Societal Transformation: AI and Big Data Journal
Vol: 3(1): 50-64,2025
DOI:10.20547/aibd.253104

Real-time Trend Analytics in Periodic Literature

Alishba Masood[1]

Abstract: With the growing interest in academics and industry towards research and the availability of large number of publishers, the frequency of publications is also on the rise. It is found that researchers are very careful about the selection of the related periodicals and the proceedings with good research index and well-known publishers. In this direction, this research performed a quantitative analysis of the publication by various authors in leading journals. We identified their interests/ selection of the periodicals for their research publication. For this purpose, a scrapping tool was developed, and data was scrapped for scientific publications published in fifteen different prominent journals of IEEE, Elsevier, springer and ACM. The journals were related to computer architecture and performance optimization. The dataset contains 1323 research papers from 2400 different authors. Clustering has been performed to identify the authors and their preferred avenues for research. The analytics trend shows that about 98% of researchers try to remain associated with a single journal. This reveal limited cross-journal engagement even with same publisher. They are reluctant to publish in other journals and periodicals. This may imply the loyalty of authors to a specific outlet that may have implication for publication strategies, editorial policies, and research dissemination diversity. Even the authors try to avoid different journals of the relevant field of same publisher. For the analytics, R-studio was used where different cluster plots of journals against the authors are made and trends are visualized.

Keywords: *literature analytics, data science, periodic analytics, cluster analytics, dominant authors, research trends*

1. Introduction

As the research work in every field is increasing rapidly so this project focuses on the interest of IEEE/ACM/Springer research paper publishers in their field of interest so the publisher can conclude better results from this type of data analytics for their future work. It is also helpful for readers to see or conclude the influences of publishers towards specific fields. The idea is to analyze the data which will be

[1] NED University of Engineering and Technology, Karachi

gathered for specific countries authors, area of interest of publishers in IEEE explore/ACM/ Springer journals.

Consider that journal quality depends on being covered in the Science Citation Index (SCI). It is found that smaller journals might not be selected by authors if one relies primarily on publication count. For that reason, other metrics such as journal impact factor (JIF) have been developed.[12]

There are several papers which are published in the field of computer architecture and big data, but it is extensive for more fields as per user requirement. These analytics can be done by *periodical literature.* Periodic literature area also termed periodical publication or simply a periodical. They are a published work that is scheduled regular in new edition.[11]. Examples of periodical are magazines. They are scheduled for publication on weekly, monthly, or quarterly basis.

This research paper analyzes or performs trajectory analysis of research publications by authors in emerging disciplines. It is evident that as the volume of data generated grows exponentially, the demand for more advanced tools to process, analyze, and secure this information has led to significant strides in fields like data science, cybersecurity, and privacy protection. [16]

Below, we discuss emerging trends or the disciplines for which this paper analyzes the publication patterns. Emerging trends in data science research include data privacy and security. With the explosion of data, privacy concerns have become paramount. Techniques like differential privacy are being employed by tech companies like Apple and Google to collect and analyze data while ensuring that individual users' information remains private [17].

In the healthcare sector, AI models are being used to predict patient outcomes based on vast datasets. For example, AI-powered data analytics is also used in finance, where algorithms analyze trading patterns and market sentiment to predict stock prices.[18].

Despite primitive techniques for analyzing the trajectory of publication, more advanced analytics such as unsupervised clustering, descriptive analytics are required. K-means clustering attempts to divide *n* observations into *k* clusters. The property of each observation is that it is fit to the cluster with the nearest mean.

In this paper, for analysis of influences of publication in specific research organizations here the k-mean clustering is done for authors' journals columns. Rest of the sections of the paper is organized as follows. We will first provide related work. Then, the methodology is presented. This is followed by results and conclusion.

2. Related Work

2.1. Literature Review

There have been various studies done on identify patterns of publication in leading scientific journals. This section cites some of these studies. One of the studies presented in [1] performed analysis of text of the articles that are published in different journals. The articles were scrutinized on the basis of authors' nationality, their research genre and the title of the publication. It has been observed surprisingly that authors in native English countries, are the ones who contributed primarily for the publications. However, the authors from other non-English countries also have an important role on the published work.

In 2012, Association for Computing Machinery (ACM) devised a classification system for computing [13]. The authors in [14] employed this novel classification to investigate publication patterns in computing discipline for the years ranging from 1951 to 2017. The novelty lies in performing time series analysis. The authors identified how different patterns have emerged over different time. In another work [15], the authors have discovered the co-authorship network of one journal titled Journal of Universal Computer Science. This helps to discover interdisciplinary collaboration between ACM categories. For this purpose, these authors used the help of graph theory and employed a directed graph of computer science disciplines.

