3-26-2019

Application of author bibliographic coupling analysis and author keywords ranking in identifying research fronts of Indian Neurosciences research

Sevukan Rathinam
sevukan2002@yahoo.com

Sri Amudha Sankar
sriamudhas@gmail.com

Application of author bibliographic coupling analysis and author keywords ranking in identifying research fronts of Indian Neurosciences research

Sri Amudha S[1]
Research Scholar

R Sevukan[2]
Associate Professor & Head
[1-2]Department of Library and Information Science
Pondicherry University
Puducherry – 605 014
Email: sriamudhas@gmail.com

Abstract

Probing research fronts identification unfailingly delivers interesting results in any field due to its decisive nature. Citation analysis is an acclaimed method used in this process among which more successful results backing Author Co-citation Analysis (ACA) and Author Bibliographic Coupling Analysis (ABCA). The current study opted to combine author bibliographic coupling network analysis and author keywords to explore and display a graphical representation of prominent research areas' evolution over the study period in Indian Neuroscience research domain. Application of hierarchical clustering to author bibliographic coupling networks for all non-overlapping consecutive years included in the study period were performed and analysed in VOSviewer mapping software. The powerful Lin/log modularity normalization was chosen for determining distance based similarity while clustering the network units. Results of the study unfolded ten prominent research subfields with more emphasis on Epilepsy' and 'Parkinson's disease' research. Depression was identified as one of the upcoming prominent area in recent years. Apart from its cruciality in framing national level mental health policies, the study will also prove ABCA to be an effective method in identifying prominent research areas.

Keywords

Author bibliographic coupling, Author keywords, Research fronts, Indian Neuroscience, Science mapping, VOSviewer

Background

Neuroscience is a medical domain concerning neurological disorders and diseases. A pentad of years ago in 2013, neurological disorders were declared 'global epidemic', a global health issue in priority by World Health Organization (WHO). Neurological disorders vastly include epilepsy, headache, Parkinson's, Alzheimer's and other dementias, brain injuries etc. among which epilepsy, headache and stroke are claimed to be prevalent in India. Although the disproportionate neurologists and patients ratio is a common global issue, Indian scenario of the same highlighted a bleak ratio of 800 neurologists for more than 1 billion people. Achieving a modest ratio is a big challenge as our country trains only 80-90 neurologists annually. Moreover, these trained neurologists are heavily concentrated in metro cities leaving the large needy patients deprived in rural areas of India. Adding to this grim setting is a recent finding of a three dimensional bibliometric study that 'Neurosciences' is one among the seven areas where no institutions published more than 500 publications each during the period 2008-2012. Clearly, Indian educational research was claimed to be biased towards engineering and physical sciences with little allocated to medical field. Upgrading the country's science policy can aid to take leap towards the betterment of the alarming situation and for delivering superior health care services. Identification of research fronts regained popularity recently due to its significance in science policy development. Citation based approach is a globally approved technique used to obtain overview of a domain's research frontiers perceiving research topic similarity between the cited and citing documents. Especially, bibliographic coupling being an evidence based measure of intellectual relatedness classified under citation based approach; the study adopts author-aggregated bibliographic coupling method to trace the knowledge network of Indian neuroscience.

Literature review

Citations are the reflections of different connectivity among subfields and citation pattern provide details on the evolution of the intellectual structures of a domain. Co-citation analysis (CA) was predominantly used to map knowledge networks of any chosen research field. Later, Vladutz and Cook (1984) examined and suggested that BC can produce accurate results in terms of subject relatedness. It was Zhao and Strotmann (2008) who initially highlighted the limitations of ACA in identifying emerging subfields by explicating that it reflects the older publications of established authors. They claimed that ACA network reflects the older works of influential authors. Van Raan studied the age dependency effects of the references used in building bibliographically coupled (BC) publication network structure and found that more recent the common references better the accuracy in their intellectual relatedness. Following this, an another study proposed a age-sensitive bibliographic coupling and claimed that not only the number of common references but also the age of the references affects the intellectual relatedness between the documents. Kuusi and Meyer (2007) reported the suitability of bibliographic coupling in anticipating technological breakthroughs. Boyack (2009) successfully mapped the knowledge structure of Chemistry by applying bibliographic coupling (BC). Boyack and Klavans (2010) investigated the accuracy of different citation based approaches and concluded that BC is comparatively better. Consecutively, many evidences (Jarneving 2007; Morris, Yen, Wu, & Tesfaye, 2003) were found claiming Bibliographic coupling as an appropriate method for research fronts identification in mapping science.

