Analysis of the Citation Patterns, Keywords Clustering and Collaboration Networks in Presented Papers of PLANNER Conventions

Pallab Pradhan ***Hitesh Solanki*** ***Swapnil Patel***

Abstract

In this paper, we present analysis of the papers that were presented in the last nine PLANNER conventions held during the period 2003 to 2014 in North East Indian Universities. The data for this study i.e. number of papers, their bibliographical data along with the keywords were collected from "IR@INFLIBNET"repository as well as from the "Convention's Proceedings" and the citation data was collected from "Google Scholar". A total of 445 full papers were selected for this study excluding the abstract papers in which 172 papers have received a total of 481 citations with an average of 1.08 per paper and the remaining 273 papers had no citation. Further, this paper demonstrates year-wise citation patterns, most prominent topics presented, key-word clustering, authorship and col-laboration networks of authors from different states from the presented papers in conventions.

Keywords: PLANNER Conventions, Citation, Keywords Clustering, Co-Authorship Networks, Collaboration Networks

1. Introduction

Promotion of Library Automation and Networking in North-Eastern Region (PLANNER) is a biannual regional convention with a special focus to uplift academic libraries in the North-Eastern region, India in the aspect of library automation and networking. It was launched in 2003 at North-Eastern Hill University (NEHU), Shillong. The event is organized by Information and Library Network (INFLIBNET) Centre, Gandhinagar in collaboration with the universities located in the respective North-Eastern States (http://www.inflibnet.ac.in/planner2016/). It is one of the most popular and happening conventions in the LIS field held in North-Eastern region of India. A total of 9 PLANNER conventions have been organized till date (August, 2016) during the period 2003 to 2014.

The collection of presented papers in PLANNER conventions provides a wealth of information enabling us to calculate and have an idea about the research trends, impact of the conventions with the help of citations, networks of LIS authors / researchers that contribute papers, prominent and emerging LIS topics, etc. specially in North-East India.

The present study is an attempt to analyse and explore PLANNER-wise / year-wise citation patterns i.e. cited and un-cited papers, real average citation, highly cited papers; which are the mostly presented topics in the conventions with the help of keyword analysis and clustering; and lastly, the authorship and collaboration networks of all the contributed authors.

© INFLIBNET Centre, Gandhinagar, Gujarat

2. Literature Review

The authors found and reviewed a lot many literatures available on the web which were based on bibliometric analysis of conferences / conventions / journals papers mostly on LIS field. For instance, Janakiramaiah and Doraswamy (2011) examined all the convention papers published in CALIBER during 2005 - 2006. Tsay (2011) analysed the publications of Journal of Information Science (JIS) over the period 1998 - 2008 and explored the journal bibliometric characteristics. Pareek (2013) studied the bibliometric aspects of the literature of IFLA Journal during the period 2001-2010. Gopalakrishnan, Gopalkrishnan, Bathrinarayanan and Tamizhchelvan studied the cited and un-citedness, relative growth rate of un-cited publications retrieved from SCOPUS during the period 1970 - 2013. Dash, Sahoo and Mohanty (2015) analysed the various bibliometric dimensions of the publications reflected in the Library Assessment Conference (LAC) proceedings from 2014 to 2016. Sordo, Oghihara and Wuchty (2015) analysed the evolution of research groups and topics by utilizing all the manuscripts in the proceedings of the ISMIR conference from 2000 - 2014.

However, the authors could not find any study done on citation analysis, keywords clustering and authorship networks on the publications of any LIS conventions / conferences / seminars or journals. Thus, this study was carried out to analyse the papers of PLANNER conventions.

3. Objectives

Some of the objectives of this study are:

- Identify the cited / un-cited papers presented in the PLANNER conventions,
- Explore the trends and ratio of cited vs. un-cited vs. total presented papers,
- Study the PLANNER wise citation and authorship patterns of papers,
- Find out the top cited papers and pattern of citations distribution,
- Explore and visualise the networks of co-authors,
- Discover the most prominent research topics presented, and
- Find out the author collaboration networks from different states.