In the past few years, various studies have been published for multiple domains. However, interdisciplinary articles that integrated computer science techniques to resolve issues in cross disciplines such as economics, biology, and sociology are not only accepted in various computer science journals and conferences, but they are published in top-tier publication venues [19-20]. It is also observed that researchers from cross-disciplines collaborate on different interdisciplinary ideas to help themselves remain up-to-date with the latest trends in technology. Several research articles have analyzed networks of science to identify and correlate the relationship between different disciplines of science. The authors in [21] attempt to discover the relationship between biotechnology for energy and nanotechnology for energy fields. For this purpose, they employed citation networks to discover interesting patterns.

2.2 Research gap analysis

Having done the analysis of the literature, it has been found that these studies primarily employ citation network analysis, co-authorship analysis or broad-content analysis. What is missing is a fine-grained journal specific perspective via systematic parsing and clustering to identify researchers' publication trajectories over time. This may allow one to move from static analysis to longitudinal patterns of researcher's loyalty and mobility across journals. This research work is based on such analysis that differentiates it from the studies proposed in the past.

2.3. Contributions of the study

It is also noted that this research is not focused on any algorithmic innovation but application of established methods to a unique dataset curated specifically for this study. The selected domain of this study has not yet been explored for trajectory analysis. The employment of clustering and descriptive analytics for the said purpose adds a unique contribution in this domain i.e. application of AI and big data to scientific research publication.

3 Methodology for Trend Analytics

Figure 1 shows the methodology adopted for the research. The first step comprises scraping data from the various sources. Once the data is scrapped, it is analyzed using various approaches such as data analytics and clustering. The number of clusters were identified using Silhouette coefficient analysis and elbow method. The convergence is said to be achieved when subsequent changes in clusters were not observed (<0.0001) or after a maximum number of iterations (300 iterations). The Silhouette coefficient showed a reasonably good separation between clusters.

,

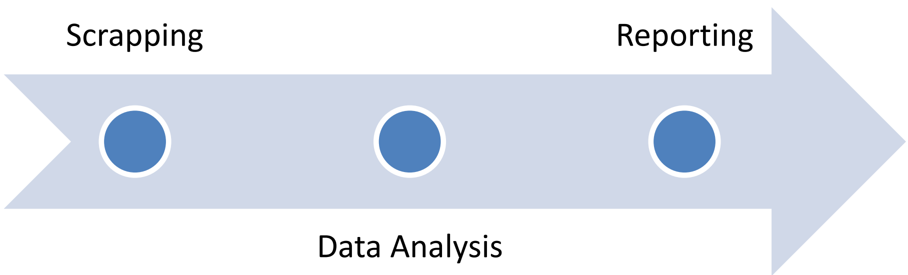

Figure 1: Methodology adopted for research

3.1 Data Collection

For trend analytics, data is collected first and the data collection is done by parser implemented in python and manually in the form of meta data. Figure 2 shows the various types of data scrapped, and the data sources i.e. the journal from which the data is scrapped. The journals are included from IEEE, SPRINGER, ELSVIER and ACM.

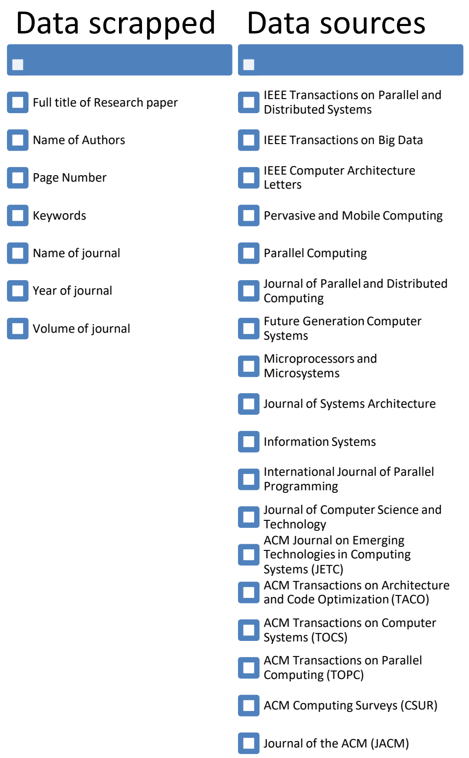

Figure 2: The data scrapped and the sources of the data

3.2 Parser for Data Collection

The collection of data is done by parser using python and done manually by research paper's Meta data. Firstly, data is collected in Meta data form. This data is about the research paper publishers of IEEE explorer/ACM/Springer journals, which are published in computer architecture and dig data field. Having input file formats HTM, TXT, BIB file of paper citation. The missing values were handled by imputation. In case when imputation is not possible, the missing values were removed. They are logged in and not considered for clustering. The parser outputs the data in a CSV file having columns as shown in Table 1. This includes the following information:

