collaborative connections. Asif Ekbal follows closely with 15 documents and a total link strength of 20, highlighting active research contributions and collaborations. The table also shows that Erik Cambria has received the highest number of citations (1700 citations for 12 documents), indicating the high impact research despite a moderate number of publications. Soujanya Poria (1283 citations for seven documents) and Preslav Nakov (1126 citations for six documents) also demonstrate high research impact. Bhattacharyya and Ekbal are the most prolific authors with strong collaborative networks. Cambria, Poria, and Nakov are high-impact researchers with significant citations. Despite contributing papers, some authors have lower collaboration scores, indicating independent or limited research partnerships.

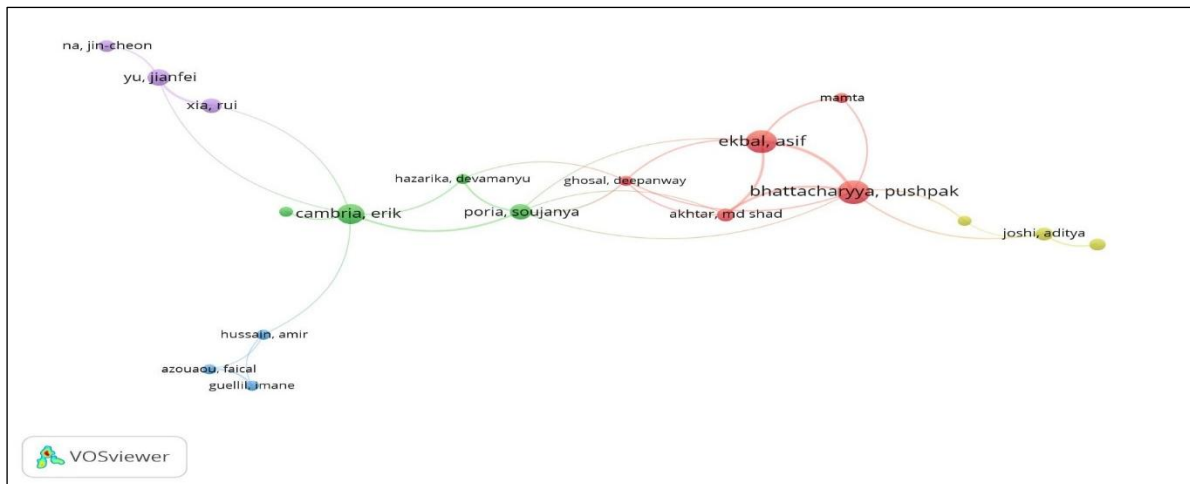


Figure 1. Network visualisation of the co-authorship of authors

Table 2. Co-authorship of the top ten organisations

Sl. no	Organization	Documents	Citations	Total link strength
1.	Indian Institute of Technology, Patna, India	6	285	1
2.	Nanjing University of Science and Technology, China.	6	244	1
3.	Brawijaya University, Indonesia	5	63	2

4.	Nanyang Technological University, Singapore	5	97	7
5.	National University of Singapore, Singapore	5	69	6
6.	School of Economics and Management, Nanjing University of Science and Technology, China	5	41	2
7.	School of Information Management, Nanjing University, China	5	22	5
8.	Harvard University, United States	4	155	2
9.	Columbia University, New York, United States	4	139	2
10.	Department of Library and Information Science, Mizoram University, India	4	17	3

Table 2 indicates the top ten organisations' research productivity, impact, and collaboration networks. IIT Patna (India) and Nanjing University of Science and Technology (China) lead with six publications each, followed by Brawijaya University, Nanyang Technological University, and National University of Singapore with five publications each. The IIT Patna (285 citations), Nanjing University of Science and Technology (244 citations), and Harvard University (155 citations) have the most citations, indicating strong academic influence. The institutions are from India, China, Singapore, Indonesia, and the United States, with Chinese institutions contributing the most papers. However, Singaporean institutions have a more balanced mix of productivity, impact, and collaboration.

