An Analysis of Co-relation between Social Network Citation with H-Index and its Variants

Bilal Ahmad[1], Ahthasham Sajid[2], Haque Nawaz[3], M.M. Rind[3], Basit Hassan[3], Abdullah Maitlo[4], S.A. Ruk[5]

[1]Department of Computer Science, Capital University of Science & Technology, Islamabad, Pakistan
2Department of Computer Science, BUITEMS, Quetta, Pakistan
[3]Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan
[4]Department of Computer Science, Shah Abdul Latif University, Khairpur, Pakistan
[5]Department of Computer Science, Shah Abdul Latif University, Ghotki Campus, Pakistan

Summary
Altmetrics is complementary to the traditional bibliographic impact measurement techniques. This study explores the extent to which Altmetrics co-relate with the bibliographic impact measurement techniques: H-index and its variants. Analysis is accomplished based on around 70,000 published papers of 45categories from the field of mathematics. Number of tweets containing title, URL, or DOI of research papers has been compared with the H-index, G-index, HC-index, M-quotient and HW-index of the scholar. Analysis has been done for around 57,155 authors and shows that with the very low twitter citation rate that is 1.47%, almost no correlation exists between Twitter citation data and conventional impact measurement techniques. It implies that the social media metrics (Altmetrics) does not reflect the same kind of metrics as bibliometric indices. It further implies that Altmetrics should also be considered along with bibliometric indices to access an author's popularity

Key Words:
Altmetrics, Bibliometrics, Co-relation, H-Index variants, Tweetation, Social media Analysis.

1. Introduction

Altmetrics is the study of non-traditional scholarly impact measurement techniques that are based on activity in web-based environments[1]. Altmetrics is an emerging field, which unlike the traditional citation impact metrics, such as impact factor and h-index, does not rely just on citation counts, but also takes into consideration other features of impact such as the number of knowledge bases that referred to the work, the number of times the work was viewed/downloaded, and/or mentions in social/news media [2,3]. Because of its potential of measuring impact of scholarly work on both scholarly as well as non-scholarly community, measuring social media impact is gaining attention from researchers, reflecting the significance of Altmetrics. Indicators, e.g. research council arguing to use Altmetrics for evaluation of authors [4] and scholars wishing to include Altmetrics into curriculum vitae [5], express the potential of these metrics. Activities on the social media platform like CiteULike, Facebook and Twitter can be monitored to predict the impact of scholarly article [6]. Studies show that social media platforms like Twitter are used for dissemination [7], science popularization[8], and scholarly product promotion[9]. However, the authenticity and reliability of these metrics is questioned [10]. As compared to other social media platforms, Twitter is much more extensive[1] with 288 million active users[12], is one of the most popular platforms for dissemination of scholarly articles[13] and a commanding tool to disseminate pointers (e.g. links) to information[14] with hash tag, @messages and retweet [14]. Terms like Tweetations[15] and citation tweets [16] are used for tweet count. Studies have also explored the involvement of research community for dissemination of links or documents via Twitter[11]although a detailed study has not yet been performed to support these claims. This research provides a comprehensive analysis to estimate the potential of Altmetrics in gauging scientific impact by analyzing mathematical publications. Bibliographic impact measuring techniques ignore the impact created by an article in non-scholar community. To measure the impact of a scholarly research in general community: scholar as well as non-scholar community, Altmetrics came into existence. Social media platforms are now the best way to share anything at any time. When it comes to scholarly research, it becomes critical to measure the impact of scholarly document and the way it is helping the non-scholarly community. Micro blogging platform Twitter is one of the best known mean to predict the impact of scholarly article in the near future.

Articles are generally cited more frequently on social networks than on published papers. Being cited shows the quality of the work produced by the specific author; there are many methods available to perform citation analysis on the basis of citations of the paper. H-index, G-Index, M-quintet, HC-index and HW-index are common impact factor measurement techniques which reflect the productivity of an author with citation count. Articles are also cited online on social networks, this reveals the quality of that article due towhich it has been cited. So author should also be given appreciation on his work getting mentioned on social platforms. Keeping in view the

Manuscript received May 5, 2018
Manuscript revised May 20, 2018

importance of social network citation, we perform an analysis of both traditional bibliographic citation indexes (H-index, G-index, M-quintet,HC-index and HW-index) and Twitter citation (Altmetrics). In this work, our focal point is to identify authors that have higher indices value and their articles are trended on the Twitter as well. We try to figure out if the author with higher index value gets higher citation on Twitter as well? That would be estimated with the help of co-relation between bibliographic impact measurement techniques and Altmetrics. We also try to answer: how closely related they both are and, the other way around i.e. how different they are?