4. Methodology

4.1 Sources of Data

The "IR@INFLIBNET" (http://ir.inflibnet.ac.in) which is INFLIBNET Centre's institutional repository and the PLANNER convention's proceedings available in INFLIBNET Centre's library were used to collect the year wise data on papers presented in all nine PLANNER conventions. "Google Scholar" was used to collect the citations data of those papers.

4.2 Data Collection

A total of 445 papers were collected for nine PLANNER conventions during 2003 - 2014 by excluding the abstract papers, and were selected for this study. All the bibliographical details i.e. titles, author's name, author's / institute's belonging state, and keywords associated with each paper was collected. Subsequently, the authors thoroughly checked, corrected the spelling errors of all titles and author's name, and tallied them with both the sources available. The number of citations of each collected papers were retrieved from "Google

Scholar" on Wednesday / 24th August, 2016 which were amounted to a total of 481 citations.

4.3 Data Analysis

After all the corrections made, the data was saved in MS Excel file and all the statistical calculations and tabular interpretations were performed on that. Bibexcel, the bibliometric analysis software was used to perform all the required bibliometric operations. Further, visualization of the results of keywords clustering, and collaboration networks was performed in Pajek visualization software.

5. Findings of the Study and Discussions

5.1. Papers, Authors and Citations

As mentioned before, a total of 445 papers were presented and considered for this study which have received a total of 481 citations. The below Table-1 shows the year wise host universities name, PLANNER wise articles presented in conventions, their number of authors along with their year wise citations data during the period 2003 - 2014. The Table-1 reveals that a total of 828 authors had written and presented those 445 articles in conventions without calculating author's frequencies. The total authorship average was 1.86 authors per paper while the received citations average is 1.08 per paper. While considering the year wise authorship pattern,PLANNER 2008 hosted at Nagaland University had the maximum collaborative papers with 2.39 authors per paper, followed by PLANNER 2004 at Manipur University with 2 authors per paper. The papers presented in PLANNER 2007 hosted at Gauhati University and PLANNER 2008 hosted at Nagaland University have the maximum number of citations received with 1.45 citations per paper. The papers presented in first PLANNER 2003 hosted at North-Eastern Hill University have the 2nd lowest rank in citations received with 0.07 citations per paper while the papers presented at the very last PLANNER 2014 at Dibrugarh University have received very low citations with 0.03 citations per paper.

Table 1: Papers, Authors and Citations Average (Year wise and Total)

Host University Name	Year	Papers (P)	% of Total Papers	Authors (A)	% of Total Authors	Avg. Authors per Paper (Year)	Avg. Authors per Paper (Total)	Citations (C)	% of Total Citations	Avg. Citations per Paper (Year)	Avg. Citations per Paper (Total)
North-Eastern Hill University	2003	45	10.11	70	8.45	1.56	0.16	29	6.03	0.64	0.07
Manipur University	2004	34	7.64	68	8.21	2.00	0.15	47	9.77	1.38	0.11
Assam University	2005	60	13.48	113	13.65	1.88	0.25	43	8.94	0.72	0.10
Mizoram University	2006	66	14.83	113	13.65	1.71	0.25	88	18.30	1.33	0.20
Gauhati University	2007	47	10.56	75	9.06	1.60	0.17	68	14.14	1.45	0.15
Nagaland University	2008	44	9.89	105	12.68	2.39	0.24	64	13.31	1.45	0.14
Tezpur University	2010	50	11.24	96	11.59	1.92	0.22	60	12.47	1.20	0.13
Sikkim University	2012	54	12.13	107	12.92	1.98	0.24	69	14.35	1.28	0.16
Dibrugarh University	2014	45	10.11	81	9.78	1.80	0.18	13	2.70	0.29	0.03
Total		445	100.00	828	100.00		1.86	481	100.00		1.08

