The Current State of Leadership Research: Identifying Key Trends through Natural Language Processing and Machine Learning Analysis

Marco Ferreira Ribeiro[1,*] , Carla Gomes da Costa[2] and Filipe R. Ramos[3]
[1]Faculty of Social Sciences, Universidade Europeia, 1500-210 Lisbon, Portugal
[2]CETRAD, Universidade Europeia, 1500-210 Lisbon, Portugal
[3]CEAUL—Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, 1649-004 Lisbon, Portugal

ferreiraribeiro.marco@gmail.com
carla.costa@universidadeeuropeia.pt
frramos@ciencias.ulisboa.pt

Abstract: The impact of leadership on organizational performance and success is a well-established theme, leading to a significant expansion in related scientific studies. This has made identifying and understanding the current state of literature challenging. Most studies focus on specific aspects of leadership, offering a fragmented view of the field. This study employs Natural Language Processing (NLP) techniques and a machine learning approach to analyze 19,429 research papers from Web of Science and Scopus. We aimed to identify and interpret current trends in leadership research, providing a comprehensive overview of the leadership domain, highlighting five key research trends: (1) Leadership and Digital Transformation Research (LDTR); (2) Leadership and Organizational Performance Research (LOPR); (3) Educational Leadership Research (ELR); (4) Leadership Practices and Development Research (LPDR); and (5) Gender and Diversity Leadership Research (GDLR). This overview not only delineates current trends but also aids researchers in identifying significant gaps in existing literature, thereby facilitating the development of innovative research projects and initiatives.

Keywords: Leadership; Research trends; Research gaps; Natural Language Processing (NLP); Latent Dirichlet Allocation (LDA) Analysis.

1. Introduction

The recognition of leadership's significance in organizations has become increasingly salient within the rapidly evolving global context (Fareed *et al.*, 2022; Piwowar-Sulej and Iqbal, 2023). This recognition is particularly pertinent in light of the challenges posed by globalization, technological advancements, and socio-cultural shifts (Sum, 2022), all of which accentuate the critical role of effective leadership in enhancing organizational performance (Hilton *et al.*, 2023; Piwowar-Sulej and Iqbal, 2023). Consequently, leadership is not only acknowledged as a pivotal element for organizational success (Fareed *et al.*, 2022; Piwowar-Sulej and Iqbal, 2023) but is also increasingly valued across various organizational domains (Fladerer and Braun, 2020), leading to a significant increase in studies on leadership.

However, the increasing volume of leadership studies has made identifying and understanding the current state of the literature challenging. Most studies focus on specific aspects of leadership, offering a fragmented view of the field. This fragmentation is exacerbated by the diversity of theoretical perspectives and methodological approaches employed by researchers and the emergence of new topics and areas of investigation (Badura *et al.*, 2022; Oc *et al.*, 2023), which complicates efforts to understand the current state of leadership research.

This study employs Natural Language Processing (NLP) techniques and a machine learning approach to identify and analyze research trends in leadership research. By managing and interpreting extensive large volumes of data, this study provides a comprehensive overview of the current state of research in this field. This overview not only delineates current trends but also aids researchers in identifying significant gaps in existing literature, thereby facilitating the development of innovative research projects and initiatives.

2. Methodology

We conducted a comprehensive search using Scopus and Web of Science about the keyword “leadership”, excluding literature reviews and conference proceedings. 133078 peer-reviewed research publications were found by this search. A total of 42847 research publications were found to be duplicated. To find research publications unrelated to leadership (e.g.: price leadership, market leadership, fashion leadership, supply chain leadership, brand leadership, technological leadership, and others), a manual human review was carried out, which led to removing 33 684 research papers, improving the quality of the data. A total of 56547 research

papers were obtained. To identify trends, we restricted the study to research papers published from 2021 to 2023, which corresponded to 19429 research papers. This study used research paper abstracts, which are easily accessible, provide a succinct synopsis of main ideas, and facilitate preprocessing and NLP analysis due to their standard format (Kim and Kim, 2022).

