The evolution of Influence maximizationstudies: A scientometric analysis

Apparao Devineni [1], Edukondalu[2], V Subhashini[3], P V Lakshmi Durga[4] ,M Srikanth Reddy[5]

[1]Department of CSE-AIML,Joginpally B R Engineering College, Hyderabad, India, 500075.
[2]Department of CSE, Shadan College of Engineering and Technology, Hyderabad ,India,500086.
[3]Department of CSE-DS, J B Institute of Engineering and Technology, Hyderabad, India 500075.
[4]Department of CSE, Nimra College of Engineering and Technology, Vijayawada, India 521456.
[5]Department of Information Technology, Joginpally B R Engineering College, Hyderabad, India, 500075.

ABSTRACT

Recent decades have witnessed a surge in research on influence maximization, yet rigorous investigation in this area remains limited. This study, analyzing Scopus-indexed publications from 2005 to 2024, reveals a growing research output, with China, the United States, Australia, and Singapore as key contributors. Chinese researchers dominate influential authors, with collaborations prominent between these nations and countries like the United States and India. Chen emerges as the most influential author. "Social networking" is a significant and emerging area of study, emphasizing the importance of understanding information diffusion and identifying key players within networks. This research aims to provide a comprehensive overview of the available literature on influence maximization for researchers in the field.

Keywords: Influence maximization,Scientometric analysis,information dissemination,community detection,Heuristic methods.

1 Introduction

Many complex networks[1], such as transportation systems, information networks, and social networks, are integral to our daily lives and can be modeled for analysis. Understanding and managing these networks is essential for societal progress. These complex systems are not only valuable for scientific research but also for data representation. Recently, studying the micro-level features of networks—such as vertices and edges formed by numerous interconnected nodes with diverse and complex interactions—has gained significant attention in network science. A social network can be represented as the graph G=(V,E), where V is the set of individuals or entities, and E defines the connections between them. When two nodes are connected socially, whether through friendship, follower relationships, or professional ties, an edge is drawn between them. Identifying influential nodes within social networks is a key area of focus in network science, with applications in areas such as marketing, revenue maximization[2], public health[3], and opinion management.

The rise of social media platforms has significantly sped up the flow of information and news in the digital age. The exploration of information spread was advanced by Domingos and Richardson's (2001) [4] application of the Markov Random Field (MRF) for simulation. A critical issue in information dissemination is influence maximization, which was first framed as a discrete optimization problem by Kempe et al. (2003) [5]. They proposed a near-optimal solution that achieved 63% of the optimal result, opening the door for further research on influence maximization. The topic has since attracted increasing scholarly attention, being applied in areas such as rumor control (Yang et al., 2020b), community detection (Li et al., 2017) [6], and word-of-mouth advertising (Li et al., 2018) [7]. Seed mining algorithms have become a primary research focus, as it's commonly believed that identifying the best seed node can maximize benefits. However, the chosen seeds do not always guarantee the desired results.

Section 2 presents the relevant area literature after the introductory section. Subsequently, in Section 3 discussed the Research Method and Data Collection. The findings of the scientometric study are presented in Section 4. In Section 5, draw final conclusions.

2 Related works

Until recently, the Influence Maximization (IM) problem [8][9] primarily focused on regular networks. Solutions to the IM problem can be classified into several categories, including approximation algorithms, heuristics, and community-based methods. In 2003, Kempe et al. [5] introduced the greedy algorithm, which guarantees an approximation rate of $(1-\frac{1}{e}-\in)$ for selecting seed nodes. However, the algorithm can be inefficient due to its stringent filtering conditions, leading to significant time consumption. As a result, new algorithms have been developed to optimize approximation solutions, balancing both efficiency and effectiveness. Approaches based on diminishing marginal returns, such as CELF and CELF++ [10], were introduced. Despite these improvements, running these algorithms on large-scale networks still requires considerable computational time. In response, more efficient heuristic algorithms, like IRIE by Jung et al. [11], were proposed. To ensure effectiveness in the IC model and its extended IC-N version, IRIE integrates seed influence spread estimation with the influence ranking process. Heuristic techniques based on centrality measures, such as degree and PageRank centrality [12], are also used to address the IM problem by selecting highly connected nodes as seed nodes. Community-based strategies, such as C2IM [13], LKG [14], and INCIM [15], aim to reduce influence overlap among seeds efficiently. Over the past decade, numerous new algorithms have emerged, exploring the IM problem from various perspectives. For example, Li et al. [16] approached the IM problem by considering the emotional responses of crowds, while Kumar et al. [17] examined the issue from a social network perspective, using a label propagation model.