- Names of authors of each research paper,
- Title of research paper,
- Journal name,
- Volume and Number,
- Date of publication

Table 1: Summary of the data Collected by Parser

Author1	Author2	Author3	Author4	Title	Journal	Volume	Number	Page Number	Month
Diegues	Nuno and Romano	Paolo		Time-Warp: Effici	ACM Trans. Para	2	2	12:1--12:44	jun
Narendra	Siva G.			Challenges and D	J. Emerg. Techn	1	1	7--49	mar
Xiao Zhang	Yinrun Lyu	Yanjun Wu	Chen Zhao	MixHeter: A globa	Journal of Paralle	111		93-103	January
Guangwei Xu	Zhifeng Sun	Cairong Yan	Yanglan Gan	A rapid detection	Journal of Paralle	111		115-125	January
Chunyan Yu	Qi Lian	Dong Zhang	Chunming Wu	PAME: Evolution	Journal of Paralle	111		136-151	January
Libing Wu	Biwen Chen	Kim-Kwang Raymon	Debiao He	Efficient and sec	Journal of Paralle	111		152-161	January
Roger W Hockney	Ian J Curington			f12: A parameter	Parallel Computi	10	3	277-286	May
Miroslaw Malek	Eli Opper			The cylindrical ba	Parallel Computi	10	3	319-327	May
C Holt	A Stewart			A parallel thinning	Parallel Computi	10	3	329-334	May
James R.A Allwrig	D.B Carpenter			A distributed imp	Parallel Computi	10	3	335-338	May
Jong-Chuang Tsay	Young-Chang Hou			Generating functi	Parallel Computi	10	3	347-356	May
M.P Bekakos	D.J Evans			Relative performa	Parallel Computi	10	3	357-364	May
Alan George	Joseph W.H Liu	Esmond Ng		Communication r	Parallel Computi	10	3	287-298	May
Robert M Hyatt	Bruce W Suter	Harry L Nelson		A parallel alpha/b	Parallel Computi	10	3	299-308	May
Giuseppe A De Bia	Paolo Ciucci	Marco Cottone		Vectorized algori	Parallel Computi	10	3	339-346	May
Tao Li				Parallel impleme	Parallel Computi	10	3	309-318	May
James J Hack				On the promise o	Parallel Computi	10	3	261-275	May
D.J. Evans				Parallel S.O.R. ite	Parallel Computi	1	1	1--3	August
Wolfgang Gentzsch				Numerical algorit	Parallel Computi	1	1	19-33	August
M.J. Kascic Jr.				Vorton dynamics	Parallel Computi	1	1	35-44	August
Paul N. Swarztrauber				FFT algorithms fo	Parallel Computi	1	1	45-63	August
Wolfgang RÃ¶nsch				Stability aspects	Parallel Computi	1	1	75-98	August
Karl Solchenbach	Ulrich Trottenberg			SUPRENUM: Sy	Parallel Computi	7	3	265-281	Septem
Dennis Parkinson	Marvin Wunderlich			A compact algori	Parallel Computi	1	1	65-73	August
O Kolp	H Mierendorff			Performance esti	Parallel Computi	7	3	357-366	Septem
I Lehmann	F HÃ¶pfl			A model of distrib	Parallel Computi	7	3	395-401	Septem
W RÃ¶nsch	H Strauss			A linear algebra p	Parallel Computi	7	3	413-418	Septem
Ulrich Kremer	Heinz-J Bast	Michael Gerndt	Hans P Zima	Advanced tools a	Parallel Computi	7	3	387-393	Septem

Meta data of the scrapped information contained 1323 research paper entries. These are from the various journals of IEEE, Springer, Elsevier and ACM. The journal's year and volume information are shown in Table: 2.

Table 2: Meta data of the scrapped journals

S. No.	Journal	Volume	Year
1	J. Emerg. Technol. Comput. Syst.	1	2005
2	ACM Trans. Parallel Comput.	2	2015
3	Parallel Computing	1,	1984,
4	Pervasive and Mobile Computing	1,2,3,5	2005,2006,2003,2009
5	ACM Trans. Comput. Syst.	1,2,4,7,8,16,17,31	1983,1984,1986,1989,1990,1998, 1999, 2013
6	IEEE Computer Architecture Letters	1,2,3,4,5,6,7,8	2002,2003,2004,2005,2006,2007,2008,2009