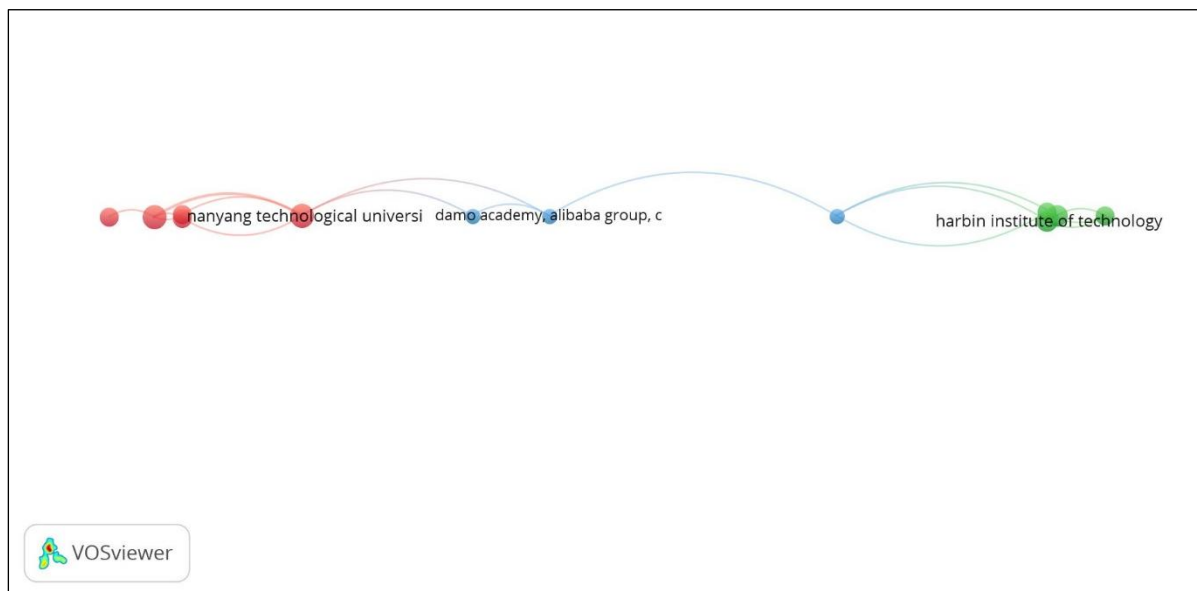


Figure 2. Network visualisation of co-authorship of the organisations

Table 3. Co-authorship of the top ten countries

Sl.no	Country	Documents	Citations	Total link strength
1.	China	406	9274	139
2.	India	378	8358	90
3.	United States	311	21337	137

4.	Spain	136	2738	67
5.	United Kingdom	133	8318	127
6.	Indonesia	113	763	15
7.	Malaysia	79	1265	61
8.	Italy	71	4917	28
9.	Germany	64	3884	46
10.	Canada	60	4280	50

Table 3 indicates the co-authorship network of the top ten countries based on documents (publications), citations (research impact), and total link strength (collaboration level). China (406 documents) and India (378 documents) lead in research output, followed by the United States (311 documents). The United States (21,337 citations) has the most citations, making it the most influential country. China (9,274 citations) and India (8,358 citations) follow. China (139), the U.S. (137), and the U.K. (127) exhibit the highest total link strength, indicating strong international research partnerships. The U.K., Germany, and Canada show a good balance between publications, impact, and collaboration strength. Indonesia and Malaysia have growing research contributions but show lower citations and weaker collaboration networks. Overall, China and India dominate in productivity, the U.S. leads in impact, and European countries maintain strong collaborations.