In the early stages of research on influential mining vertices in complex social networks, methods such as degree centrality [18], PageRank [12], eigenvector centrality [19], and K-shell [20][21] were employed to identify key vertices. These node ranking techniques help determine the importance of individual nodes. However, with the advent of the "big data" era, real-world networks have become increasingly complex and difficult to analyze. As a result, research has shifted from assessing the influence of individual nodes to focusing on making a group of nodes as influential as possible, a problem known as influence maximization, which is NP-Hard [5]. To address this challenge, researchers have introduced four key methodologies: (i) simulation-based, (ii) heuristic, and (iii) community-based approaches.

The simulation-based method[22][23] is not well-suited for large-scale real-world networks due to the high demands of exhaustive Monte Carlo simulations. However, these methods can offer higher-quality solutions. To address the computational challenges and avoid repeated Monte Carlo simulations, heuristic methods[24][25][26][11][7] were introduced. While these approaches improve scalability and efficiency, they tend to sacrifice solution quality[23]. Community-based algorithms[27][28][29] are typically faster than traditional greedy algorithms. However, evaluating the marginal gain of a node within its community still requires Monte Carlo simulations, which can be time-consuming and limit the applicability of these algorithms in large-scale networks.

Community-based strategies have been developed to enhance productivity and scalability in tackling the IM problem. Numerous studies[30][13][31][32] have focused on uncovering the underlying community structure within social networks. These methods primarily emphasize community integrity, based on the assumption that relationships among members are consistent. However, in the context of IM, the level of influence within a community can vary. This highlights the need for a method that can both accurately identify community structures and provide reliable results. Due to these challenges, traditional community detection methods are not sufficient for solving the IM problem. As a result, several community-based approaches have been introduced to address the IM issue[33][34][35].

3 Data collection& Methodology:

Scopus, Google Scholar, and Web of Science are widely used academic databases, each with its own advantages and limitations. According to Orduna-Malea et al. [36], while Google Scholar offers extensive coverage of published works, the quality of its data is often criticized, and academics may need to manually remove duplicates. The data coverage of Scopus and Web of Science is quite similar [37]. This study relies on literature from Scopus, as the data analysis tools employed have produced a more accurate knowledge map using this database.

To address the question, "What are the research clusters, state-of-the-art advancements, and emerging trends in social networks?" this study performs a scientometric analysis of the current literature on Influence Maximization (IM). Drawing from previous studies in fields such as IoT in healthcare [38][39], post-occupancy evaluation [40], and convolutional neural networks[41], this approach allows researchers to better visualize qualitative data, leading to a deeper understanding of the topic.

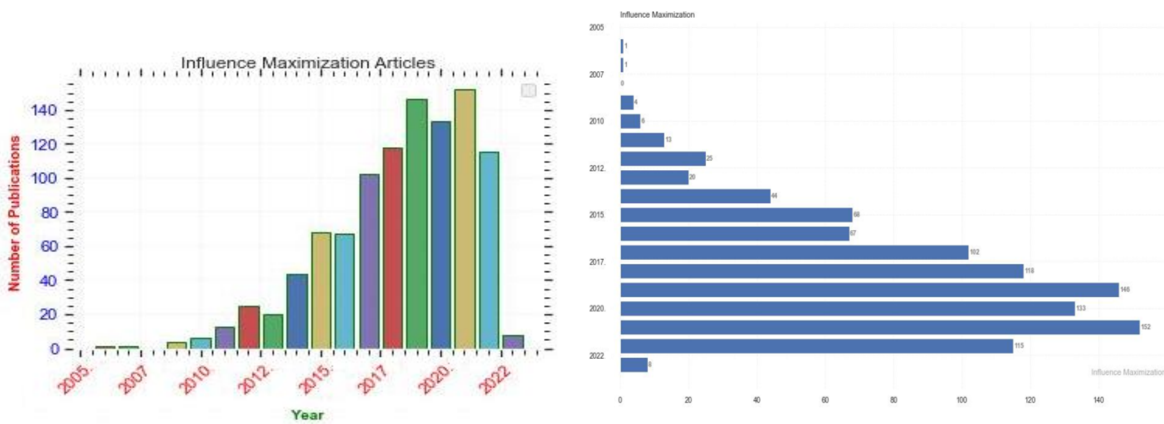


Fig.1(a)Fig. 1(b)

Fig. 1(a),(b). The number of Influence maximization articles in the Scopus database, 2005–2023.