5.2 Citations Distribution

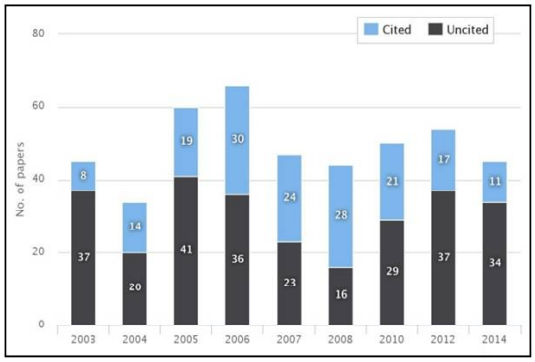
The below Table-2 presents the distribution of number of citationsreceived for each article. It indicates that 66 (14.83%) papers have received only one citation each and 4 (0.90%) papers have received more than 10 citations each which are the highly cited papers presented in PLANNER conventions. 102 (22.92%) papers have received citations between 2 to 9 times. If, we consider the actual 172 papers out of 445 papers which have only received citations then the average number of citations per paper is 0.36 against 1.08 citation per paper of total average.

Table-2: Citations Distribution

Citations (Unit)	Papers (P)	%
0	273	61.35
1	66	14.83
2	38	8.54
3	25	5.62
4	17	3.82
5	10	2.25
6	1	0.22
7	6	1.35
8	4	0.90
9	1	0.22
10+	4	0.90
Total	445	100.00

5.3 Distribution of Cited & Un-cited Papers

The 445 papers used in the study have received a total of 481 citations including the self-citations. The Chart-1 and 2 shows the trend and graphical representations of year wise distribution of cited and un-cited papers from the total papers presented in conventions during 2003 - 2014. Out of 445 papers, 172 (38.65%) papers are cited and 273 (61.35%) papers have received no citations till date. It shows the citation trend of papers.



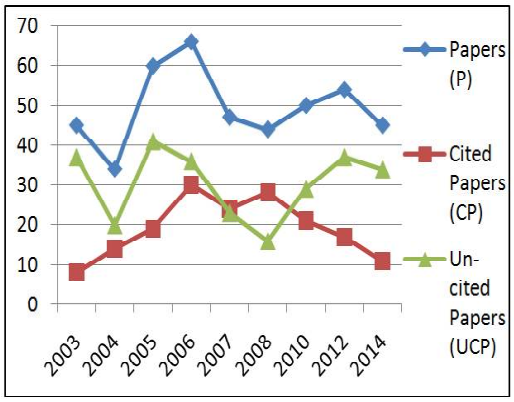


Chart 1 and 2: PLANNER'sYear Wise Distribution of Cited & Un-cited Papers

5.4 Calculation of Cumulative, Real Average Citation and Ratio

For further clarification on citations and papers, the year wise cumulative statistics of cited and un-cited papers has been calculated and presented in below Table-3. The Table-2 shows the calculated cumulative percentage of cited and un-cited papers, ratio in between cited vs. un-cited vs. total papers. The cited papers ranges between 1.80 and 6.74 from

total papers, whereas the un-cited papers ranges between 3.60 and 9.21. The study indicates that the PLANNER / year wise no. of un-cited papers are on higher side with an average of 61.35 as compared to the cited papers with an average of 38.65 during the period. Further, real average citations (RAC) varies year wise from the maximum of 4.06 for 2012 to 1.18 for 2014 which represents the average citations rates of paper having at least one citation over a period of each past 10 years. The average RAC during the study is 2.80 which indicates that RAC for the year 2005, 2008 and 2014 are on below average.

Table 3: Year wise Cumulative, RAC & Ratio in between Cited vs. Un-cited vs. Total Papers