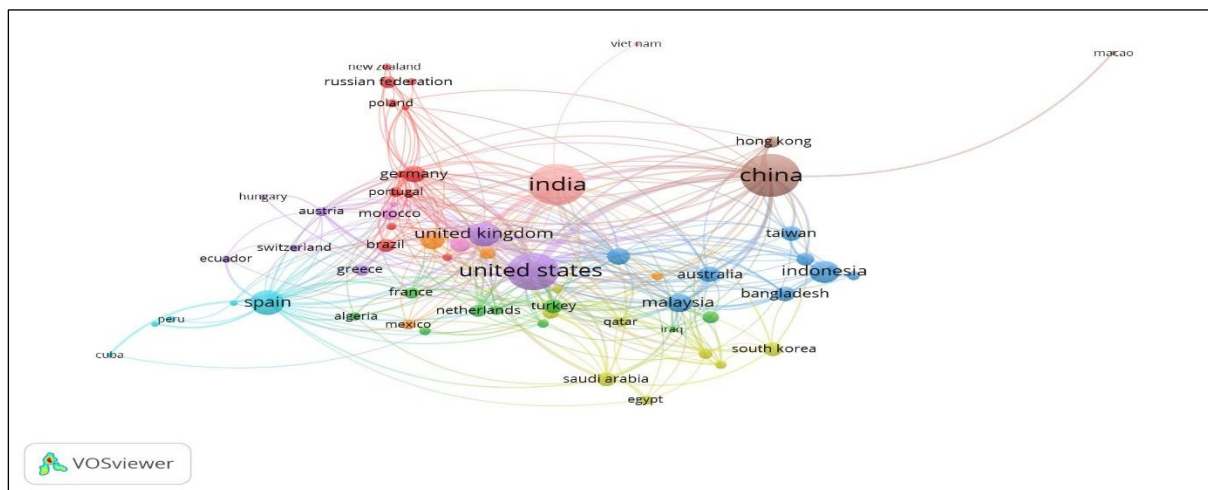


Figure 3. Network visualisation of co-authorship of countries

Table 4. Co-occurrence of top ten keywords

Sl.no	Keyword	Occurrences	Total link strength
1.	Sentiment Analysis	2026	13215
2.	Social Networking (online)	444	4090
3.	Data Mining	373	3036
4.	Social Media	350	2610
5.	Classification (of information)	280	2741
6.	Semantics	277	2193
7.	Machine Learning	276	2360

8.	Computational Linguistics	271	1993
9.	Deep Learning	241	2154
10.	Learning Systems	240	2582

Table 4 shows the co-occurrence of the top ten keywords. It is observed from the table that the keyword “Sentiment Analysis” dominates the field with the highest occurrences (2026) and strongest link strength (13,215), making it the core research area. Key related topics include “Social Networking (Online)” (444 occurrences), “Data Mining” (373), and “Social Media” (350), highlighting a strong focus on analysing online interactions. AI-driven techniques such as “Machine Learning” (276), “Deep Learning” (241), and “Computational Linguistics” (271) are widely used in sentiment analysis. The research is highly interdisciplinary, combining social media analytics, AI, and natural language processing (NLP). Overall, sentiment analysis is the central theme, with AI, data mining, and computational linguistics playing a crucial role in advancing research in this area.

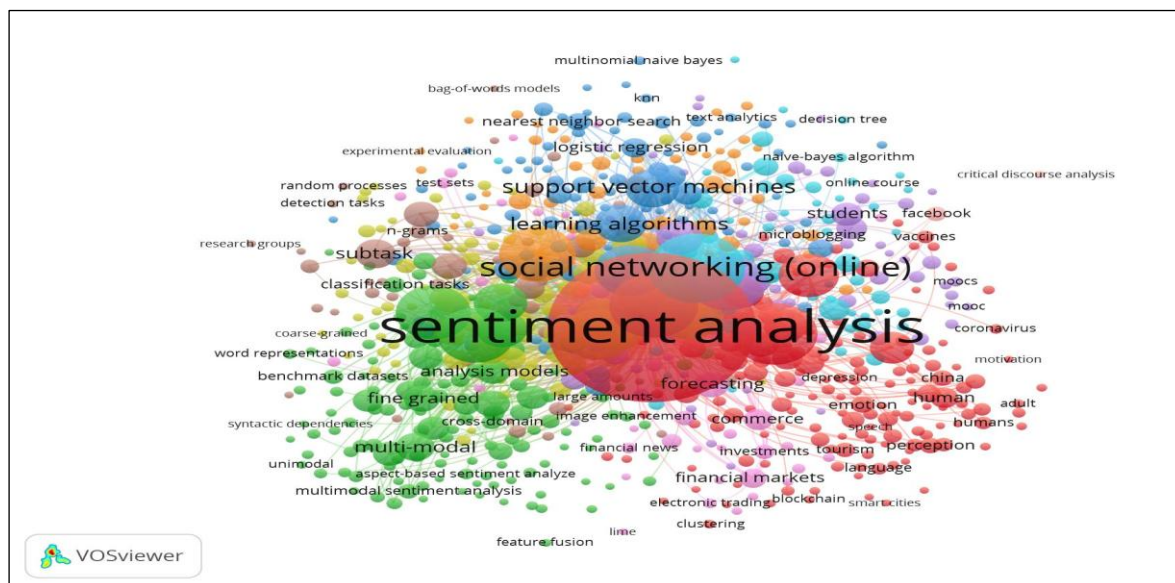


Figure 4. Network visualisation of the co-occurrence of keywords

Table 5. Co-citation of the top ten cited references

Sl. no	Cited references	Citations	Total link strength
1.	Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. <i>Foundations and Trends® in information retrieval</i> , 2(1–2), 1-135.	92	188
2.	Liu, B. (2012). Sentiment analysis and opinion mining. <i>Synthesis lectures on human language technologies</i> , 5 (1), 1–167.	78	185
3.	Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. <i>Ain Shams engineering journal</i> , 5(4), 1093–1113.	62	122