The results of keyword searches conducted in bibliographic databases were used to build a literature database focused on influence maximization. Scopus, the database from Elsevier, was selected for this research due to its user-friendly interface and comprehensive coverage. Scopus is a high-quality resource containing around 84 million records, 26,000 active titles, and 243,400 volumes. It also provides a set of advanced search and analytical tools designed to facilitate future data retrieval and analysis. Additionally, Scopus allows users to export data in multiple formats compatible with popular scientometric analysis software.

The primary goal of this research is to examine the research trends, clusters, and interconnections within the existing literature on influence maximization. As a result, the keywords "influence" and "maximization" were selected for the search. The final query string used for the search was: TITLEABS-KEY (“influence” AND “maximization”).

The literature search was conducted in November 2024, covering articles published from January 2009 to November 2023. Some papers were excluded from the analysis due to missing data, such as unspecified document types and authors. The search returned 1,026 publications from the Scopus database, including conference papers, journal articles, conference reviews, book chapters, and books. To ensure compatibility with the chosen data analysis tool, VOSviewer, the full bibliographic record for each article was exported as a CSV file. This file contained citation details, bibliographical information, author affiliations, abstracts, index keywords, and other relevant data.

VOSviewer, developed by the Centre for Science and Technology Studies at Leiden University, is one of the most widely used tools for bibliometric network generation and visualization. It has been applied in various fields, including Agricultural pollution [42], risk management in construction [43], and climate change adaptation studies [44], for scientometric research. In this study, the literature database on influence maximization was analyzed using VOSviewer to create several visual representations, such as co-authorship network maps, citation-based network maps, and co-occurrence network maps, to facilitate the visualization of qualitative data. To ensure the accuracy and validity of the findings, multiple validation steps were conducted. These included double-checking the input data, rerunning the software, and randomly selecting some outputs for review. However, several factors could potentially affect the results of the analysis. These include: (a) the possibility that the selected bibliographic repository and search terms may not capture all relevant articles related to the study's objectives, (b) the use of a fuzzy search, which could increase the likelihood of irrelevant publications being included in the database, and (c) the potential for unconscious bias from the author to influence the results.

Table 1.*Quantitative findings were obtained from the retrieved data.*

Data source	***Scopus repositories***
Covered Time Frame	1995-2024
Countries Included	83
Publication Count	1026
Number of organizations	1741
Number of Articles	496
Conference papers	497
Book chapters	16
Review	5

4. Results

4.1 Observations

The Influence Maximization literature database comprises 1,026 works published between January 2005 and November 2024, authored by 1,782 researchers from 1,741 organizations across 83 different countries (Table 1). The number of articles on influence maximization has shown a noticeable upward trend (Fig. 1(b)), with a significant exponential increase from 68 publications in 2015 to 152 in 2021. Notably, around 88.6 percent of all the publications in the database were released between 2015 and 2023.

The majority of selected publications were conference papers (49%) and journal articles (48.9%). Figure 1(c) illustrates that only 2.5% of the publications consist of grey literature, such as reviews and book chapters. The leading subject areas of the selected papers are computer science (86%), mathematics (30%), and engineering (25%). The top three most productive countries in the field are China (p = 484, accounting for 47.5%), the USA (p = 227, 22.6%), and India (p = 83, 8.5%). Wang Y, Chen W, and Zhang are the three most prominent authors in influence maximization research, contributing 3.5%, 3.02%, and 3.11%, respectively, to the selected literature. Figure 1(c) also shows the publication distribution, with 48.9% of papers published as articles, 49.0% as conference papers, 1.6% as book chapters, and 0.5% as reviews.