Year	2003	2004	2005	2006	2007	2008	2010	2012	2014	Total
Cited Papers (CP)	8	14	19	30	24	28	21	17	11	**172**
(Ratio Cited R=CP/P) (Year wise)	17.78	41.18	31.67	45.45	51.06	63.64	42.00	31.48	24.44	**38.65**
% of Total Cited Papers	4.65	8.14	11.05	17.44	13.95	16.28	12.21	9.88	6.40	**100.00**
% of Total Papers (Total)	1.80	3.15	4.27	6.74	5.39	6.29	4.72	3.82	2.47	
Cummulative of Cited	8.00	22.00	41.00	71.00	95.00	123.00	144.00	161.00	172.00	
Cum. %	4.65	12.79	23.84	41.28	55.23	71.51	83.72	93.60	100.00	
Ratio CP/UCP	0.22	0.70	0.46	0.83	1.04	1.75	0.72	0.46	0.32	**0.63**
Real Average Citation (RAC)	3.63	3.36	2.26	2.93	2.83	2.29	2.86	4.06	1.18	**2.80**
Un-cited Papers (UCP)	37	20	41	36	23	16	29	37	34	**273**
(Ratio Un-cited R=UCP/P) (Year wise)	82.22	58.82	68.33	54.55	48.94	36.36	58.00	68.52	75.56	**61.35**
% of Total Un-cited Papers	13.55	7.33	15.02	13.19	8.42	5.86	10.62	13.55	12.45	**100.00**
% of Total Papers (Total)	8.31	4.49	9.21	8.09	5.17	3.60	6.52	8.31	7.64	
Cummulative of Un-cited	37.00	57.00	98.00	134.00	157.00	173.00	202.00	239.00	273.00	
Cum. %	13.55	20.88	35.90	49.08	57.51	63.37	73.99	87.55	100.00	
Ratio UCP/CP	4.63	1.43	2.16	1.20	0.96	0.57	1.38	2.18	3.09	**1.59**

5.5 Highly Cited Papers

Citations are considered as a positive indicator for a published work / paper of an author after his / her nos. of publications. The below Table-4 shows the list of most cited papers, i.e. top 5 ranks by citation counts during the study period.

Table-4: List of Highly Cited Papers

Title	Authors	PLANNER Year	Citations	Rank
Activities and Reasons for Using Social Networking Sites by Research Scholars in NEHU: A Study on Facebook and ResearchGate	Chakraborty, Nirmali	2012	20	1
Concept of Quality in Library Services: An Overview	Thakuria, Pranjit Kumar	2007	13	2
UGC- Infonet: Its Availability and Use in Universities of Assam	Borthakur, Jyotika; Das, Rumi; Gohain, Anjan	2010	12	3
Cloud Computing for Libraries: A SWOT Analysis	Pandya, Miteshkumar	2012	12	3
Studies on the scenario of Internet use pattern of Assam University Community and Local population of barak valley : A survey	Sinha, Manoj Kumar	2004	9	4
Planning, Problems and Solutions for Automation and Networking of University Libraries in North Eastern Region: A Case Study of Assam University Library	Sinha, Manoj Kumar; Bhattacharjee, Jayant	2003	8	5
Electronic Theses and Dissertations for Indian Universities : A Framework	Vijayakumar, J K; Murthy, T A V; Khan, M T M	2004	8	5
Assessment and Evaluation of Usage of UGC Infonet E-journals Consortium in North East Universities	Chand, Prem; Devi, Th Satyabati; Chauhan, Suresh K	2006	8	5
The Role of Open Source Software in Building Institutional Rrepository	Deka, Dipen	2006	8	5

5.6 Co-Authorship Networks

Co-authorship is joint works by two or more authors on a paper. The below network diagram shown in below figure-1 represents co-authorship network. In this study, there were 535 unique authors who presented 445 papers in PLANNER conventions. Among them, only 69 authors have more than 2 papers each which has been considered as cut-off for this analysis. Further, among 69 unique authors, 17 authors had written papers by themselves without any collaboration with other authors. Thus, only those 52 unique authors have been considered to generate this co-authorship networks diagram.In this resultant diagram, each node represents author and edge represents collaborative paper(s) written by those networked authors which is mentioned as the weight of edge. However, size of each node indicates no. of papers written by a particular author which is mentioned in square braces []. From this network diagram, we could see that two authors: Murthy, TAV and Devi, Th. Madhuri have presented highest 12 nos. of papers each in PLANNER conventions. Two authors: Murthy, TAV (12) and Chand, Prem (10) have a strong network and collaborations with other authors.