7	IEEE Transactions on Big Data	1,2,3	2015,2016,2017
8	International Journal of Computer Information Sciences	1,2,9,20,21,38,40,41,42,43, 45	1972,1973,1980,1991,1992, 2010,2012,2013,2014,2015,2016,2017
9	ACM Trans. Archit. Code Optim.	2,3,6,7,8,9,10,17	2005,2006,2009,2010,,2011,2013,2014
10	ACM Comput. Surv.	24,25,26,27,30, 33,45	1992,1993,1994,1998,2001,2012
11	ACM Trans. Parallel Comput.	2,3	2015,2016
12	IEEE Transactions on Parallel and Distributed Systems	1,2,3,.4,5,6,7,8, 9,10,11,12, 13,	1990,1991,1992,1993,1994,1995, 1996,1997,1998,1999,200, 2001,2002
13	J. Emerg. Technol. Comput. Syst.	5,8	2009,20012
14	Parallel Computing	1,7,10	1984,1988,1989
15	Computer Architecture Letters	9	2010

3.3 Trends Analysis

Once the data is collected, analytics are done in following trends:

1. Identify the dominating authors which have most publication in the domain
2. Identifying authors which have publications in more than one co-related journal
3. Interests of authors in terms of research organization

These trends are analyzed in R-Studio by importing excel files of collected data. Through the studio table function, data frame is made from 1st, 2nd, 3rd author of their number of published papers in specific journal. Structured Query Language (SQL) quires are run against all authors' data frames for highest number of papers. The collected are then processed via a clustering algorithm. The clusters are made for each journal for finding authors' dominance.

4 Analytics in R Studio

Through R-Studio table function data frame is made of 1st, 2nd, 3$^{rd\ and}$ 4th author of their number of published papers in specific journal. SQL quires are run against all authors' data frames for highest number of papers. Clusters are made for each journal to find the authors dominancy and interest of research in specified research organization. From the SQL query, it is identified that there are clusters of research papers of authors in a specific journal.

Figure 3 provides a result for a data frame comprising authors who have paper publications in more than one journal. Only one author has cross-journal publication.

Most of the authors are accustomed to publishing in a single place. This might indicate a stronger degree of loyalty and can impact the diversity of research dissemination.

	Authors	ACM Comput. Surv.	ACM Trans. Archit. Code Optim.	ACM Trans. Comput. Syst.	ACM Trans. Parallel Comput.	Computer Architecture Letters	IEEE Computer Architecture Letters	IEEE Transactions on Big Data	IEEE Transactions on Parallel and Distributed Systems	International Journal of Computer Information Sciences
2622	J. Wang	0	0	0	0	0	0	4	0	0
2623	G. J. Qi	0	0	0	0	0	0	4	0	0
2624	N. Sebe	0	0	0	0	0	0	4	0	0
2618	Y. R. Lin	0	0	0	0	0	0	3	0	0
2619	H. Tong	0	0	0	0	0	0	3	0	0
2620	J. Tang	0	0	0	0	0	0	3	0	0
2621	C. C. Aggarwal	0	0	0	0	0	0	3	0	0
2611	F. Xia	0	0	0	0	0	0	2	0	0
2612	Q. Yang	0	0	0	0	0	0	2	2	0

Figure 3: Authors which have paper publication in more than one journal/ avenue

Interests of authors by their number of publications are extracted by SQL queries on data farm. The data frame of authors who have 10 or more publications are shown in Table 3. The table analyzes the interest of leading authors by their number of publications in various journals/ avenues. K.G. Shin has almost all publications in IEEE Transactions on Parallel and Distributed Systems. We performed further analysis and found out that no author with high output in terms of publications spans across multiple journals. This pattern reinforces the trend of specialization.

Table 3: Authors which have 10 and greater than ten research papers

Authors	ACM Trans. Comput. Syst.	ACM Trans. Archit. Code Optim.	ACM Trans. Comput. Syst.	ACM Comput. Surv.	ACM Trans. Parallel Comput	IEEE Transactions on Parallel and Distributed Systems	J. Emerg. Technol. Comput. Syst.	Parallel Computing	IEEE Transactions on Big Data
K. G. Shin	0	0	0	0	0	18	0	0	0
S. Olariu	0	0	0	0	0	11	0	0	0

For performing deep analytics, clustering is performed for dominating authors i.e. those who have one or more co-related journals of different research organization/ avenues. For cluster analysis, following major steps are performed: mean values according to journals, plotting clusters information, validation for clusters. k-means of authors for IEEE Transactions on Parallel and Distributed Systems (TPDS) and IEEE Transactions on Big Data (TPD) are performed. k-means clustering resulted in 2 clusters of sizes 1068 and 1556. The k-means value of authors are shown in Table 4. It can be deduced that cluster 1 comprises authors who frequently publish in TPDS and somewhat in TBD. Cluster 2 comprises authors with almost zero publication in

TPDS and minimal in TBD. These findings also substantiate that authors prefer one dominant journal and rarely publishes across both journals.