Figure 1 illustrates the overview of the steps processed. Key Python packages used included NLTK for preprocessing (Bird *et al.*, 2009) and Gensim for applying LDA (Rehurek and Sojka, 2010). Preprocessing the data to ensure suitability for unsupervised machine learning is essential before implementing the Latent Dirichlet Allocation (LDA) algorithm. Preprocessing involved lowercasing, lemmatization, stop word removal, tokenization, n-grams evaluation, and removal of numerals and punctuation. Pointwise Mutual Information (PMI) scores were calculated for n-grams to select the most relevant ones. An LDA function was created and a grid search approach was employed to optimize LDA parameters, evaluating the number of passes, topics, and hyperparameters alpha and beta. The coherence scores (C_v) and perplexity scores (Ps) were used to choose the best model.

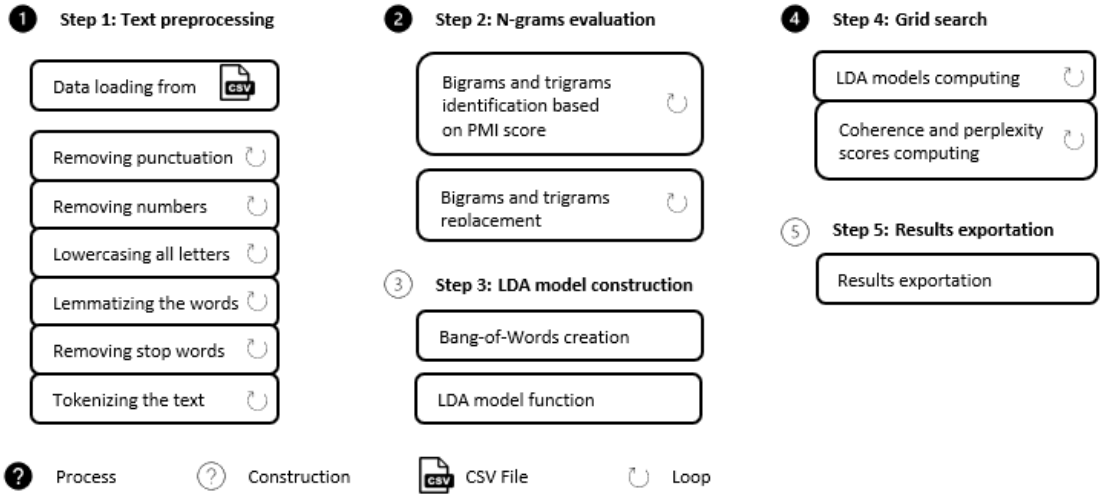

Figure 1: Study method overview.

3. Findings

A wide range of set values was covered by these grid search parameters, and 675 distinct potential LDA models were generated. The coherence and perplexity scores for the combination of the number of trends and passes are displayed in Figure 2.

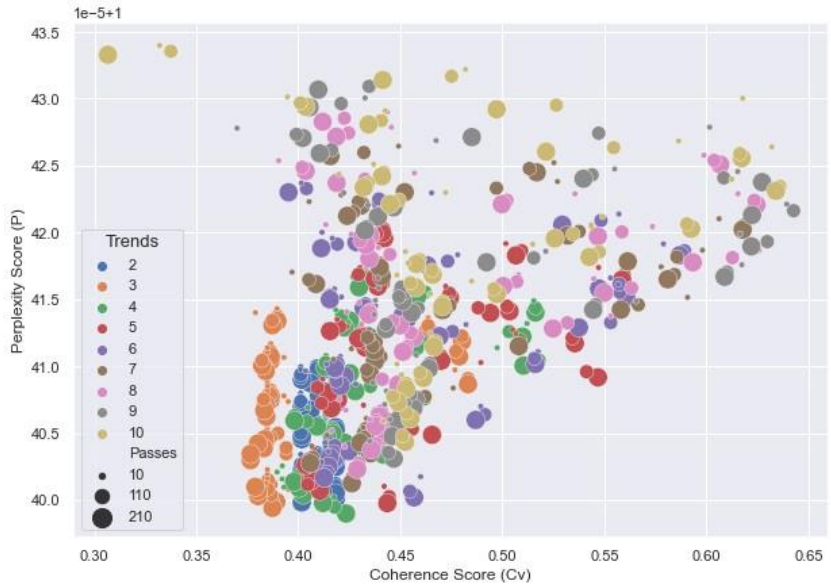

Figure 2: Coherence and Perplexity Scores by Number of Trends and Passes.