Author bibliographic coupling was proposed as an extended version of bibliographic coupling especially aiming to overcome the limitations of author co-citation technique. The study claimed that author bibliographic coupling (ABC) network reflects the work of active authors in the field and thus appropriate to identify emerging research fronts. Ji, Jeong and Youl (2013) successfully examined the intellectual structure of Korean Library and Information Science using author bibliographic coupling analysis. Recently, author keywords are taking center stage in presenting the research fronts. Li and Chu (2016) examined a novel technique of enhanced co-word analysis and succeeded in exploring research fronts. Further many works (Chen, Chen, Wu, Xie, & Li, 2016; Hu, Zhang, 2015; Khasseh, Soheili, moghaddam & Chelak, 2017) were published proving the advantage of using keywords/keyword networks in identifying research fronts. A notable study by Wanying, Jin, & Kun (2018) explored the relationships among different ranking metrics, including one frequency-based and six network-based metrics, in order to understand the impact of network structural features on ranking themes on co-word networks. They claimed that coreness is ideal for categorizing keywords. Therefore, Author bibliographic coupling analysis was considered to be combined with author keywords ranking in identifying the knowledge structure and the research fronts of Indian Neuroscience domain.

Data collection and Research Methodology

The study used scientific literature published by Indian authors on Neuroscience discipline for 27 years during the period 1989-2015. Full bibliographic data along with citation data for a total of 7966, 10801 records were retrieved from Web of Science and Scopus respectively in the below formats:

Web of Knowledge -Win tabbed (.txt file)

Scopus - Plain text (.ris file)

For the convenient use of VOSviewer, Scopus data were converted to Web of Knowledge tabbed format using. Uniformity in the data from both databases is thus achieved after which all the units of the bibliographic dataset were processed and rearranged in the desired order using Bibexcel software for the researcher to manually eliminated redundancy in dataset. Finally, the data is ready for the scientometric analysis.

Step 1: Scopus data files are converted from .ris file format to Web of science plain text in .txt file format using conversion software 'Scope2wos.exe' and further converted to Web of Science tabbed file format using 'Bibexcel' software. **Step 2:** Bibliographic data from two databases are combined and compared manually to eliminate redundancy. **Step 3:** Bibliographic data is submitted as input to VOSviewer software for generating bibliographic coupling network on author aggregator level. **Step 4:** The largest connecting component of the network was identified and used for further analysis so as to eliminate publications with negligible or no citation relationships. **Step 5:** Various clusters were made on the extracted largest connecting network using Linlog modularity parameter that uses (bottom-up) agglomerative hierarchical clustering. Grouping of data is done on the basis of nearest distance measure by applying ward's methods where the sum of squared Euclidean distance is minimized. **Step 6**: Each clusters were separately analyzed. Naming of each cluster to identify emerging research fronts was done based on the list of authors in each cluster and their corresponding author keywords. **Step 7:** step 5 was repeated separately for every year and the grouping of authors were done manually to generate the evolution map of total clusters in the field. Prominent clusters were extracted from the total number of clusters based on the duration period each cluster exists. Naming of each cluster was done based on the author keyword frequency method. **Step 8:** Eventually, the emerging research fronts were identified by comparing both the software generated clusters and manually generated clusters. **Step 9:** The network files (.net) generated from VOSviewer are submitted to the network analysis software Pajek for macro analysis - Component size, Average distance between nodes, degree and cluster coefficient of the generated networks. **Step 10:** The network files (.net) are submitted as input to NetDraw network analysis software for micro analysis – performance of the individual nodes level units involved in the networks.

Analysis

Bibliographic data of the data retrieved for the study period was submitted to the VOSviewer as an input to generate a network of authors based on bibliographic coupling relationship. Most productive authors were assumed to be found in the main component of the network and had less distance to other authors in comparison to the less creative authors. For this reason, the author bibliographic coupling network was limited to high productive authors by applying a minimum threshold of 15 publications for each author. A count of 531 authors was found to fulfill the minimum threshold and the generated network was further limited to largest connecting component of the network. It was found that the largest connecting component of network, which is the main component of the network, comprised of 500 nodes and it was considered for further analysis. The network was mapped using the VOSviewer mapping algorithm and clustered based on their subject relatedness. Distance between the nodes in the network is directly proportional to their subject relatedness. Thickness of the lines between the network nodes represents the bibliographic coupling strength between them. After the application of mapping and clustering algorithm, visualization of the generated author bibliographic coupling network is displayed in Fig. 1.