Table:4: K-Means of authors

IEEE Transactions on Parallel and Distributed Systems	1	1.192884
	2	***0.000000***
IEEE Transactions on Big Data	1	0.8670412
	2	***0.1529563***
Within cluster sum of squares by cluster	1861.3858 201.5964	

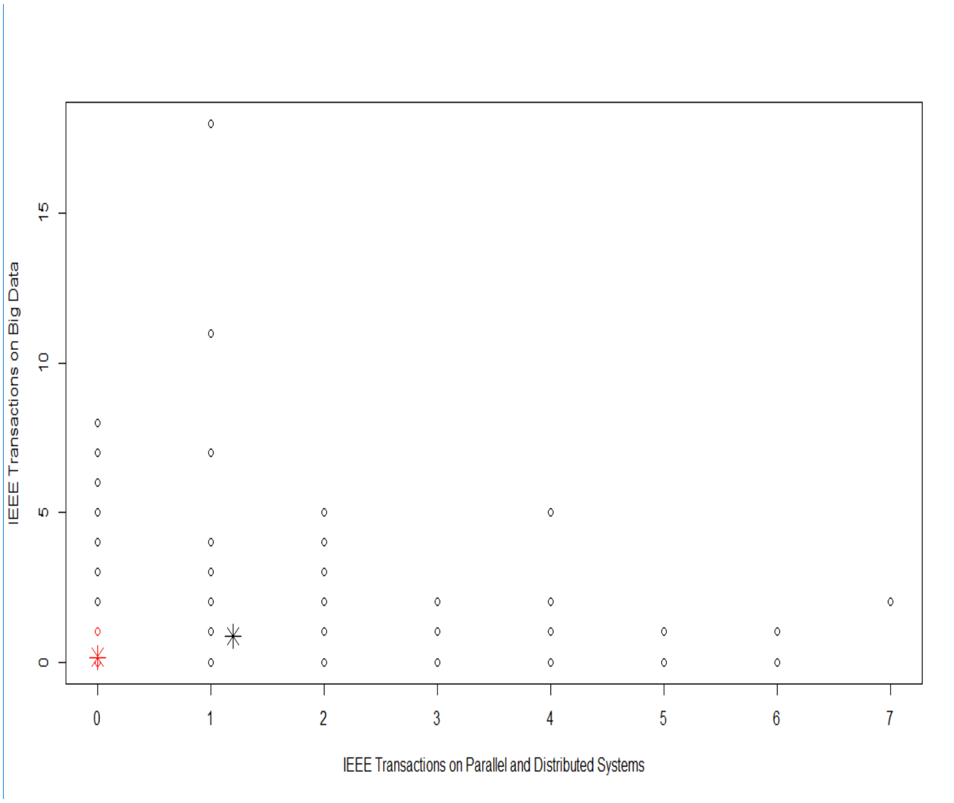

Figure 4: K-Means of authors

Eclipse shape clusters for IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Big Data and IEEE Computer Architecture Letters are shown in Figure 5. They are refined visualization of partitioning methods including k-means, Observations are represented by points in the plot, using principal components for an ellipse is drawn around each cluster. Cluster 1 in red shows moderate publishing activity that are spread across a limited number of journals. Cluster 2 in green shows authors who dominantly and strictly publish in one journal. Cluster 3 in blue shows contributors who change their avenues of publishing. This substantiated that researchers normally cluster around specific journals based on avenues.

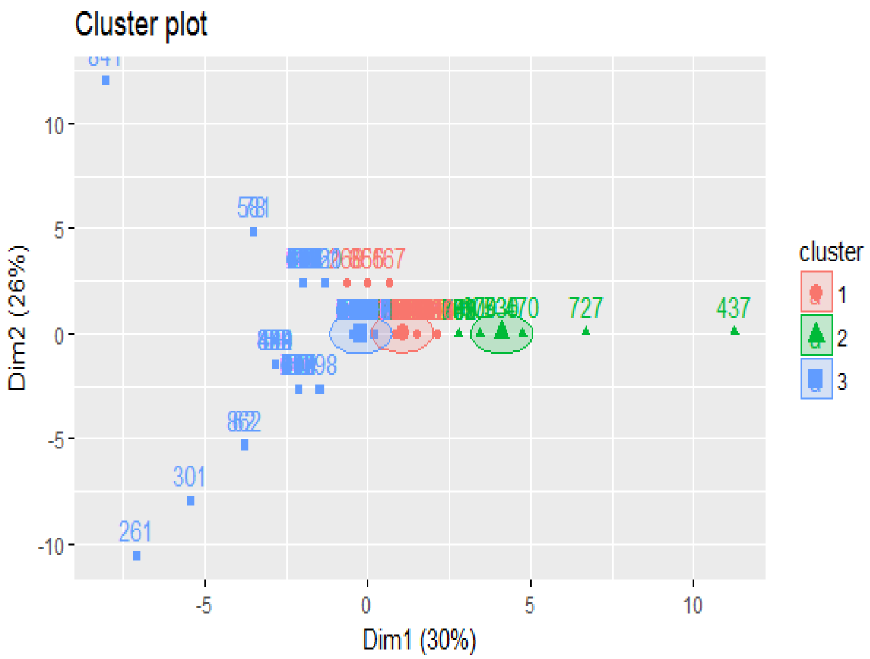

Figure 5: Eclipse shape clusters for IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Big Data and IEEE Computer Architecture Letters

Cluster plot for Journal of Computer Science and Technology, International Journal of Parallel Programming and Journal of Parallel and Distributed Computing is shown in Figure 6**.** The high dispersion of cluster 3 confirms that inter-disciplinary authorship is rare and scattered.