In selecting the LDA model, the 9-trend model (K=9) had the highest coherence score (Cv=0.643), while the 2-trend model (K=2) had the lowest perplexity score (Ps=43.399e-5 + 1), indicating statistical robustness. However, a 5-trend model (K=5; Cv=0.503; Ps= 41.863e-5 + 1) was chosen for its balance between topic coherence and model perplexity, ensuring coherent and interpretable topics without oversimplification or excessive complexity. The term-topic matrix for the 5-trend model is presented in Table 1.

Table 1: Term-topic matrix

Trend Number	Terms expressed as lemmas									
1	Digital	Digital Transformation	Digital Technology	Charismatic	Strategizing	Century	Sometimes	Text	Chief Executive	Discursive
	(0.004)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
2	Leadership	Study	Employee	Leader	Relationship	Performance	Behavior	Organizational	Effect	Organization
	(0.039)	(0.026)	(0.018)	(0.013)	(0.011)	(0.010)	(0.008)	(0.007)	(0.007)	(0.007)
3	School	Teacher	Student	Principal	Learning	Instructional	Educational	Teaching	School Principal	Reform
	(0.045)	(0.038)	(0.020)	(0.015)	(0.014)	(0.006)	(0.005)	(0.005)	(0.004)	(0.003)
4	Leadership	Leader	Practice	Study	Change	Development	Approach	Process	Experience	Management
	(0.028)	(0.013)	(0.009)	(0.007)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)
5	Woman	Female	Gender	Position	Diversity	Career	Sport	Coach	Athlete	Academic
	(0.018)	(0.010)	(0.010)	(0.007)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.003)

Note: The terms are listed according to their decreasing probability of occurring in the trend k.

Based on Table 1, the results suggest the following research trend topics in the leadership field.

Trend 1: Leadership and Digital Transformation Research (LDTR) – This trend focuses on the relationship between leadership and digital technology in organizations. Keywords like "digital" and "digital transformation" indicate research on how leadership adapts to digitalization, addressing its benefits and challenges. The terms "charismatic" and "strategizing" highlight leaders' personal attributes and strategies in the digital landscape.

Trend 2: Leadership and Organizational Performance Research (LOPR) – This trend examines the connection between leadership and organizational performance, using keywords like "performance," "organizational," and "leadership." Studies focus on how leadership impacts employee performance and organizational success, aiming to improve overall performance and cultivate strong work relationships.

Trend 3: Educational Leadership Research (ELR) – This trend highlights the role of leadership in educational environments, with keywords like "school," "teacher," "student," and "principal." Research explores how school leaders influence student performance, teaching quality, and educational reforms, emphasizing leadership's impact on learning outcomes and the educational process.

Trend 4: Leadership Practices and Development Research (LPDR) – This trend investigates the implementation and development of leadership, with keywords like "leadership," "practice," and "development." It delves into leadership tactics, the evolution of methods, and the processes of enhancing leadership competencies, focusing on adaptive leadership and change management.

Trend 5: Gender and Diversity Leadership Research (GDLR) – This trend explores gender and diversity in leadership, using keywords like "woman," "female," "gender," and "diversity." It examines the obstacles and opportunities for women and diverse groups in leadership roles, particularly in sectors like sports, highlighting the importance of inclusivity and the benefits of diverse leadership.

4. Discussion

The increasing complexity and globalization of organizations have highlighted the need for effective leadership, leading to substantial growth in leadership research with new areas and research topics (Badura *et al.*, 2022; Oc *et al.*, 2023). We identified five key trends: LDTR, LOPR, ELR, LPDR, and GDLR. Each trend represents unique research interests and offers directions for future investigation.

The LDTR trend focuses on leadership's role in digital transformation, a niche yet critical area given the rapid adoption of digital technologies. Future research could explore how traditional leadership styles adapt to digital contexts and develop new competencies for managing digitally mediated environments. The LOPR trend emphasizes the link between leadership practices and organizational performance. Future research could

expand to include destructive leadership behaviors and their negative impacts. The ELR trend focuses on the impact and role of leadership within educational settings. Future research could explore how different leadership styles influence student engagement and success, and develop new practices tailored to educational contexts. The LPDR trend highlights leadership practices and development. Future research could investigate the mechanisms of leadership development and how competencies are acquired and refined over time. Additionally, the GDLR trend addresses diversity in leadership, highlighting the barriers faced by underrepresented groups and the benefits of diverse leadership teams (Zattoni *et al.*, 2023). Intersectionality and the experiences of marginalized leaders are key areas for further study.