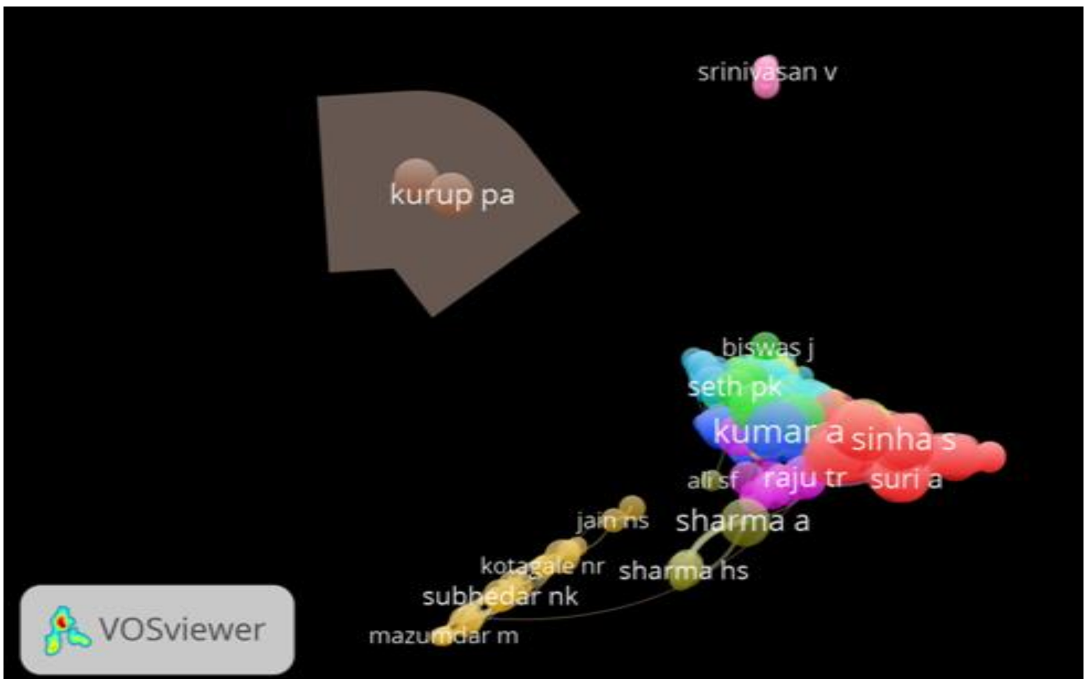

Fig. 1 Visualization of Author bibliographic coupling network

The average degree of centrality of this network was computed as 260.45. Density of the network was figured out to be 0.522 which shows 52% cohesiveness in the network between the authors. Interestingly, the value of cluster coefficient for the network was computed as 0.5515 i.e. 55.14%. Cluster coefficient illustrates the chance of nodes in the network to establish new connections in the future. Therefore the estimated clustering coefficient value 0.757 illustrated higher possibility of authors extending their subject scope and establish more new connections. Average distance among the reachable pairs was 1.4782 which connoted the information flow from one vector to any other vector in the network can be averagely reached in less than 2 nodes. However, the most distant vertices were Ahmed S (13) and Lafuente JV (229) with the distance value 3. This network presented 91 k-cores ranging from 17 to 259 depicting good collaboration which correlates with the collaboration indices computed. Micro level metrics Degree, Closeness, Betweenness and Eigenvector computed for the network showed that a list of 10 Authors namely Kumar A, Kumar S, Singh S, Sharma S, Singh A Kumar P, Shankar SK, Das S, Gosh S, and Mohankumar KP played the highly influential role in this network.

Authors/nodes in the network were further clustered by applying lin/log modularity normalization in the VOSviewer mapping software. Application of clustering algorithm resulted in 12 different clusters. For the purpose of naming these 12 prominent clusters, authors listed in each clusters were associated with their corresponding author keywords. Based on the frequency of the author keywords, each cluster was assigned names. Authors with high bibliographic coupling score in each cluster were identified as the prominent authors of the clusters. Thus identified cluster names and their corresponding prominent authors are presented in Table 1. It was found that the authors research work concerning 'Migraine' and 'Delirium' had high and strong bibliographic coupling strength representing high level of subject relatedness.