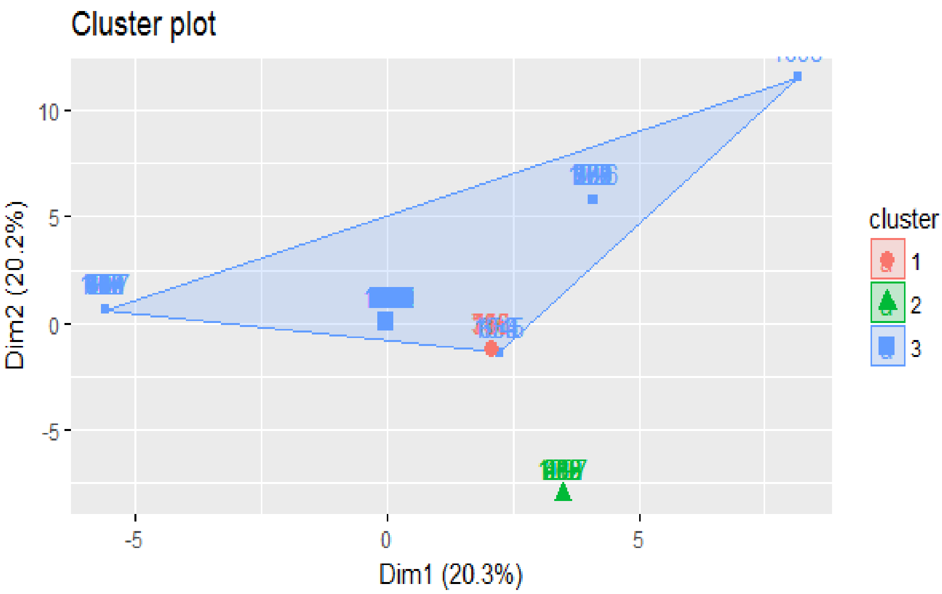

Figure 6: Cluster plot for Journal of Computer Science and Technology, International Journal of Parallel Programming and Journal of Parallel and Distributed Computing

Cluster plots for various journals are shown in Figure 7. Each color and symbol represent a distinct journal cluster. Cluster on the left (green, blue, cyan) represent IEEE and ACM journals. Clusters on the right represent Springer and Elsevier. Publisher influence is evident from the graph.

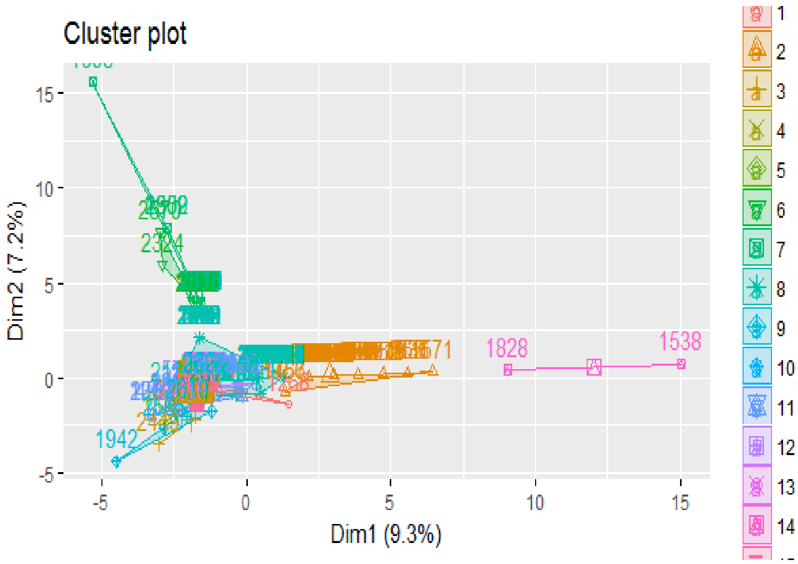

Figure 7: Cluster plot for all selected journal for trend analytics

Cluster plot for selected journals with validation is shown in Figure 8.