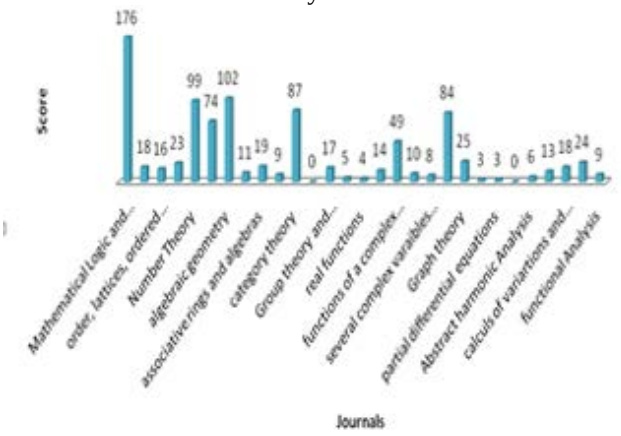

Fig. 1 Research Papers tweeted from various journals.

2. Literature Review

In the past, author's research has been accessed by the number of times it has been mentioned by other authors in their research articles called citation. Now with advancement of the web, there is a huge number of authors creating their contents and mentioning others on the digital web. Numbers in the citation database "WebOfScience" and "Scopus" show how well an author is perceived in scholarly community but it ignores the impact of that scholarly product on non-authors [17]. To cater this issue, terms like "webometrics" [18], "Scholarometer"[19]and "Altmetrics" [16] come into play.Term "Altmetrics" was first used by Jason Priem in his tweet on 29 September 2010 which is a short form of Alternative (Citation) Metrics [20]. Although Altmetrics claim to capture impact from a broader public but still it cannot replace the bibliographic scholarly impact [10], nonetheless it acts as a compliment to the traditional citation system[21]. More ever some scholars do not encourage correlation coefficient for comparing Altmetrics and different citation indicators for papers published in different time period due to time constraints.

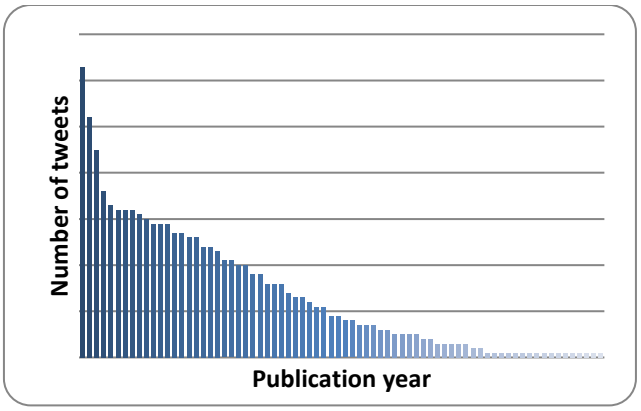

Fig. 2 Shows that latest papers get more attention on the Twitter as compare to the older one. As publication year gets older the number of tweets decreases.

Social bookmarking systems [22] are used to tag, share and bookmark an article over the social web and Twitter more frequently and with shorter time span [23]. "Altmatrics.com"and"Impactstory.org" are two main sources that provide social impact data from different social platforms, including how many data and knowledge bases have referred to a work, it's article views, number of its downloads, and its mentions in social and news media.

S.NO	AUTHOR ID	AUTHER First Name	AUTHER's Last Name	TWEETS	G-Index	H-Index	CORRL-h	CORL-g
1	7866	L	Zhongkui	10000	1	1	-0.3758512	-0.3759116
2	20457	L	Bo	10000	2	1		
3	40238	GA	Barnard	10000	2	2		
4	57250	S	Barrilai	10000	1	1		
5	71	G	Longo	4948	10	7		
6	57697	EJ	Horvitz	4117	7	7		
7	21609	JH	Palmieri	4107	3	3		
8	59161	A	Alves	4088	1	1		
9	59162	Y	Panis	4088	1	1		
10	59163	D	Trancart	4088	1	1		
11	59164	JM	Reginbeau	4088	1	1		
12	2259	SO	Hansson	3547	2	2		
13	2260	T	Grüne-Yanoff	3547	1	1		
14	12899	AV	Stepanov	2815	3	2		
15	22238	C	Petronio	2762	4	3		
16	410	P	Schuster	2755	5	5		
17	228	T	Coquand	2754	14	8		
18	2261	H	Lombardi	2754	7	5		
19	54203	H	Formby	2471	2	1		
20	23400	R	Steiner	2353	13	13		

Fig. 3 Top twenty authors who score highest on the Twitter along with their H-Index and G-Index. Correlation coefficient

3. Research Methodology

3.1 Data Collection

For analysis, the dataset of around 70,000 papers of about 57533 authors and 45 categories of math was borrowed froma recent research by Imama. The dataset was collected majorly from Google Scholar.