Table 1 Names of 12 Research Clusters Generated Through VOSviewer

Cluster No.	Representing Author	Research Fronts
1	Paulose CS	Parkinson's
2	Sinha S	Epilepsy
3	Kumar A	Oxidative stress
4	Raju TR	Tremor
5	Gangadhar BN	Schizophrenia
6	Islam F	Parkinson
7	Subhedar NK	Seizure
8	Mohanakumar KP	Parkinson
9	Sharma A	Delirium
10	Srinivasan V	Depression
11	Kamal MA	Alzheimer's/ dementia
12	Kurup PA	Migraine

Manual Clustering of Indian Neuroscience by Tracing the Evolution of Clusters

Author bibliographic coupling (ABC) network was generated for all the years of the study period 1989 to 2015, in the sequence of 1989, 1989-1990, 1989-1999 and likewise. 27 networks were generated with authors who qualify the software generated minimum threshold for each year. Authors/nodes in the network were further clustered by applying lin/log modularity normalization in the VOSviewer mapping software. Due to recurrent emergence of new authors in the field, tracing the continuity of author occurrence in the clusters, based on cluster colours generated by VOSviewer happened to be an intricate process. Henceforth, the names of authors identified in the network listed in each clusters of every individual years were manually compared with the authors' list in the clusters of the consecutive year to plot the continuity of author's occurrence in a particular cluster. By repeating the above process, an evolution map of various research clusters in Indian neuroscience was developed. In this fashion, there were 39 different clusters found in Indian neuroscience field among which 10 were found to be prominent, based on the regularity of its occurrence. The evolution map thus generated is displayed in Fig. 2. Visualizing the evolution of various clusters in the neuroscience field is displayed in Fig. 3(a-c). Taking into consideration that emergence of new clusters is the portrayal of specific research emerging as prominent subfield in Neuroscience domain, it is evident that 1992, 1994, 1996 were the years found to be significant with emergence of 3, 5 and 7 new clusters each respectively. For the purpose of naming the identified 10 prominent clusters, the core author nodes identified by NetDraw software listed in each clusters were associated with their corresponding author keywords. Based on the frequency of the author keywords, each cluster was assigned appropriate cluster name. Authors with high bibliographic coupling score in each cluster were identified as the prominent authors of the clusters.

To verify and finalize the research fronts, these clusters in Table 1 were compared with the manually identified 10 prominent clusters from the evolution map. Comparison between these two lists is presented in Table 2. Comparison of these two research fronts showed clusters in both the Tables matched except that the manually identified cluster C2 representing Parkinson's disease occurred repeatedly in C1, C6, C8 software identified prominent clusters. As a result, 'Parkinson's disease', 'Schizophrenia', 'Tremor', 'Seizure', 'Epilepsy', 'Dementia', 'Migraine', 'Oxidative Stress', 'Depression' and 'Delirium' were identified to be the prominent research fields in Indian neuroscience field. Research domain density and relatedness among other research fronts could be visualized in Fig. 4. Visual representation of research fronts denoted Parkinson's disease research as core research field in Indian neurosciences field. Obviously, all dementia related disorders namely 'Alzheimer's disease', 'Parkinson's disease', 'Schizophrenia', and 'Delirium' were closely located displaying higher level of subject relatedness. 'Migraine' and 'Depression' were also identified as prominent research fields located in distant from the core research field.

Table 2 Comparison of Clusters Generated and Filtered

Manual Cluster No.	Cluster Names
c2	Parkinson's
c6	Schizophrenia
c8	Tremor
c17	Seizure
c22	Epilepsy
c26	Dementia/ Alzheimer's
c30	Migraine
c36	Oxidative stress
c38	Depression
c39	Delirium

Software Cluster No.	Cluster Names
1	Parkinson's
2	Epilepsy
3	Oxidative stress
4	Tremor
5	Schizophrenia
6	Parkinson's
7	Seizure
8	Parkinson's
9	Delirium
10	Depression
11	Dementia/ Alzheimer's
12	Migraine

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39
1989	1	2																																					
1990	1	2																																					
1991	1	2																																					
1992	2	5	1	3	4																																		
1993	2	1	3	4		5																																	
1994	5	7	6	9		2	1	3		8	10	11																											
1995	6	3	11	8		2	5	7	1	4	13		12	10	9																								
1996	4	1	14	9		2		5	2	7	15		11		10	8	12	13	16	17	6	3																	
1997	9	1		14		3		8	7	10	13		5			4	11	12			6	2																	
1998	5	1				3		8		9	12		7			2	10	6				4	11																
1999	6	1				2		8		9	14		7			3	11	5				4	13	12	10														
2000	8	1		13		3		6		12	15		7			2	10	4				5	11		9	14													
2001	9	1		12		3		6		7	14		4			2	10	5				13	11			8	15												
2002	10	1		14		3		6		8	16		9--7			2	12	4				15	13			5	17	11	18	19									
2003		1				3		6		9			5			2	11	4				10	8			7	13		12	14									
2004		1				3		4		10			6			2	9	5								7			11	13	12	8							
2005		1				3		5		10--8			6			2	7	4								11--9			12	13	14								
2006		2				4		3		9--8--11						1	10	5								12--6			13	15			7	14					
2007		1				5		3								11--2	8					9				6			10	12			7						
2008		1				4		3								2	9					8				6			10	11					5	7			
2009		1				4		3								2	8					7				6			9	10					5				
2010		1				4		3								2	8					5				12,6			10	13					9	7	11	11	
2011		10 2				3		6								7	9					1				8			12	13					5	4		11	
2012		9 7--1				4		3									6					2				8			12	13						5		11	
2013		8 5--2				9		3									6					1				7			12	13						4		11	10
2014		7--5--1				6		4									8					2--10				9				13						3		11	12
2015		6--8--1				5		4									7					2				11				12						3		10	9