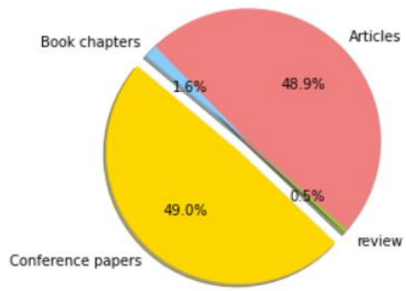
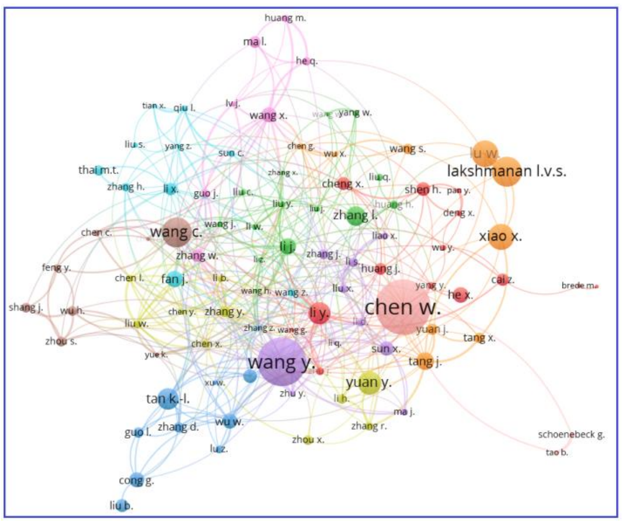


Fig 1(c)Document type of the Influence maximization literature.

4.2 Co-author analysis

This part creates a co-authorship and authorship map to illustrate the strength of the co-authorship links with other scholars.



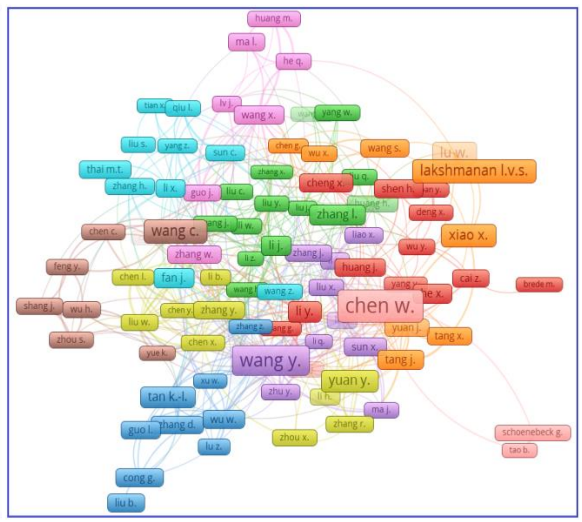
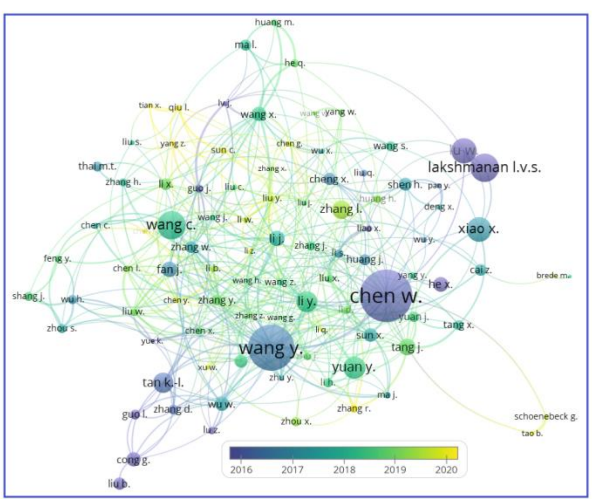


Fig. 2(a)Fig. 2(b)



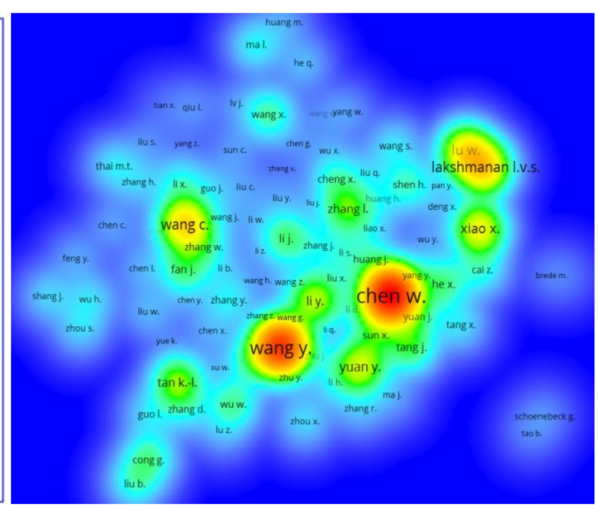


Fig. 2(c) Overlay visualizationFig. 2(d) Density Visualization

Fig 2(a),2(b),2(c), and 2(d) explains for each of the 151 authors,the total strength of the co-authorship links with other authors will be calculated. The author with the greatest total link strength will be selected. Out of all the authors, the highest link strength is 69 by wang published 36 papers and a total number of citations 4299.

