

Preface to the Workshop of the 2nd Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (JCDL2024)

Zhongyi Wang¹, Haihua Chen², Chengzhi Zhang³, Yi Bu⁴, Wei Lu⁵, Jian Wu⁶

¹ Central China Normal University, Guizishan, Wuhan, 430079, China

² University of North Texas, Denton, Texas, 76201, USA

³ Nanjing University of Science and Technology, No. 200, Xiaolingwei, Nanjing, 210094, China

⁴ Peking University, Yiheyuan, Beijing, 100091, China

⁵ Wuhan University, Luojiashan, Wuhan, 430072, China

⁶ Old Dominion University, Norfolk, Virginia, 23529, USA

Abstract

The Workshop of the 2nd Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (JCDL2024; <https://imsc-committee.github.io>) was held in Hong Kong, China and online. This workshop focuses on the discussion for innovation measurement and produces enlightening outcomes. We will engage broad audiences to share their ideas and pre-productions, enabling an interdisciplinary approach to exploring frontier areas. The workshop features a comprehensive agenda, including keynotes from leading experts, oral presentations showcasing cutting-edge research, and poster sessions for in-depth discussions. The primary topics covered in the proceedings encompass the definitions for innovation and methods for its measurement, as well as the applications of innovation measurement.

Keywords

Innovation novelty, Disruption, Scientific communication, Academic data mining, Digital library

1. Introduction

Innovations drive scientific and technical advancements. Measuring and tracking innovations is a core task in informetrics and the Science of Science. In the big data era, opportunities and challenges arise as the sheer volume of published papers complicates the identification of truly novel work. Advances in artificial intelligence, particularly in natural language processing and knowledge reasoning, provide promising solutions.

Numerous methods have been developed to measure scientific innovation. For instance, Wang et al. evaluated the quality and novelty of research papers using a combination of coarse features, knowledge entity networks, and semantic similarity analysis [1,2]. Bu et al. quantified the scientific novelty of doctoral dissertations with a combinatorial approach, incorporating the pre-trained Bio-BERT model [3]. Zhao et al. identified that teams with more thought leaders tended to generate less disruptive ideas [4]. Luo et al. introduced a novel approach to measure scientific novelty by examining the interplay of research questions and methods, employing life-index and semantic similarity metrics [5]. Wu et al. analyzed scientific collaboration through a cost-benefit perspective, highlighting its complex dynamics [6]. However, there is still a lack of unified frameworks, high-quality datasets,

Joint Workshop of the 2th Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (IMSC2024), Dec 20th, 2024, Hong Kong, China and Online

Corresponding author.

✉ wzywzy13579@163.com (Zhongyi Wang); haihua.chen@unt.edu (Haihua Chen); zhangcz@njust.edu.cn (Chengzhi Zhang); buyi@pku.edu.cn (Yi Bu); weilu@whu.edu.cn (Wei Lu); jwu@cs.odu.edu (Jian Wu)



© 2024 Copyright 2024 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

effective methods, and robust practical applications. To address this gap, we propose this workshop to convene researchers and practitioners to establish a collaborative platform for exchanging ideas, sharing pilot studies, and defining future directions in this cutting-edge field.

The Workshop of the 2nd Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data (JCDL2024) was held in Hong Kong, China and online on December 20th, 2024. This workshop aims to engage the research community in exploring key issues related to the definition and measurement of innovation in scientific communication, with a focus on big data analytics. The goal is to bridge theoretical and practical gaps, offering a comprehensive understanding of innovation definitions and methodologies for quantifying innovation across disciplines. The workshop will examine AI-empowered informetric models and big data techniques designed to improve the robustness, scalability, and adaptability of existing measurement methods. It will also explore the applications of innovation measurement, from assessing the impact of scientific publications to identifying emerging trends and breakthroughs. By integrating advanced data-driven approaches with established models, the workshop seeks to enhance the interpretability of innovation measurement, ensuring these technologies meet practical needs in real-world applications. This collaborative effort aspires to provide innovative solutions to the challenges of measuring scientific innovation in today's data-driven research landscape.

2. Overview of the papers

This workshop received 22 submissions for peer review, and accepted 11 papers, which are collected in this proceeding. It includes 2 long papers, 5 short papers, and 4 power talks. The workshop also featured one keynote talk on *Team: Power, Leadership, and Diversity*.

All contributions and slides in the workshop are available on the IMSC2024 workshop website <<https://imsc-committee.github.io/JCDL2024-IMSCworkshop/paper/>>. The following section provides a brief overview of the keynote and the 11 accepted submissions.

2.1 Keynote

The keynote in this workshop highlights the dynamics of team power, leadership, and diversity in driving innovation and success.