Fig. 2 Prominent research clusters in Indian Neuroscience research

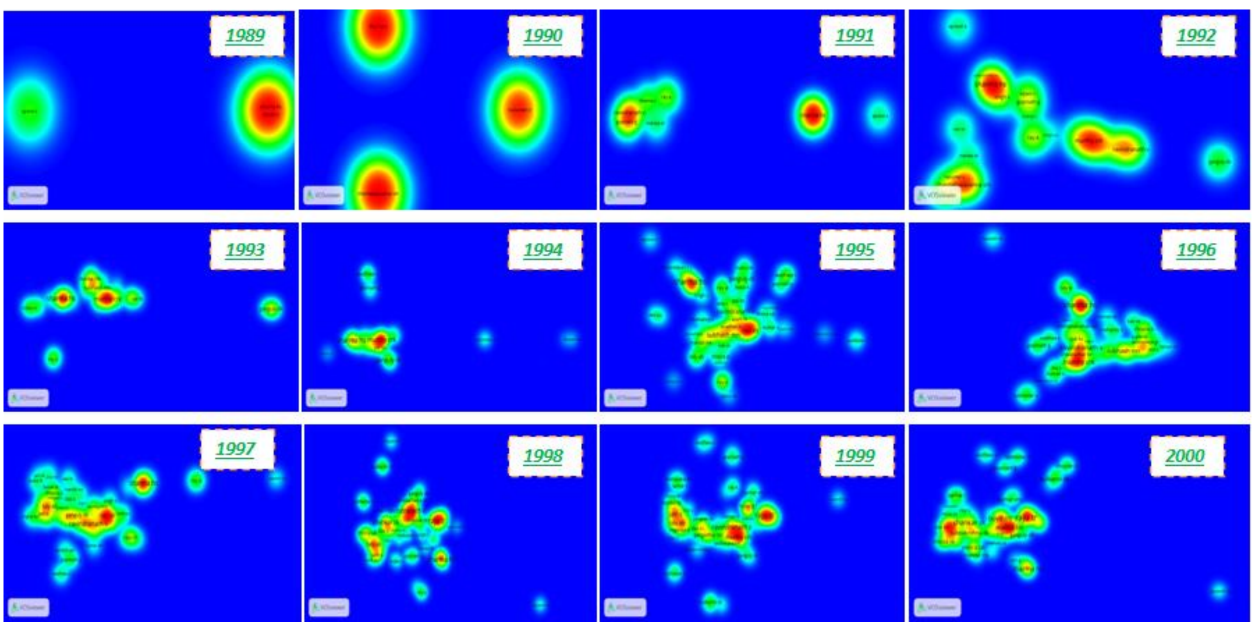

Fig. 3(a) Visualizing the evolution of various clusters in the neuroscience field, 1989-2000