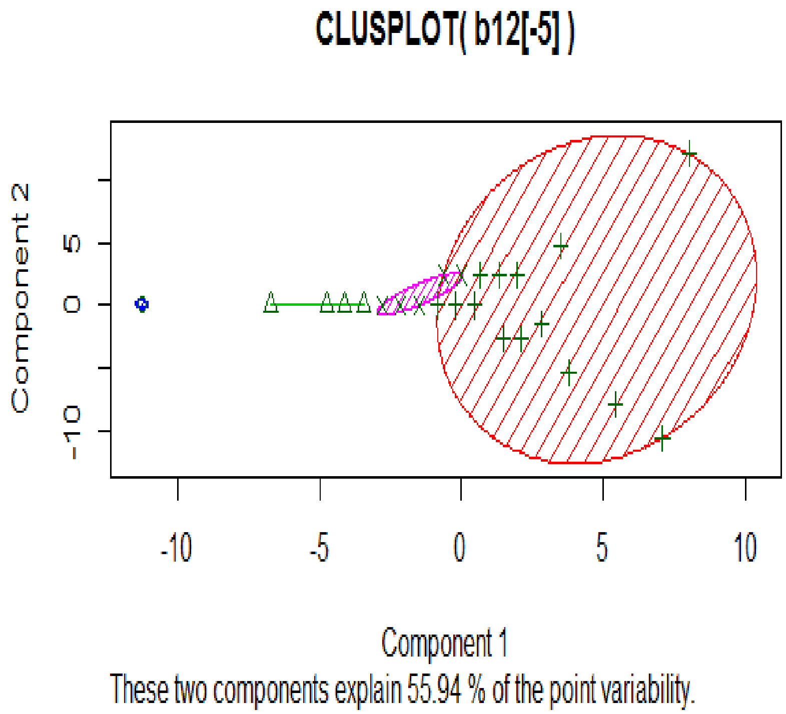

Figure 8: Cluster plot for three journals with validation

In the above Figure, there are three clusters of Journals, clusters are made according to the entries of authors' research papers. CLUSPLOT for all Journals in Ellipses are then drawn to indicate the clusters. The further layout of the plot is determined by the optional arguments. Cluster plot for all fifteen journals with validation is shown in Figure 9.

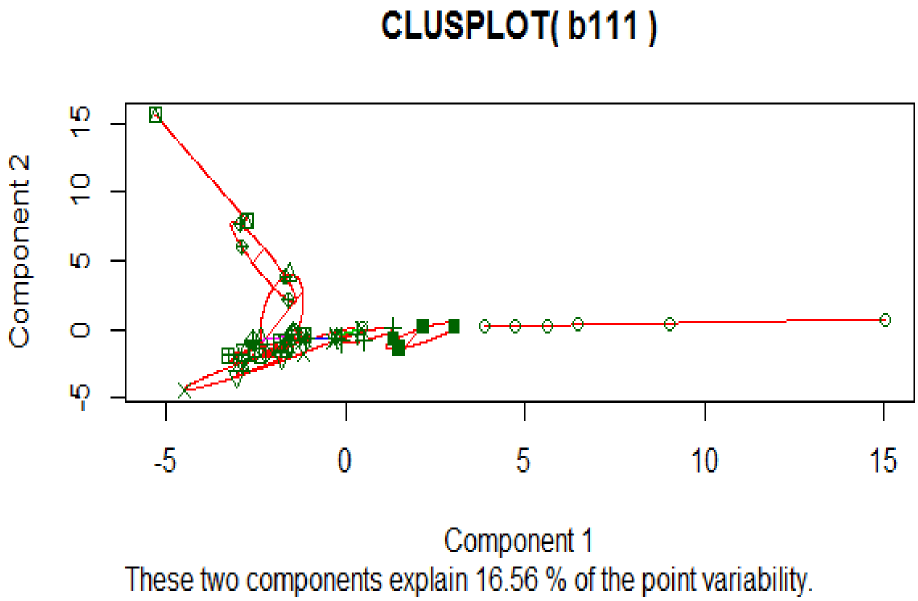

Figure 9: Cluster plot for all fifteen journals with validation

The primary objective of the study is the descriptive analysis of research journal publication patterns. Descriptive analytics are considered as the valid first step identifying structural features of scholarly communication. The future work can focus on hypothesis testing, statistical testing and significance modelling. It is found that 98% of researchers remain associated with a single journal that can be validated further in future studies using advanced data analytics.

5 Conclusion

In this research paper, periodic trend analytics were performed to identify the preferences of the authors in publishing their journals at various avenues. Results and trends were found through various data science techniques, SQL querying and clustering. These analyses were performed over the meta data collected about research papers in leading computer science journals of IEEE, Elsevier, and Springer and ACM. These trends were analyzed on collected data, which have 1300 research papers of fifteen leading journals and 2400 authors. It is concluded that most of the authors are very much interested/ accustomed to the publishing with their preferred publishers. Most of the authors prefer for publishing in that a specific journal of a publishing body or publishing in other related journals but with same publisher. Also, authors don't prefer working with different publishers, they primarily rely on one publisher.