4.	Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory, <i>Neural Computation</i> , 9(8)1735-1780.	53	118
5.	Hu, M., & Liu, B. (2004, August). Mining and summarising customer reviews. In <i>Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining</i> (pp. 168–177).	52	157
6.	Kenton, J. D. M. W. C., & Toutanova, L. K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In <i>Proceedings of naacL-HLT</i> (Vol. 1, No. 2). (pp. 4171–4186)	50	134
7.	Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. <i>Computational linguistics</i> , 37(2), 267–307.	43	140
8.	Liu, B. (2020). <i>Sentiment analysis: Mining opinions, sentiments, and emotions</i> . Cambridge University Press.	41	90
9.	Hu, M., & Liu, B. (2004, August). Mining and summarising customer reviews. In <i>Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining</i> (pp. 168–177).	38	88
10.	Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In <i>Mining text data</i> (pp. 415–463). Springer, Boston, MA.	38	91

Table 5 highlights the key co-citation patterns in sentiment analysis research. The most influential work is Pang & Lee (2008), with the highest citation count (92) and total link strength (188), establishing it as a foundational paper. Liu (2012) follows closely with 78 citations and a TLS of 185, reinforcing its significance in opinion mining. Survey papers by Medhat et al. (2014) and Liu & Zhang (2012) demonstrate broad influence, while Hu & Liu (2004) appeared twice, emphasising its role in customer review mining. Taboada et al. (2011) remained relevant with a high TLS (140) for lexicon-based sentiment analysis. Deep learning models, such as LSTM (Hochreiter & Schmidhuber, 1997) and BERT (Kenton et al., 2019), suggest a shift from traditional lexicon-based approaches to advanced AI-driven sentiment analysis. Overall, this co-citation analysis reveals three major research clusters: traditional sentiment analysis (Pang & Lee, Liu, Hu & Liu), lexicon-based approaches (Taboada et al.), and deep learning-based advancements (LSTM, BERT). Foundational works continue to shape the field, while modern AI techniques are gaining prominence.

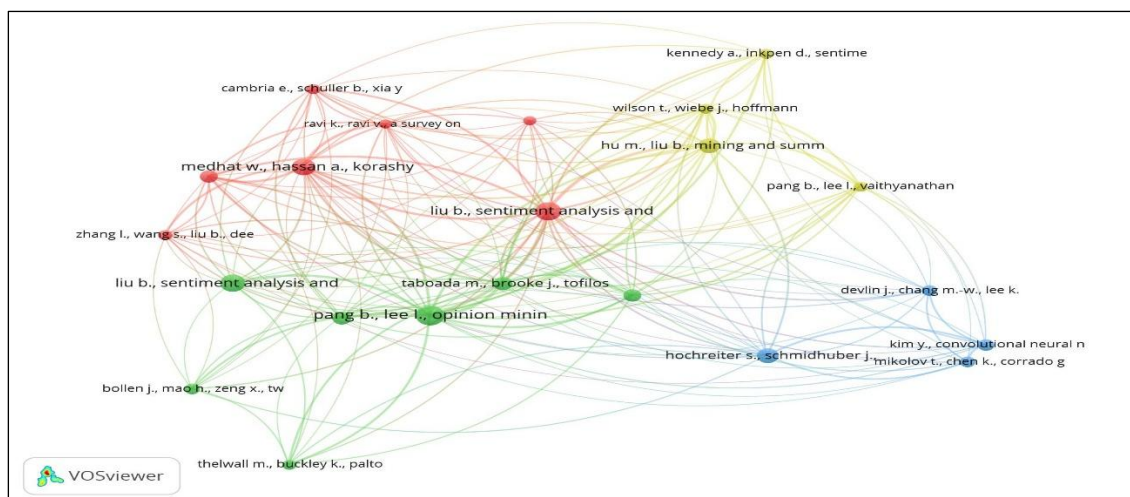


Figure 5. Co-citation network visualisation of cited references

Table 6. Co-citation of the top ten cited sources

Sl.no	Source	Citations	Total link strength
1.	IEEE Access	708	13218
2.	Expert Systems with Applications	504	10241
3.	PLOS One	333	4518
4.	Information Processing & Management	309	7136
5.	Knowledge-Based Systems	308	7316
6.	Neurocomputing	292	7401
7.	Sustainability	267	4406
8.	Advances in Neural Information Processing Systems	262	3002
9.	Procedia Computer Science	248	4093
10.	Tourism Management	246	6743