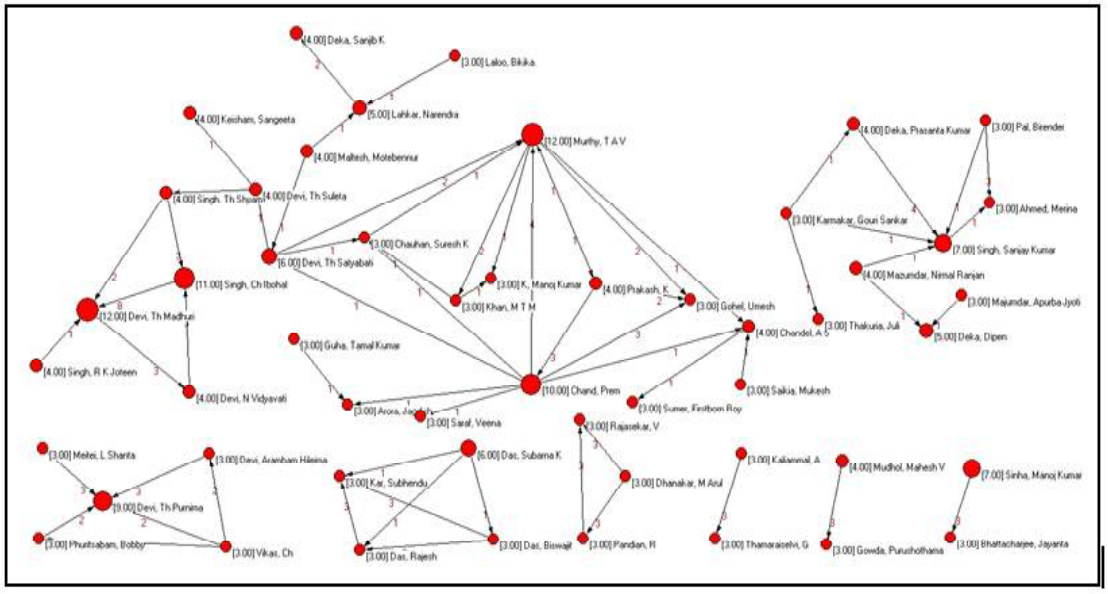


Figure1: Co-Authorship Networks

5.7 Keywords Clustering

A proper list of keywords provided in a research work / paper basically represents the main research interests of that concerned paper. The Network Figure 2 represents keywords clustering of author defined keywords retrieved from all 445 papers considered for this study. A total of 1002 unique keywords were retrieved from all papers and the keywords which were appeared in more than 3 papers were identified. Thus, there were a total of only 77 keywords considered to generate this network after limiting those to 3 papers.