4.3 Citation analysis by the organization:

Yangzhou University in China, Microsoft Research in China, and the University of Texas at Dallas are the three most frequently mentioned organizations in the field of influence maximization. According to the top ten most-cited entities, about half of these organizations are based in China and the USA (p = 4), with the remaining ones from Italy (p = 1) and Singapore (p = 1). Additionally, most of the organizations (p = 8) that meet the classification criteria are educational institutions, while only two, including Microsoft Research, are entrepreneurial research organizations. The findings highlight the significant contribution of the academic community, especially Chinese educational institutions, in advancing the study of influence maximization.

5. Conclusion

This study performs a scientometric analysis of 1,026 papers on influence maximization published over the past 20 years. The research not only provides a comprehensive synthesis of the current state of the field but also draws insights from the extensive body of literature to construct a model of influence maximization and its evolution. The analysis identified three main approaches currently used in influence maximization: (i) simulation-based, (ii) heuristic, and (iii) community-based methods. Since 2005, the field has undergone two major developments, driven by the integration of new technologies such as deep learning and metaheuristic methods. Overall, the findings reveal that research in influence maximization has steadily expanded over the last two decades.

Furthermore, the findings of this study reveal that most research on influence maximization in fields such as social sciences, astronomy, agriculture and biology, medicine, and transportation has been carried out in developed countries. There is a need for more studies to be conducted in developing nations, which face the most pressing challenges in these areas. Despite this, various factors could assist researchers in addressing the issue more effectively.

It would appear that these findings might provide researchers with a means to more readily specify the domains in which they wish to focus their subsequent study.

References

[1] D. J. Watts and S. H. Strogatz, "Strogatz - small world network Nature," *Nature*, vol. 393, no. June, pp. 440–442, 1998, [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/9623998

[2] Y. W. Teng, C. H. Tai, P. S. Yu, and M. S. Chen, "Revenue maximization on the multi-grade product," *SIAM Int. Conf. Data Mining, SDM 2018*, pp. 576–584, 2018, doi: 10.1137/1.9781611975321.65.

[3] J. N. Rosenquist and J. Murabito, "Article annals of internal medicine the spread of alcohol consumption behavior in a large," *Ann. Intern. Med.*, vol. 152, no. 7, p. 426, 2010, [Online]. Available: http://www.annals.org/content/152/7/426.short

[4] P. Domingos and M. Richardson, "Mining the network value of customers," *Proc. Seventh ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 57–66, 2001, doi: 10.1145/502512.502525.

[5] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2003, pp. 137–146. doi: 10.1145/956750.956769.

[6] J. Li, X. Wang, K. Deng, T. Sellis, J. X. Yu, and X. Yang, "Discovering influential community over large social networks," in *Proceedings of the 2017 IEEE International Conference on Data Engineering*, 2017, pp. 871–882.

[7] S. Galhotra, A. Arora, and S. Roy, "Holistic influence maximization: Combining scalability and efficiency with opinion-aware models," *Proc. ACM SIGMOD Int. Conf. Manag. Data*, vol. 26-June-20, pp. 743–758, 2016, doi: 10.1145/2882903.2882929.

[8] L. Qiu, X. Tian, J. Zhang, C. Gu, and S. Sai, "LIDDE: A differential evolution algorithm based on local-influence-descending search strategy for influence maximization in social networks," *J. Netw. Comput. Appl.*, vol. 178, no. July 2020, p. 102973, 2021, doi: 10.1016/j.jnca.2020.102973.

[9] M. Xie, X. X. Zhan, C. Liu, and Z. K. Zhang, "An efficient adaptive degree-based heuristic algorithm for influence maximization in hypergraphs," *Inf. Process. Manag.*, vol. 60, no. 2, 2023, doi: 10.1016/j.ipm.2022.103161.