Professor Ying Ding is Bill & Lewis Suit Professor at School of Information, and adjunct Professor at Department of Population Health at Dell Medical School, University of Texas at Austin. She delivered a keynote on *Team: Power, Leadership, and Diversity*.

Prof. Ying Ding explored the dynamics of team formation and performance, emphasizing the critical role of power, leadership, and diversity in driving innovation. With extensive expertise in AI, knowledge graphs, and data-driven science, she led pioneering research at the intersection of health, informatics, and team collaboration. In her keynote, she unveiled insights from a large-scale study analyzing millions of teams across academic fields, shedding light on the factors that contributed to team success. She highlighted the importance of flat team structures, heterogeneous shared leadership, and rich diversity as key pathways for fostering high-performing teams. Her work demonstrated how understanding team dynamics empowered leaders, researchers, and professionals to cultivate effective, innovative, and resilient teams.

2.2 Research papers and posters

We organized the 11 submissions in the following two sections.

2.2.1 Session 1: Definitions for innovation and methods for its measurement

This session includes five papers.

In their paper “What is Academic Innovation: A Concept Analysis”, Li et al. [7] analyzed the concept of academic innovation by identifying its antecedents, attributes, and consequences. They found that academic innovation stems from new knowledge combinations, is defined by novelty, value, contextuality, and cumulativeness, and results in knowledge creation and paradigm shifts, while distinguishing it from related concepts.

In the paper “Identifying Emerging Topics in Specific Domains via Novelty Analysis of Entities in Future Work Sentences from Academic Articles”, Yang et al. [8] identified emerging NLP topics by analyzing future work sentences, assessing entity novelty, and filtering key research directions. They found that optimizing and applying pre-trained language models is a significant trend.

In their paper “Entity-Citation-Driven Academic Impact Measurement in Scientific Papers”, Gao et al. [9] proposed an entity-citation-driven approach to measuring academic impact by analyzing citations referring to specific knowledge entities within a paper. Using knowledge entity recognition and citation detection, they demonstrated that entity citations enhance the precision and interpretability of impact evaluation, improving ranking in knowledge retrieval.

The last paper in this section is by Qiu and Li [10], “Research on Paper Semantic Novelty Measurement Based on Large Language Model”, they proposed a semantic novelty measurement model for scientific papers using a large language model to generate question and method words. Enhanced by LoRA and prompts, the model achieved high precision and recall, proving effective and robust, with optimal cost-effectiveness at 3,000 training samples.

2.2.2 Session 2: Applications of innovation measurement

This session includes six papers.

The work by Ren et al. [11], entitled “Novelty Assessment of Chinese Academic Articles in Information Resources Management: A Comparison of Knowledge Entity and Reference-Based Methods” assessed the novelty of Chinese academic articles in Information Resources Management (IRM) using fine-grained knowledge entities and references. Their findings show that entity-based novelty scores are generally lower than reference-based ones, with both being skewed. Themes like University Libraries and Bibliometrics exhibit high novelty across both perspectives.

In the paper “Freshness and Informativity Weighted Cognitive Extent and Its Correlation with Cumulative Citation Count”, Wang and Wu [12] introduced Freshness and Informativity Weighted Cognitive Extent (FICE), refining cognitive extent with lifetime ratio and informativity of scientific entities. They modeled entity lifetimes using Gaussian profiles and found that FICE strongly correlates with cumulative citation counts, confirming prior trends in scientific entity growth.

In the paper “Relationship Between Paper Authorship Roles and Novelty from a Gender Perspective: Evidence from 81,137 PLOS ONE Articles”, Zeng et al. [13] analyzed 81,137 PLOS ONE articles to explore the relationship between authorship roles, gender, and paper novelty. They found that greater participation in writing and software development correlates with higher novelty. Women contribute more to investigation and data curation, while men dominate supervision and funding. Novelty links to methodology and visualization only for male authors.

In “Revealing the Research Deviation of AI Research Between China and the U.S.”, Sun and Chen [14] analyzed AI research differences between China and the U.S. using co-occurrence and vector semantic fields. They identified distinct research focuses and content preferences, providing insights into research distribution and potential collaboration opportunities between the two nations.

The paper by Zhang and Chen [15], entitled “Technology Topic Evolution from the Perspective of Patent Validity”, analyzed technology topic evolution by categorizing patents based on legal status. They proposed a two-dimensional evolution trajectory to distinguish between maturing, saturated, and lagging technologies. Experimental validation in 3D printing confirmed the effectiveness of this approach.