Table 6 highlights the most influential sources in sentiment analysis research. IEEE Access leads in citations (708) and total link strength (13,218), indicating its central role in the field. Expert Systems with Applications (504 citations, TLS 10,241) follows closely, emphasising the application of AI in sentiment analysis. Key AI and deep learning journals like Neurocomputing, Knowledge-Based Systems, and NeurIPS show the growing importance of machine learning-based sentiment analysis. Information Processing & Management (TLS 7,136) reinforces the significance of information retrieval and NLP. Application-driven sources such as Tourism Management and Sustainability highlight sentiment analysis's real-world use in tourism, social sciences, and environmental studies. Procedia Computer Science indicates the strong presence of ongoing research in conference proceedings. AI, deep learning, and interdisciplinary applications strongly influence sentiment analysis research. IEEE Access, Expert Systems with Applications, and Neurocomputing serve as foundational sources, while sustainability and tourism research showcase the field's expanding impact beyond computer science.

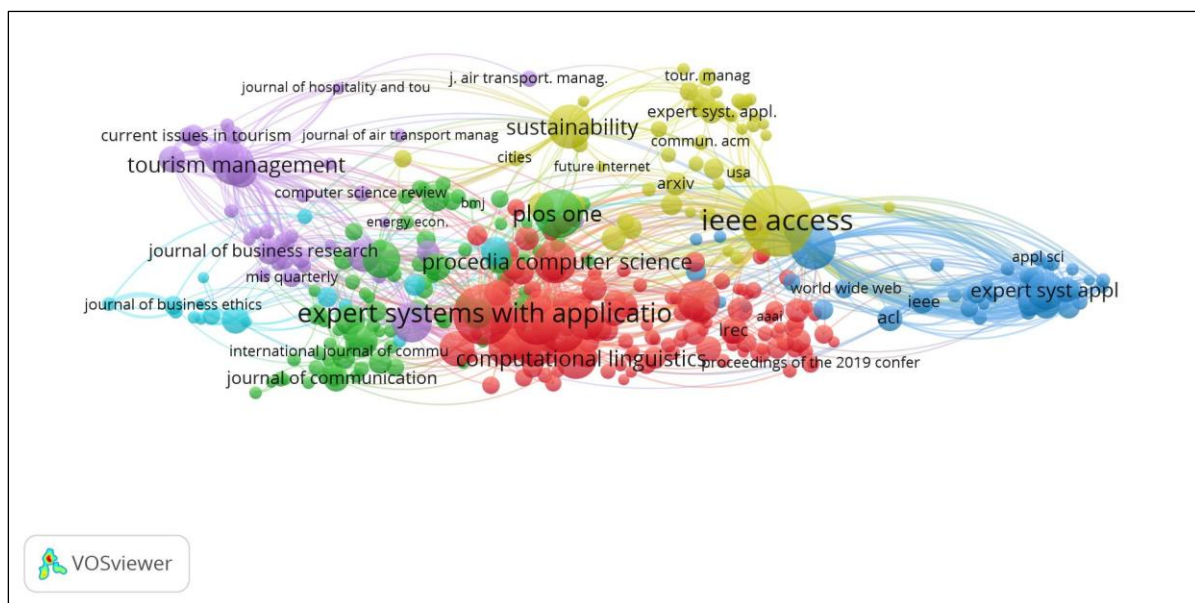


Figure 6. Co-citation network visualisation of cited sources

Table 7. Co-citation of the top ten cited authors

Sl.no	Author	Citations	Total link strength
1.	Liu, B.	1157	82553
2.	Cambria, E.	1011	112047
3.	Lee, L.	664	34609
4.	Pang, B.	649	34098
5.	Poria, S.	449	53123
6.	Wiebe, J.	429	28706
7.	Zhang, Y.	415	39944
8.	Wang, Y.	393	32740
9.	Liu, Y.	377	33332
10.	Wilson, T.	347	23244
11.	Wang, J.	340	30699