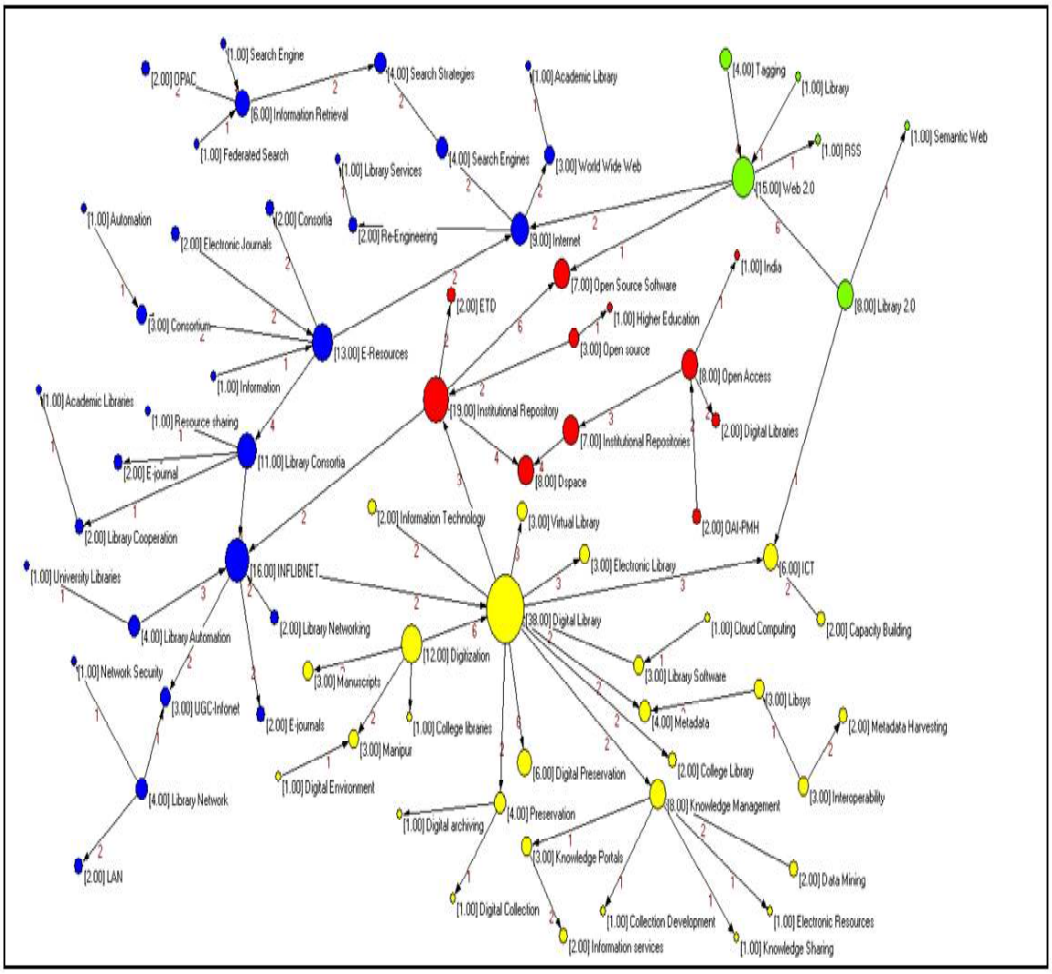


Figure 2: Keywords Clustering

In this resultant network Figure 2, each node represents keywords, and edge of connected nodes represent keywords defined in a particular paper. The size of each node indicates the number of papers wherekeyword has been defined which is mentioned in square braces []. The weight of each edge indicates number of papers where connected keywords have been appeared. From this diagram, we can observe that keywords like digital library (38), institutional repository (19), INFLIBNET (16), Web 2.0 (15) and e-resources (13) were appeared on top 5 respectively and frequently used by authors in their papers. It means, these 5 topics were the most preferred by authors of PLANNER conventions.

The PLANNER / year wise clustering of keywords has been represented in these below nine network diagrams which are named as per their year from 2003 to 2014 respectively.