5. Conclusion

The identified research trends in leadership studies highlight the current dynamics of leadership in response to technological advancements, shifting organizational structures, and societal expectations, indicating a critical juncture for addressing contemporary organizational challenges. These trends emphasize the need for adaptive, innovative leadership practices to navigate modern workplace complexities. Future research should integrate these diverse insights to foster a holistic understanding of leadership, exploring synergies and interdependencies among the trends for more robust findings.

Some limitations can be pointed out in our study. This research's reliance on Web of Science and Scopus may exclude pertinent studies from other databases, and the use of the LDA algorithm might result in less accurate topic classifications, missing subtle meanings discernible through human analysis, that would not be possible given the volumes of data analyzed.

References

Badura, K.L., Galvin, B.M. and Lee, M.Y. (2022), "Leadership emergence: An integrative review", *The Journal of Applied Psychology*, J Appl Psychol, Vol. 107 No. 11, pp. 2069–2100, doi: 10.1037/APL0000997.

Bird, S., Klein, E. and Loper, E. (2009), *Natural Language Processing with Python - Analyzing Text with the Natural Language Toolkit*, O'Reilly Media, Inc.

Caniëls, M.C.J. (2023), "How remote working increases the importance of positive leadership for employee vigor", *Frontiers in Psychology*, Vol. 14, pp. 1–10, doi: 10.3389/fpsyg.2023.1089557.

Fareed, M.Z., Su, Q., Almutairi, M., Munir, K. and Fareed, M.M.S. (2022), "Transformational leadership and project success: The mediating role of trust and job satisfaction", *Frontiers in Psychology*, Vol. 13, pp. 1–14, doi: 10.3389/fpsyg.2022.954052.

Fladerer, M.P. and Braun, S. (2020), "Managers' resources for authentic leadership – A multi-study exploration of positive psychological capacities and ethical organizational climates", *British Journal of Management*, Vol. 31, pp. 325–343, doi: 10.1111/1467-8551.12396.

Hilton, S.K., Madilo, W., Awaah, F. and Arkorful, H. (2023), "Dimensions of transformational leadership and organizational performance: The mediating effect of job satisfaction", *Management Research Review*, Vol. 46 No. 1, pp. 1–19, doi: 10.1108/MRR-02-2021-0152.

Kim, M. and Kim, D. (2022), "A Suggestion on the LDA-Based Topic Modeling Technique Based on ElasticSearch for Indexing Academic Research Results", *Applied Sciences*, Vol. 12 No. 6, p. 3118, doi: 10.3390/app12063118.

Oc, B., Chintakananda, K., Bashshur, M.R. and Day, D. V. (2023), "The study of followers in leadership research: A systematic and critical review", *The Leadership Quarterly*, Elsevier Inc., Vol. 34 No. 1, pp. 1048–9843, doi: 10.1016/j.leaqua.2022.101674.

Piwowar-Sulej, K. and Iqbal, Q. (2023), "Leadership styles and sustainable performance: A systematic literature review", *Journal of Cleaner Production*, Elsevier Ltd, Vol. 382, pp. 1–14, doi: 10.1016/j.jclepro.2022.134600.

Rehurek, R. and Sojka, P. (2010), "Software Framework for Topic Modelling with Large Corpora", *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, ELRA, Valletta, Malta, pp. 45–50.

Sum, N. (2022), "School leaders' perceptions of their roles during the pandemic: an Australian case study exploring volatility, uncertainty, complexity and ambiguity (VUCA leadership)", *School Leadership & Management*, Vol. 42 No. 2, pp. 188–207, doi: 10.1080/13632434.2022.2045268.

Zattoni, A., Leventis, S., Van Ees, H. and De Masi, S. (2023), "Board diversity's antecedents and consequences: A review and research agenda", *Leadership Quarterly*, Elsevier Inc., Vol. 34 No. 1, p. 101659, doi: 10.1016/j.leaqua.2022.101659.