[10] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "CELF++: Optimizing the greedy algorithm for influence maximization in social networks," *Proceedings of the 20th International Conference Companion on World Wide Web, WWW 2011*. pp. 47–48, 2011. doi: 10.1145/1963192.1963217.

[11] K. Jung, W. Heo, and W. Chen, “IRIE: Scalable and robust influence maximization in social networks,” *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 918–923, 2012, doi: 10.1109/ICDM.2012.79.

[12] S. Brin and L. Page, “The anatomy of a large-scale hypertextual Web search engine BT - Computer Networks and ISDN Systems,” *Comput. Networks ISDN Syst.*, vol. 30, no. 1–7, pp. 107–117, 1998, [Online]. Available: http://dx.doi.org/10.1016/S0169-7552(98)00110-X%5Cnhttp://apps.webofknowledge.com/full_record.do?product=UA&search_mode=GeneralSearch&qid=6&SID=X1pOWPMuSmOv1SlwJ6f&page=1&doc=2

[13] S. S. Singh, A. Kumar, K. Singh, and B. Biswas, “C2IM: Community based context-aware influence maximization in social networks,” *Phys. A Stat. Mech. its Appl.*, vol. 514, pp. 796–818, 2019, doi: 10.1016/j.physa.2018.09.142.

[14] A. M. Samir, S. Rady, and T. F. Gharib, “LKG: A fast scalable community-based approach for influence maximization problem in social networks,” *Phys. A Stat. Mech. its Appl.*, vol. 582, p. 126258, 2021, doi: 10.1016/j.physa.2021.126258.

[15] A. Bozorgi, H. Haghighi, M. Sadegh Zahedi, and M. Rezvani, “INCIM: A community-based algorithm for influence maximization problem under the linear threshold model,” *Inf. Process. Manag.*, vol. 52, no. 6, pp. 1188–1199, 2016, doi: 10.1016/j.ipm.2016.05.006.

[16] W. Li, Y. Li, W. Liu, and C. Wang, “An influence maximization method based on crowd emotion under an emotion-based attribute social network,” *Inf. Process. Manag.*, vol. 59, no. 2, p. 102818, 2022, doi: 10.1016/j.ipm.2021.102818.

[17] S. Kumar, L. Singhla, K. Jindal, K. Grover, and B. S. Panda, “IM-ELPR: Influence maximization in social networks using label propagation based community structure,” *Appl. Intell.*, vol. 51, no. 11, pp. 7647–7665, 2021, doi: 10.1007/s10489-021-02266-w.

[18] P. Bonacich, “Factoring and weighting approaches to status scores and clique identification,” *J. Math. Sociol.*, vol. 2, no. 1, pp. 113–120, 1972, doi: 10.1080/0022250X.1972.9989806.

[19] P. Bonacich and P. Lloyd, “Eigenvector-like measures of centrality for asymmetric relations,” *Soc. Networks*, vol. 23, no. 3, pp. 191–201, 2001, doi: 10.1016/S0378-8733(01)00038-7.

[20] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, “K-core organization of complex networks,” *Phys. Rev. Lett.*, vol. 96, no. 4, pp. 3–6, 2006, doi: 10.1103/PhysRevLett.96.040601.

[21] S. B. Seidman, “Network structure and minimum degree,” *Soc. Networks*, vol. 5, no. 3,

pp. 269–287, 1983, doi: 10.1016/0378-8733(83)90028-X.

[22] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, and N. Glance, "Cost-effective outbreak detection in networks," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 420–429, 2007, doi: 10.1145/1281192.1281239.

[23] S. R. Arora Akhil, Sainyam Galhotra, "Debunking the Myths of Influence Maximization: An In-Depth Benchmarking Study," *ACM*, 2017.

[24] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 199–207, 2009, doi: 10.1145/1557019.1557047.

[25] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "SIMPATH: An efficient algorithm for influence maximization under the Linear Threshold model," *Proc. - IEEE Int. Conf. Data Mining, ICDM*, pp. 211–220, 2011, doi: 10.1109/ICDM.2011.132.

[26] W. Chen, C. Wang, and Y. Wang, "Scalable Influence Maximization for Prevalent Viral," *16th ACM SIGKDD Int. Conf. Knowl. Discov. data Min.*, pp. 1029–1038, 2010.

[27] D. Williams, *Probability with Martingales*. Cambridge: Cambridge University Press, 1991. doi: DOI: 10.1017/CBO9780511813658.