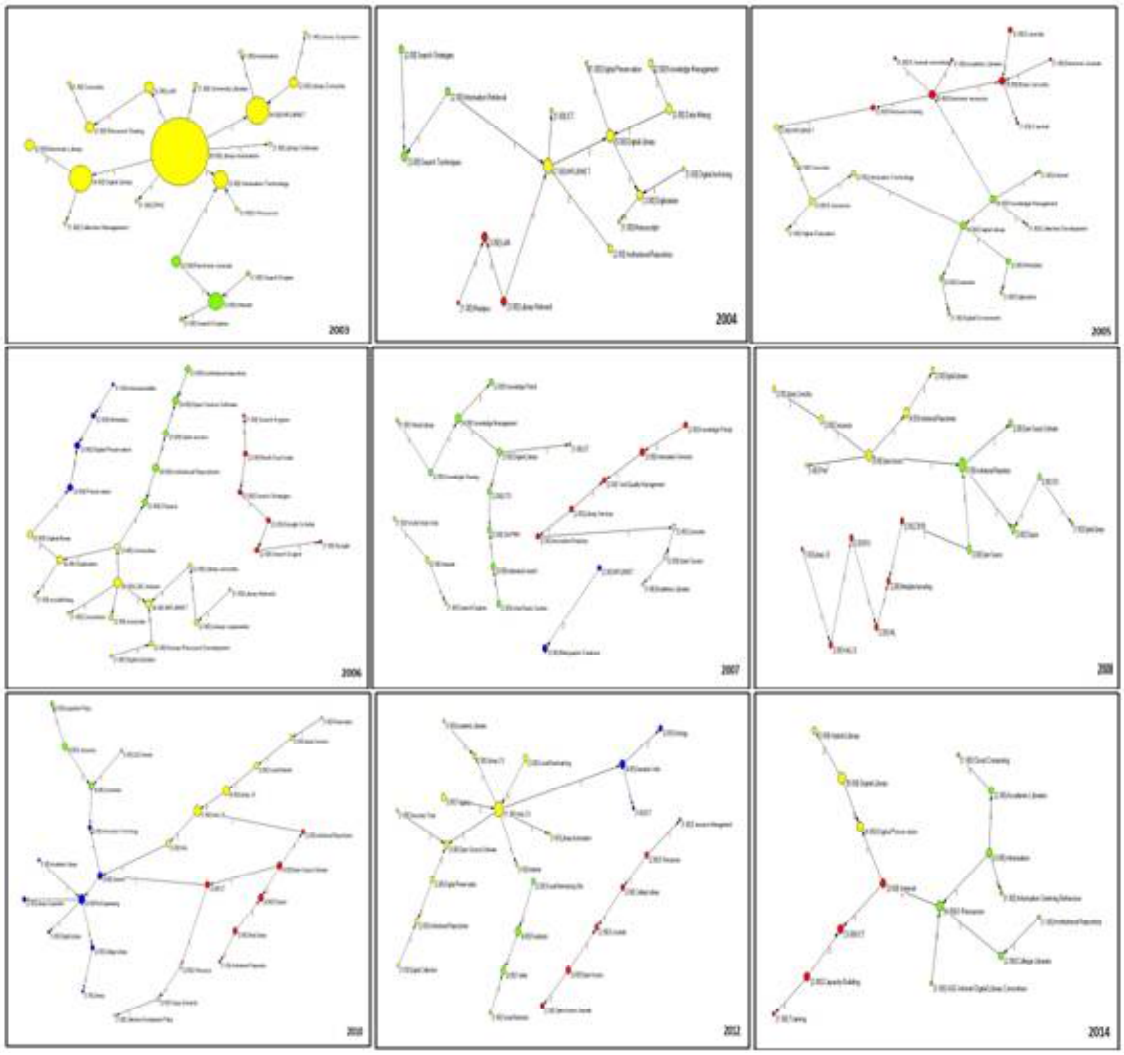


Figure 3: 2003 to 2014: Year wise Keywords Clustering

5.8 State Collaboration

The network Figure-3depicts state collaboration which reveals the collaboration trend in between states in 445 papers presented in PLANNER conventions. Collaboration was classified as state when the author(s) / authors' affiliations were from other Indian state institutions. There were authors from 28 states of India and 4 other countries (Bangladesh, England, Bhutan and Swaziland) who presented papers in the conventions. Among them, only 26 states whose have more than 2 papers each (cut-off) has been considered for this analysis without considering other countries. Further, among those 26 states, 11 states had papers without any collaboration with other states. Thus, only those left 15states have been considered to generate this state collaboration networks diagram.