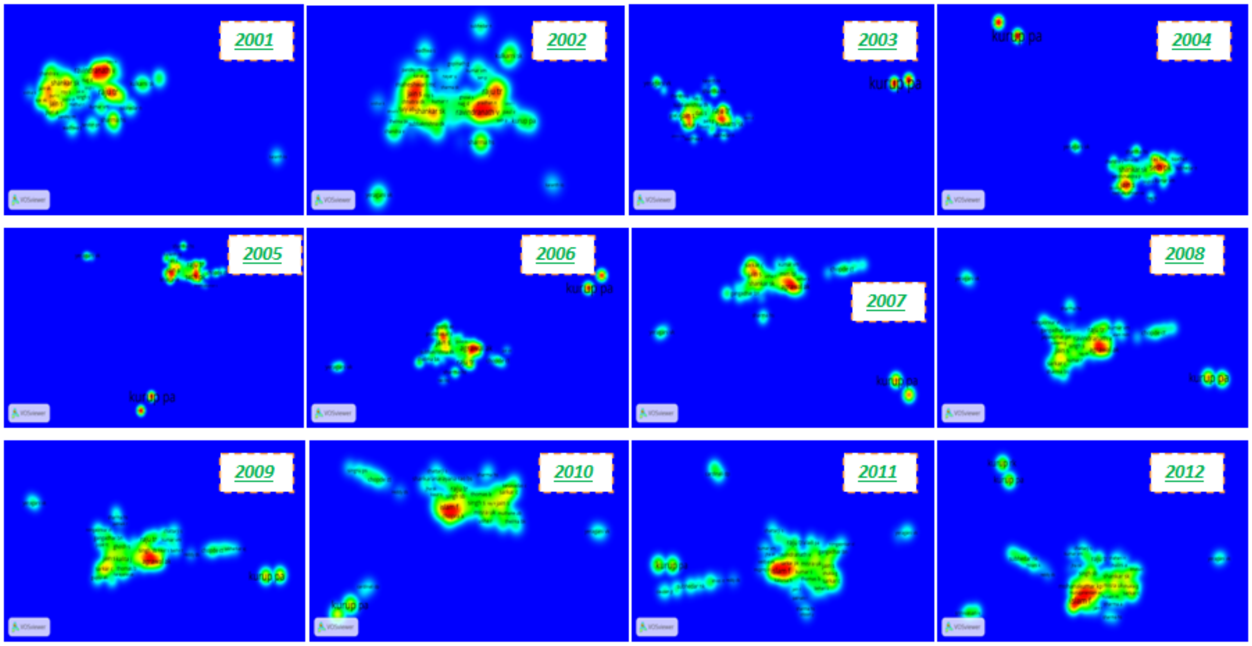

Fig. 3(b) Visualizing the evolution of various clusters in the neuroscience field, 2001-2012

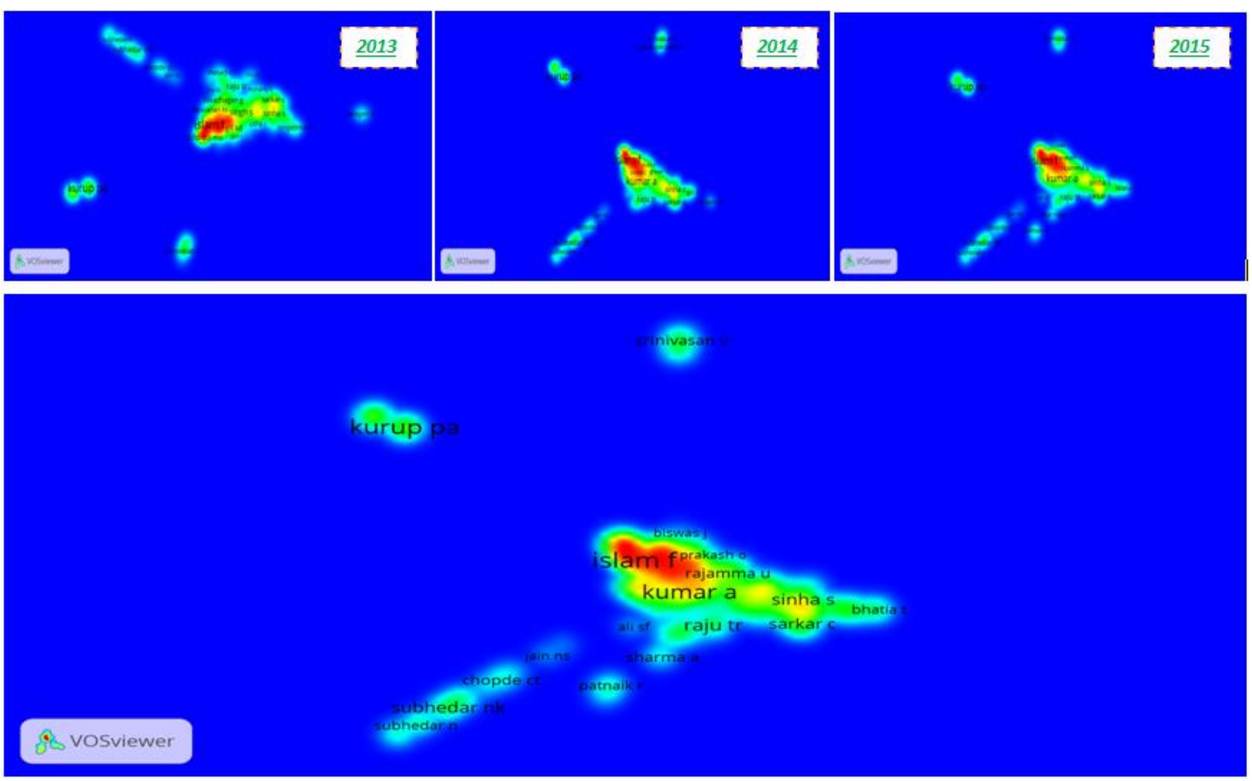

Fig. 3(c) Visualizing the evolution of various clusters in the neuroscience field, 2013-2015