[28] F. D. Malliaros and M. Vazirgiannis, "Clustering and community detection in directed networks: A survey," *Phys. Rep.*, vol. 533, no. 4, pp. 95–142, 2013, doi: 10.1016/j.physrep.2013.08.002.

[29] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010, doi: 10.1016/j.physrep.2009.11.002.

[30] M. E. J. Newman, "Detecting community structure in networks," *Eur. Phys. J. B*, vol. 38, no. 2, pp. 321–330, 2004, doi: 10.1140/epjb/e2004-00124-y.

[31] A. Biswas and B. Biswas, "FuzAg: Fuzzy Agglomerative Community Detection by Exploring the Notion of Self-Membership," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 5, pp. 2568–2577, 2018, doi: 10.1109/TFUZZ.2018.2795569.

[32] A. Kanavos, I. Perikos, I. Hatzilygeroudis, and A. Tsakalidis, "Emotional community detection in social networks," *Comput. Electr. Eng.*, vol. 65, pp. 449–460, 2018, doi: 10.1016/j.compeleceng.2017.09.011.

[33] K. X. Yu Wang, Gao Cong, Guojie Song, "Community-based Greedy Algorithm for Mining Top-K Influential Nodes in Mobile Social Networks Categories and Subject Descriptors," *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. data Min.*, pp. 1039–1048, 2010.

[34] V. R. K, M. Katukuri, and M. Jagarapu, “CIM: clique-based heuristic for finding influential nodes in multilayer networks,” *Appl. Intell.*, 2021, doi: 10.1007/s10489-021-02656-0.

[35] H. Li, S. S. Bhowmick, A. Sun, and J. Cui, “Conformity-aware influence maximization in online social networks,” *VLDB J.*, vol. 24, no. 1, pp. 117–141, 2015, doi: 10.1007/s00778-014-0366-x.

[36] E. Orduna-Malea, A. Martín-Martín, and E. D. López-Cózar, “Google Scholar as a source for scholarly evaluation: A bibliographic review of database errors,” *Rev. Esp. Doc. Cient.*, vol. 40, no. 4, pp. 1–33, 2017, doi: 10.3989/redc.2017.4.1500.

[37] A. Martín-Martín, E. Orduna-Malea, M. Thelwall, and E. Delgado López-Cózar, “Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories,” *J. Informetr.*, vol. 12, no. 4, pp. 1160–1177, 2018, doi: https://doi.org/10.1016/j.joi.2018.09.002.

[38] L. Yang *et al.*, “Satellite Altimetry: Achievements and Future Trends by a Scientometrics Analysis,” *Remote Sens.*, vol. 14, no. 14, pp. 1–22, 2022, doi: 10.3390/rs14143332.

[39] A. Belfiore, C. Cuccurullo, and M. Aria, “IoT in healthcare: A scientometric analysis,” *Technol. Forecast. Soc. Change*, vol. 184, no. August, p. 122001, 2022, doi: 10.1016/j.techfore.2022.122001.

[40] H. Jiang, M. Wang, and X. Shu, “Scientometric analysis of post-occupancy evaluation research: Development, frontiers and main themes,” *Energy Build.*, vol. 271, p. 112307, 2022, doi: https://doi.org/10.1016/j.enbuild.2022.112307.

[41] M. Daradkeh, L. Abualigah, S. Atalla, and W. Mansoor, “Scientometric Analysis and Classification of Research Using Convolutional Neural Networks: A Case Study in Data Science and Analytics,” *Electron.*, vol. 11, no. 13, 2022, doi: 10.3390/electronics11132066.

[42] B. Li *et al.*, “A scientometric analysis of agricultural pollution by using bibliometric software VoSViewer and Histcite™,” *Environ. Sci. Pollut. Res.*, vol. 29, no. 25, pp. 37882–37893, 2022, doi: 10.1007/s11356-022-18491-w.

[43] R. Osei-Kyei, T. Narbaev, and G. Ampratwum, “A Scientometric Analysis of Studies on Risk Management in Construction Projects,” *Buildings*, vol. 12, no. 9, 2022, doi: 10.3390/buildings12091342.

[44] T. Iancu *et al.*, “A Scientometric Analysis of Climate Change Adaptation Studies,” *Sustain.*, vol. 14, no. 19, 2022, doi: 10.3390/su141912945.