Table 7 highlights the most influential authors in sentiment analysis research. Bing Liu (1157 citations, TLS 82,553) is the most cited author, confirming his foundational contributions to opinion mining and sentiment analysis. Erik Cambria (1011 citations, TLS 112,047) has the highest total link strength, indicating his extensive influence in AI-driven sentiment analysis. Pang & Lee remain highly relevant, with their early work on machine learning-based sentiment classification being frequently co-cited. Poria, Zhang, and Wang represent a shift toward deep learning and multi-modal sentiment analysis, while Wiebe and Wilson continue to impact lexicon-based sentiment approaches. Sentiment analysis research is shaped by foundational studies (Liu, Pang & Lee), AI-driven approaches (Cambria, Poria, Zhang), and lexicon-based methods (Wiebe, Wilson). The high co-citation strength of Cambria and Poria reflects the increasing dominance of deep learning in sentiment analysis.

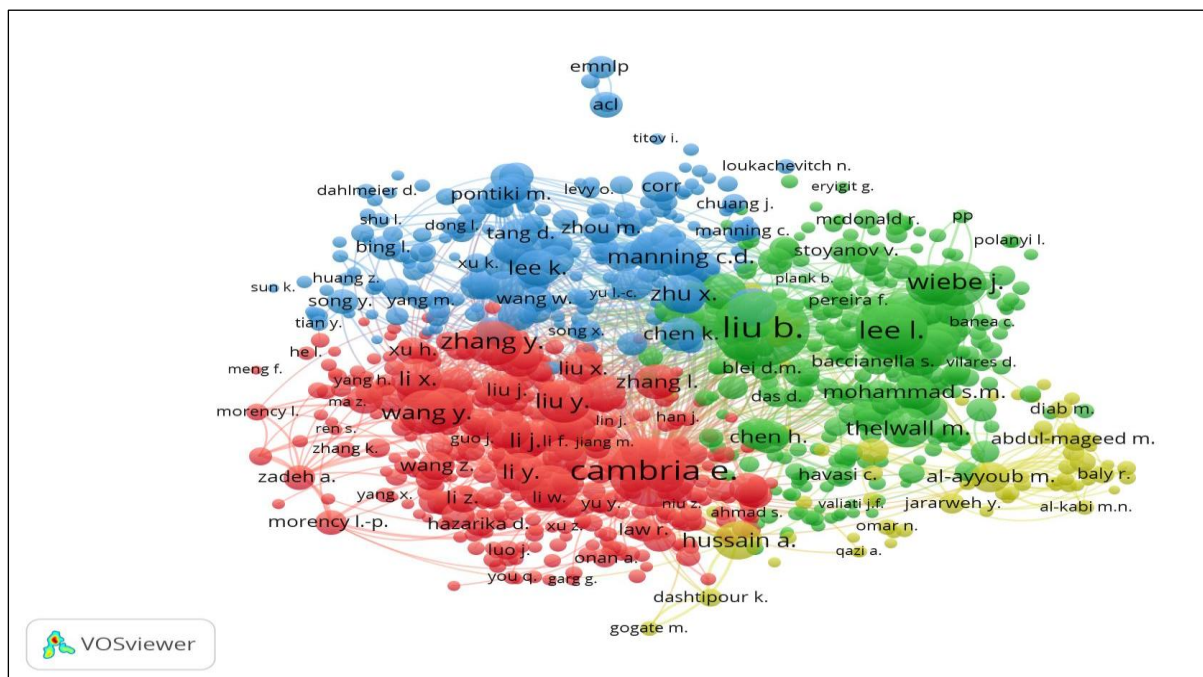


Figure 7. Co-citation network visualisation of cited authors

5. Discussion

The present bibliometric analysis provides a comprehensive understanding of the intellectual structure, collaboration patterns, and thematic evolution of sentiment analysis research. The findings reveal a dynamic and rapidly evolving field shaped by interdisciplinary contributions, technological advancements, and global research collaborations. The co-authorship analysis highlights that prolific authors such as Pushpak Bhattacharyya and Asif Ekbal demonstrate strong collaborative networks, indicating the importance of teamwork in advancing sentiment analysis research. At the same time, highly cited authors like Erik Cambria, Soujanya Poria, and Preslav Nakov exhibit significant scholarly impact despite relatively fewer publications, suggesting that quality and innovation outweigh mere productivity. This aligns with the broader trend in scientific research where impactful contributions often define the direction of a field rather than publication volume.