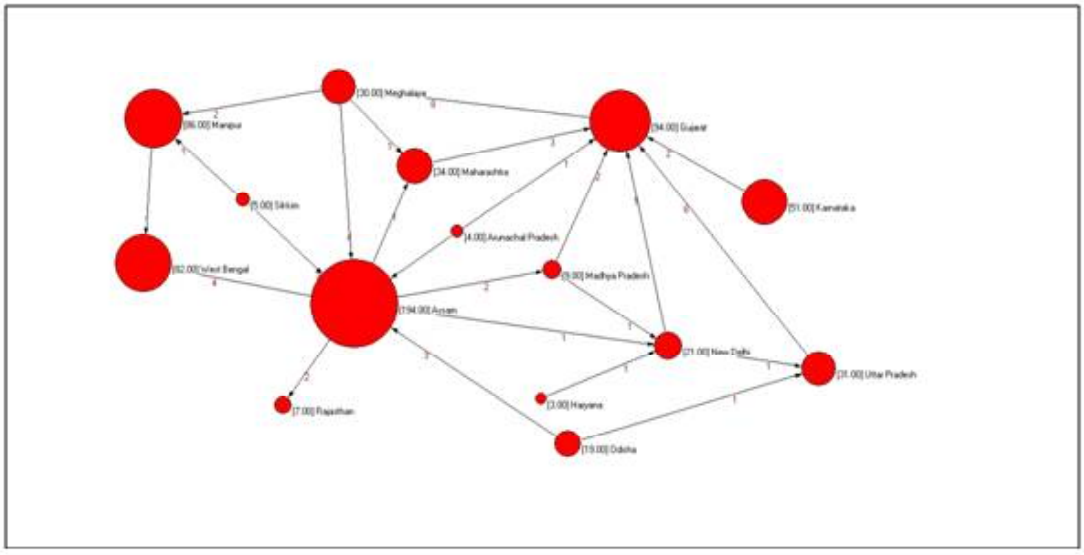


Figure 4 : State Collaboration

In the network Figure-3, each node represents affiliated state of that paper/author and edge represents collaboration between affiliated states. The size of each node indicates the number of papers belongs to that affiliated state which is mentioned in square braces []. The weight of each edge indicates number of collaborative papers between affiliated states. This network reveals that the state Assam had highest number of papers (194), followed by Gujarat (94). Assam had the most collaborative papers with other states such as: Rajasthan (2), West Bengal (4), Sikkim (3), Meghalaya (4), Maharashtra (1), Arunachal Pradesh (1), Madhya Pradesh (2), New Delhi (1), and Odisha (3). The states such as: Nagaland, Mizoram and Tripura had no collaboration with other states.

6. Conclusion

The presented study reveals that all the papers presented in PLANNER conventions were mostly collaborative in nature with an average of 1.86 authors per paper. In context of citations, there is a huge gap in between the nos. of cited and un-cited papers presented in the conventions. In terms of nos. of papers, Dr. TAV Murthy and Dr. Th. Madhuri Devi were the most prominent authors with highest 12 nos. of publications each to their credit. The keywords clustering network shows the important topics that were presented in the conventions which reveals that topics like: "digital library", "institutional repository", "Web 2.0", and "e-resources" were the most preferred and favourite topics of authors. The state collaboration networks reveal that the state of Assam had highest rate of collaboration having a strong tied-up with all other states. Further, the North-Eastern states affiliated universities/ institutions had strong collaboration with each other internally as well other parts of the country specially in library and information science field in PLANNERs.

References

1. Dash, N. K., Sahoo, J., & Mohanty, B. (2015). Evolution of Library Assessment Literature – A Bibliometric Analysis of. In 10th International CALIBER-2015, HP University and IIAS, Shimla, Himachal Pradesh, India, March 12-14, 2015.

2. Janakiramaiah, M., Doraswamy, M. (2011). Measuring Impact Of Web Resources In Conference Proceedings/ : A Citation Analysis. In 8th International CALIBER - 2011, Goa University, Goa, March 02-04, 2011.

3. Pareek, S. (2013). A Bibliometric analysis of the literature of IFLA. Library Philosophy and Practice, 1(4), 1–16. Retrieved from http://digitalcommons.unl.edu/libphilprac/954

4. Sordo, M., Ogihara, M., & Wuchty, S. (2015). Analysis of the Evolution of Research Groups and Topics in the Ismir Conference. In 16th International Society for Music Information Retrieval Conference (ISMIR 2015) (pp. 204–210).

5. Tsay, M. (2011). A Bibliometric Analysis on the Journal of Information Science. Journal of Library and Information Science Research, 5(2), 1–28. Retrieved from http://lac3.glis.ntnu.edu.tw/vj-attachment/2011/07/attach69.pdf

6. http://www.inflibnet.ac.in/planner

7. http://ir.inflibnet.ac.in

About Authors

Mr. Pallab Pradhan, Scientist B (LS), INFLIBNET Centre, Gandhinagar, Gujarat.
Email: pallab@inflibnet.ac.in

Mr. Hitesh Solanki, Scientist B (CS), INFLIBNET Centre, Gandhinagar, Gujarat.
Email: hitesh@inflibnet.ac.in

Mr. Swapnil Patel, Scientist B (CS), INFLIBNET Centre, Gandhinagar, Gujarat.
Email: swapnil@inflibnet.ac.in