There are certain practical implications associated with editors. For instance, editors can leverage the loyalty of authors via having programs such authors engagement programs, targeted call for papers. In addition, there are implications for authors as well. They may have limited visibility and networking by only focusing on specific journals.

6 Discussion and Future Direction

In the future, studies can be conducted to analyze interdisciplinary phenomena. For instance, one can analyze physical sciences, numerical sciences, and social sciences by employing a dataset covering a wide area of scientific disciplines.

In addition, the employed clustering method reveals loyalty patterns. It will be interesting to see if these patterns remain same after editorial board changes or journal policy reforms. Temporal analysis could be done to identify if the same behavior persists by authors at various stages of their career. Such researches can help in understanding deeply publication dynamics and clarify about editorial strategies and career development choices.

References

[1] Tsai, C. (2005). Research and trends in science education from 1998 to 2002: A content analysis of publications in selected journals. *International Journal of Science Education*, 27(1), 3–14. https://doi.org/10.1080/0950069042000243727

[2] IEEE Transactions on Parallel and Distributed Systems. (n.d.). Retrieved January 25, 2017, from http://ieeexplore.ieee.org/xpl/tocresult.jsp?isnumber=2627&punumber=71

[3] *Journal of Supercomputing*, 45(5). (2017, October). Retrieved March 12, 2017, from https://link.springer.com/journal/10766/45/5/page/1

[4] ACM Journal on Emerging Technologies in Computing Systems (JETC). (n.d.). Retrieved April 14, 2017, from http://jetc.acm.org/archive-toc.cfm

[5] Big data: Issues, challenges, tools and good practices. (n.d.). Retrieved January 25, 2017, from http://ieeexplore.ieee.org/abstract/document/6612229/

[6] Outlier. (n.d.). *Wikipedia*. Retrieved February 14, 2017, from https://en.wikipedia.org/wiki/Outlier

[7] Density-based clustering exercises. (2017, July 2). *R-bloggers*. Retrieved from https://www.r-bloggers.com/density-based-clustering-exercises/

[8] Merging data. (n.d.). *Quick-R*. Retrieved July 10, 2017, from https://www.statmethods.net/management/merging.html

[9] *Parallel Computing*. (n.d.). Elsevier. Retrieved March 11, 2017, from https://www.journals.elsevier.com/parallel-computing/

[10] ACM Transactions on Computer Systems (TOCS). (n.d.). Retrieved April 15, 2017, from http://tocs.acm.org/archive-toc.cfm

[11] Periodical literature. (n.d.). *Miami Dade College Libraries*. Retrieved January 3, 2017, from https://libraryguides.mdc.edu/researchperiodicalsguideHime

[12] Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA*, 295(1), 90–93. https://doi.org/10.1001/jama.295.1.90

[13] Rous, B. (2012). Major update to ACM's computing classification system. *Communications of the ACM*, 55(11), 12. https://doi.org/10.1145/2366316.2366321

[14] Cassel, L., & Buzydlowski, J. (2020). Exploring research trends in computing using the ACM computing classification system. In *Proceedings of the Northeastern Association of Business, Economics and Technology Conference (NABET)* (pp. 56–63).

[15] Ullah, M., Shahid, A., Roman, M., Assam, M., Fayaz, M., Ghadi, Y., Aljuaid, H., et al. (2022). Analyzing interdisciplinary research using co-authorship networks. *Complexity, 2022*, Article 6598743. https://doi.org/10.1155/2022/6598743

[16] Chen, X., et al. (2020). Big data: A survey. *Computer Science Review*, 37, 100312. https://doi.org/10.1016/j.cosrev.2020.100312

[17] Dwork, C. (2008). Differential privacy: A survey of results. In *Theory and Applications of Models of Computation* (pp. 1–19). Springer. https://doi.org/10.1007/978-3-540-79228-4_1

[18] Liu, X., et al. (2018). Financial market prediction using machine learning algorithms. *Financial Technologies Journal*, 2(1), 65–79.

[19] Tadmor, B., & Tidor, B. (2005). Interdisciplinary research and education at the biology–engineering–computer science interface: A perspective (Reprinted article). *Drug Discovery Today*, 10(23–24), 1706–1712. https://doi.org/10.1016/S1359-6446(05)03646-0

[20] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., et al. (2009). Computational social science. *Science*, 323(5915), 721–723. https://doi.org/10.1126/science.1167742

[21] Karunan, K., Lathabai, H. H., & Prabhakaran, T. (2017). Discovering interdisciplinary interactions between two research fields using citation networks. *Scientometrics*, 113, 335–367. https://doi.org/10.1007/s11192-017-2480-7