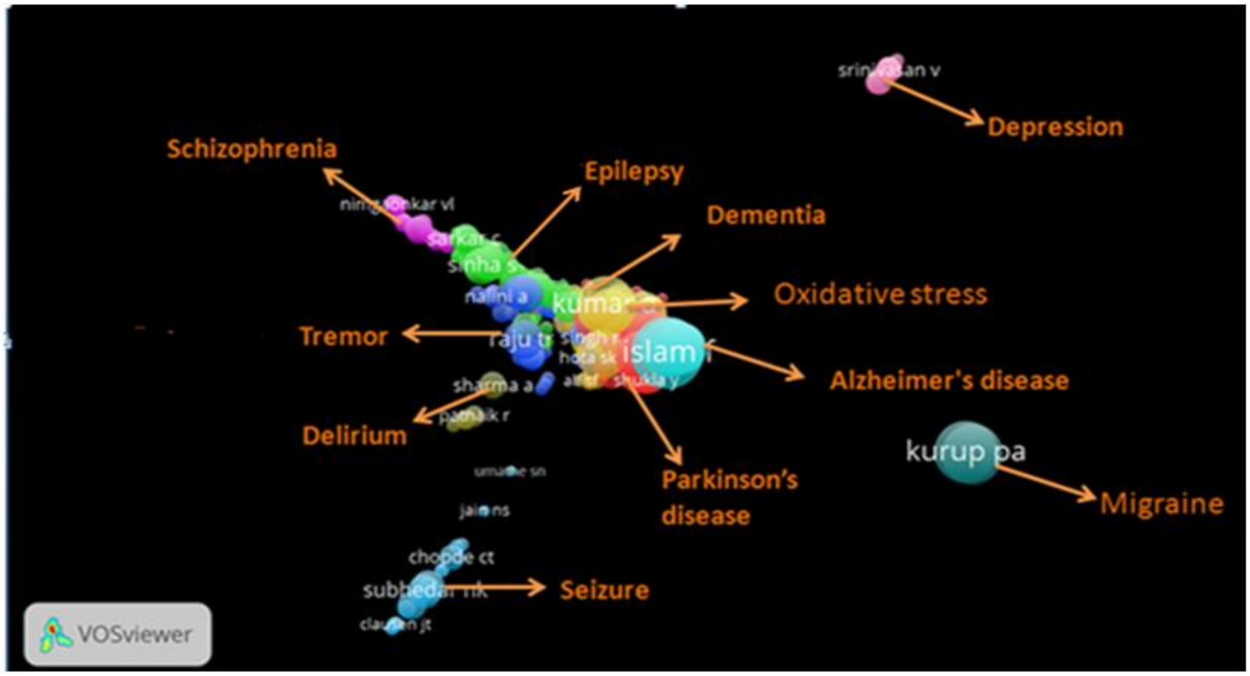

Fig.4 Visualization of Neuroscience research fronts in VOSviewer mapping software

Discussion

Findings of the current study witnessed that the Indian neuroscience research fronts ‘Epilepsy’, ‘Parkinson’s’, ‘Migraine’, ‘Dementia’ and ‘Tremor’ correspond with neuroscience disorders with high incidence rate in India. Especially it was interesting to find that ‘Epilepsy’ research had gained prominence early in the year 1996 and continued to have high growth rate until now. Growth rate analysis performed lead to the finding that ‘Epilepsy’ had the highest growth rate percentage of 71% followed by ‘Parkinson’s disease’ research with growth rate percentage of 66%. Surprisingly, growth rate of all the remaining prominent subfields had less than half of the growth rate of the top two fields. ‘Delirium’ was found to be the subfield with least growth rate percentage of 19%. Though ‘Depression’ was not identified as a prevalent disorder in the epidemiological study mentioned above, the current study identifies depression as a major research. ‘Depression’ had emerged as a prominent field only in the recent years. Moreover its growth rate was ranked at 3rd position by overshadowing the growth rate of other fields that gained prominence decades before it. Though it was identified to be one among the high incidence disorders in India since 1994, its growth rate of ‘Tremor’ was found to be quite low. This could be justified by the fact that it neither affects the patients’ everyday life nor make them dependable on other people unlike other discussed disorders. Therefore it was also quite clear that there exists no significant relationship between the year of prominence and the growth rate which means neither high growth rate nor high publications/authors can be associated with the advantage of early emergence.