Institutional analysis reveals that Asian institutions, particularly from India, China, and Singapore, dominate research output. However, institutions such as Harvard University show higher citation impact, reflecting the continued influence of Western academia in shaping foundational research. The relatively lower total link strength for some institutions suggests that institutional collaborations remain fragmented and could benefit from stronger international partnerships. At the country level, China and India lead in terms of research productivity, while the United States stands out in citation impact, indicating its leadership in high-quality and influential research. European countries such as the United Kingdom and Germany exhibit strong collaboration networks, demonstrating a balanced approach between productivity and international cooperation. These findings highlight a shift toward a more globally distributed research ecosystem, where emerging economies play a critical role in knowledge production while established nations maintain intellectual leadership.

The keyword co-occurrence analysis underscores the interdisciplinary nature of sentiment analysis. Core themes such as “sentiment analysis,” “social media,” and “data mining” dominate the field, reflecting its strong application in analysing online user-generated content. The prominence of “machine learning,” “deep learning,” and “computational linguistics” indicates a clear methodological shift toward AI-driven approaches. This transition is further supported by the co-citation analysis, which identifies three major research clusters: traditional lexicon-based methods, machine learning approaches, and advanced deep learning techniques. The co-citation results reveal that foundational works by Pang and Lee (2008) and Liu (2012) continue to play a central role in shaping the field, serving as theoretical anchors for subsequent research. At the same time, the emergence of deep learning models such as LSTM and BERT signifies a paradigm shift toward more sophisticated and accurate sentiment analysis techniques.

This evolution reflects the broader trend in natural language processing, where transformer-based models are increasingly replacing traditional approaches. Furthermore, the analysis of cited sources demonstrates that journals focusing on artificial intelligence, machine learning, and applied domains are highly influential. The presence of interdisciplinary journals such as *Sustainability* and *Tourism Management* highlights the expanding application of sentiment analysis beyond computer science into social sciences, business, and environmental studies. This diversification indicates that sentiment analysis is no longer confined to technical domains but is increasingly used for decision-making and policy analysis. The findings suggest that sentiment analysis research is characterized by strong growth, increasing interdisciplinarity, and a shift toward advanced AI techniques. However, the study also identifies gaps in collaboration and the need for greater integration between research communities across regions and disciplines.

6. Conclusion

This study presents a comprehensive bibliometric analysis of sentiment analysis research, focusing on co-authorship networks, co-citation structures, and keyword trends. The results demonstrate that the field has experienced substantial growth, driven by advancements in artificial intelligence, machine learning, and natural language processing. The analysis reveals that while countries such as China and India lead in research productivity, the United States remains the most influential in terms of citation impact. Influential authors and foundational works continue to shape the intellectual structure of the field, while emerging researchers contribute

to its ongoing evolution. The transition from traditional lexicon-based approaches to deep learning and transformer-based models marks a significant paradigm shift in sentiment analysis methodologies.

The study also highlights the increasing interdisciplinary of sentiment analysis, with applications extending into diverse fields such as business, tourism, and sustainability. This broad applicability underscores the importance of sentiment analysis as a tool for understanding human behaviour and supporting data-driven decision-making. Despite these advancements, the study identifies the need for stronger international collaborations and more integrated research efforts to enhance knowledge sharing and innovation. Future research should focus on emerging areas such as multilingual sentiment analysis, multimodal data integration, and ethical considerations in AI-driven sentiment analysis. This bibliometric exploration provides valuable insights into the evolution, structure, and future direction of sentiment analysis research. It serves as a useful reference for researchers, academicians, and policymakers aiming to understand the trends and opportunities within this rapidly growing field.

References

1. Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975.
2. Cambria, E., Poria, S., Gelbukh, A., & Thelwall, M. (2017). Sentiment analysis is a big suitcase. *IEEE Intelligent Systems*, 32(6), 74–80.
3. Devlin, J., Chang, M. W., Lee, K., & Toutanova, L. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171–4186).
4. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of Business Research*, 133, 285–296.
5. Liu, B. (2012). *Sentiment analysis and opinion mining*. Synthesis Lectures on Human Language Technologies, 5(1), 1–167.
6. Liu, B. (2020). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
7. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113.
8. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
9. Van Eck, N. J., & Waltman, L. (2014). Visualising bibliometric networks. In *Measuring scholarly impact* (pp. 285-320). Springer.
10. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253. <https://doi.org/10.1002/widm.1253>