In the paper “Can the strength of co-citation linkages be evaluated using context-aware citation network embeddings?”, Eto [16] proposed a context-aware citation network embedding method to evaluate co-citation linkage strength within a single citing document. Experimental results show that this approach outperforms traditional co-citation techniques, effectively distinguishing between weak and strong co-citation linkages.

3. Outlook and further reading

The IMSC2024 workshop series have been highly successful and garnered substantial attention from the research communities. This workshop series has made significant contributions to the literature by introducing innovative technological advancements and valuable empirical insights.

Acknowledgements

Zhongyi Wang acknowledges National Social Science Foundation of China (Grant Number 22BTQ102) and Chengzhi Zhang acknowledges National Natural Science Foundation of China (Grant No. 72074113).

References

- [1] Wang, Z., Zhang, H., Chen, H., Feng, Y., & Ding, J. (2024). Content-based quality evaluation of scientific papers using coarse feature and knowledge entity network. *Journal of King Saud University-Computer and Information Sciences*, 102119. Elsevier. <https://doi.org/10.1016/j.jksuci.2024.102119>
- [2] Wang, Z., Qiao, X., Chen, J., Li, L., Zhang, H., Ding, J., & Chen, H. (2024). Exploring and evaluating the index for interdisciplinary breakthrough innovation detection. *The Electronic Library*, 42(4), 536-552. <https://doi.org/10.1108/EL-06-2023-0141>
- [3] Liu, M., Xie, Z., Yang, A. J., Yu, C., Xu, J., Ding, Y., & Bu, Y. (2024). The prominent and heterogeneous gender disparities in scientific novelty: Evidence from biomedical doctoral theses. *Information Processing & Management*, 61(4), 103743. Elsevier. <https://doi.org/10.1016/j.ipm.2024.103743>
- [4] Zhao, Y., Wang, Y., Zhang, H., Kim, D., Lu, C., Zhu, Y., & Zhang, C. (2024). Do more heads imply better performance? An empirical study of team thought leaders' impact on scientific team performance. *Information Processing & Management*, 61(4), 103757. Elsevier. <https://doi.org/10.1016/j.ipm.2024.103757>
- [5] Luo, Z., Lu, W., He, J., & Wang, Y. (2022). Combination of research questions and methods: A new measurement of scientific novelty. *Journal of Informetrics*, 16(2), 101282. Elsevier. <https://doi.org/10.1016/j.joi.2022.101282>
- [6] Wu, L., Yi, F., Bu, Y., Lu, W., & Huang, Y. (2024). Toward scientific collaboration: A cost-benefit perspective. *Research Policy*, 53(2), 104943. Elsevier. <https://doi.org/10.1016/j.respol.2023.104943>
- [7] Li, S., Chen, D., & Hua, B. (2024). What is Academic Innovation: A Concept Analysis. *Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data*, Hong Kong, China and online.
- [8] Yang, Y., Xiang, Y., & Zhang, C. (2024). Identifying Emerging Topics in Specific Domains via Novelty Analysis of Entities in Future Work Sentences from Academic Articles. *Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data*, Hong Kong, China and online.

- [9] Gao, X., Chen, C., Li, W., & He, Y. (2024). Entity-Citation-Driven Academic Impact Measurement in Scientific Papers. Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data, Hong Kong, China and online.
- [10] Qiu, X., & Li, J. (2024). Research on Paper Semantic Novelty Measurement Based on Large Language Model. Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data, Hong Kong, China and online.
- [11] Ren, Y., Yang, C., Zhao, Y., Zhang, H., & Zhang, C. (2024). Novelty Assessment of Chinese Academic Articles in Information Resources Management: A Comparison of Knowledge Entity and Reference-Based Methods. Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data, Hong Kong, China and online.
- [12] Wang, Z., & Wu, J. (2024). Freshness and Informativity Weighted Cognitive Extent and Its Correlation with Cumulative Citation Count. Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data, Hong Kong, China and online.
- [13] Zeng, J., Zhao, Y., & Zhang, C. (2024). Relationship Between Paper Authorship Roles and Novelty from a Gender Perspective: Evidence from 81,137 PLOS ONE Articles. Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data, Hong Kong, China and online.
- [14] Sun, H., & Chen, G. (2024). Revealing the Research Deviation of AI Research Between China and the U.S. Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data, Hong Kong, China and online.
- [15] Zhang, J., & Chen, R. (2024). Technology Topic Evolution from the Perspective of Patent Validity. Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data, Hong Kong, China and online.
- [16] Eto, M. (2024). Can the Strength of Co-Citation Linkages Be Evaluated Using Context-Aware Citation Network Embeddings? Proceeding of the 2nd Workshop on Innovation Measurement for Scientific Communication (IMSC) in the Era of Big Data, Hong Kong, China and online.