More subfields were found to have emerged predominantly during the years 1996, 1994, 1992 respectively. This evinced that the decade of 1990s was named as the 'decade of the brain' as it was accountable for enormous growth in neuroscience domain (Tandon, 2000). However, the identified research frontiers were found to have gained prominence equally during and after the 1990s decade. In terms of growth rate, the top two fronts Epilepsy and Parkinson's belonged to the years 1996 and 1989s whereas the following three fronts namely 'Depression', 'Migraine' and 'Dementia' were found to be in the years 2010, 2002 and 2000. Investigation of the second group showed that the initial three fronts namely 'Schizophrenia', 'Seizure' and 'Tremor' gained prominence during 1993, 1996, 1994 whereas Oxidative stress and Delirium were in 2008 and 2013. Obviously, it is understood that the recent subfields are taking over the center stage

Conclusion

A total of ten prominent research subfields in Indian neuroscience research field and a graphical representation of their evolution over the study period were identified. It was found that the research on 'Epilepsy' and 'Parkinson's disease' were the highly productive with high growth rate. Research in the field of 'Depression', though became prominent in recent years it was found to have high growth rate. Findings also hinted a substantial increase in the two other research areas namely Schizophrenia and Tremor. Dementia and Tremor were found to correspond to neuroscience disorders with high incidence rate in India. The study's findings not limited to the above are worth considering while framing national level mental health policies to enhance neuroscience research quality and neurological health care system in India.

References:

Boyack, K. W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics*, 79(1), 27-44.

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389-2404.

Chen, X., Chen, HJ., Wu, D., Xie, Y. Li, J., (2016). Mapping the research trends by co-word analysis based on keywords from funded project. *Proceida Computer Science.91*, 547-555. Retrieved on 14.5.2017 from https://www.sciencedirect.com/science/article/pii/S1877050916313333.

Hu, J., Zhang, Y., (2015). Research patterns and trends of recommendation system in China using co-word analysis. *Information Processing & Management. 51*(4). 329-339. Retrieved on 5.7.2018 from https://www.sciencedirect.com/science/article/abs/pii/S0306457315000266.

Jarneving, B. (2007). Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics*, *1*(4), 287–307. Retrieved on 20.05.2017 from https://ideas.repec.org/a/eee/infome/v1y2007i4p287-307.html.

Jun Yoo, Y., Yun Lee, J. & Choi, S. (2012) Intellectual structure of Korean theology 2000–2008: Presbyterian theological journals. *Journal of Information Science, 39*(3), 307-318. Retrieved on 25.05.2017 from http://journals.sagepub.com/doi/abs/10.1177/0165551512466972.

Khasseh, A.A., Soheili, F., Moghaddam, H.S., Chelak, A.M., (2017). Intellectual structure of knowledge in iMetrics: a co-word analysis. *Information Processing & Management.53(*3). 705-720. Retrieved on 25.05.2018 from https://www.sciencedirect.com/science/article/abs/pii/S0306457316303338.

Kuusi, O. & Meyer, M. (2007). Anticipating technological breakthroughs: Using bibliographic coupling to explore the nanotubes paradigm. *Scientometrics, 70*(3), 759-777.

Li, M., Chu, Y. (2016). Explore the research front of a specific research theme based on a novel technique of enhancwed co-word analysis. *Journal of Information Science*.*43*(6). 725-741. Retrieved on 5.7.2018 from https://journals.sagepub.com/doi/10.1177/0165551516661914.

Morris, S. A., Yen, G., Wu, Z., & Tesfaye, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology*, *54*(5), 413-422.

Raan, Van. (2003). Reference-Based Publication Networks with Episodic Memories. Scientometrics. 63. 10.1007/s11192-005-0227-x.

Vladutz, G., & Cook, J. (1984). Bibliographic coupling and subject relatedness. *Proceedings of the American Society for Information Science*, 21, 204-207.

Wanying, Z., Jin, M., & Kun, L. (2018). Ranking themes on co-word networks: Exploring the relationships among different metrics. *Information Processing & Management, 54(*2). 203-218. Retrieved on 5.07.2018 from https://www.sciencedirect.com/science/article/abs/pii/S0306457317304430.

Zhao, D., & Strotmann, A. (2008). Author bibliographic coupling: Another approach to citation-based author knowledge network analysis. *Proceedings of American society for information science and technology, 45* (1), 1-8. Retrieved on 20.05.2017 from https://onlinelibrary.wiley.com/doi/10.1002/meet.2008.1450450292.