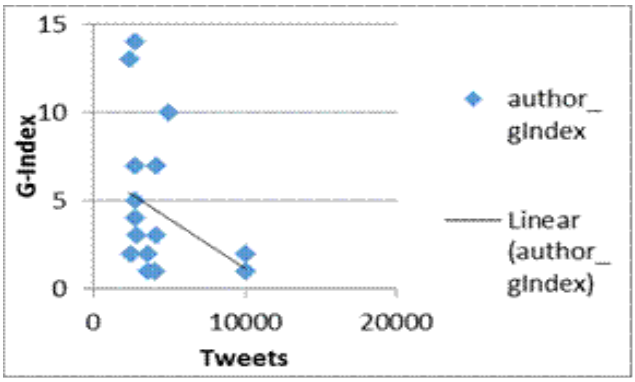

Fig. 4 Co-relation Coefficient of G-index and Tweetation of first 20 authors [Round -0.4].

Motivation Behind choosing mathematics field for analysis was that it covers vast knowledge as compared with any otherfield of study. Math classifications commonly known as MSC (Mathematics Subject Classification) had done 64 top level categories in last version of MSC classification MSC2010. 45 of them refer pure Mathematics while rest of them are related to applied mathematics.In this study we concentrate on the 45 sub categories that refer pure mathematics. To find out the correlation between the bibliographic impact factor and the Altmetrics, other data like H-index, G-index, M-quintet, HC index and HW-index value of the authors have been calculated using standard formulas used for calculating indices [23]. Tweet count has been calculated using Twitter API with title, URL and DOI of the paper.

3.2 Analysis

In different set of analysis, we first show the extent to which scholarly documents from the field of mathematics are found on the Twitter. We also explore the degree to which these citations vary over time. To perform this analysis data set of 69196 papers were used. Citation rate over Twitter, simply calculated by finding mean of tweeted papers, is calculatedand distribution of tweets per article is mentioned. We also identify top twenty papers which have been cited frequently over Twitter.

In second phase of analysis our focus is to distinguish and identify number of papers that have been tweeted at least once. Percentage of tweeted documents or Twitter courage P% tweeted and the mean that is Twitter Citation Rate T/Ptweeted [16]were calculated. For this calculation, only those articles have been considered which were tweeted at least once. Exclusion of the articles which are never tweeted leaves us with 1618 papers. Spearman correlation has been calculated between Twitter citation of papers and traditional bibliographic impact factor techniques H-Index and its variants. Finally analysis of correlation between Twitter citation count and Index value of all 57533 authors are mentioned that is calculated with statistical correlation (Spearman's) that is a common approach to validate new matrices by examining the correlation between them. Journals that are most frequently tweeted have also been listed in the results.

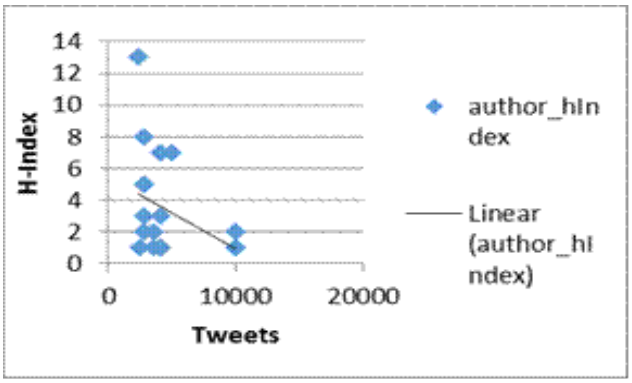

Fig. 5 Correlation Coefficient of H-index and Tweetation of first 20 authors [Round -0.4].

3.3Limitations

Replication is considered as one of the big hurdles while dealing with the Altmetrics data. First and very obvious one is Altmetrics data providers may change or suspend their presence with the passage of time. Secondly, it is nearly impossible to collect complete data from the internet sources; there is always a chance to miss some of information while collecting data from internet sources. Although, we make sure that all the data collected regarding Altmetrics must be complete and up-to-date. However, still the chances of an error cannot be denied. For tweet count we relied on the Twitter API for tweet count (which itself caries some technical limitations), which are obtained through document's title, DOI and URL. Therefore, evaluations of results associated with Twitter are based on its internal criteria on which it considers a tweet related to a specific paper. Another basic question asked about social media platforms is reliability. Reliability itself comes with many questions about completeness, authenticity, comprehensiveness etc. In addition, we did not know how download speed, time zone restrictions, and server downtime etc. would affect available data. On the bases of above mentioned limitations, we have ensured data integrity by crosschecking collected data with other Altmetrics data sources like Impactstory.org to minimize the effect of limitations

4. Results and Discussion

Analysis shows that only 1.47% (1021 out of 69196) scholarly documents get citation on the Twitter at least once which is very low tweetation rate indeed. From these 1021

documents 6.17% papers were published in 2013, 4.47% in 2012 and round 3% of papers were published in 2003, 2007, 1998, 2002, 2008 and 2006. Figure1 shows the numbers of tweets score by a document based on its publication year. With the mean of 89.69, from 1021 tweeted papers only four papers score more than 10,000 citations over Twitter which covers 0.391% of the tweeted documents. Remaining 98.5% (68175 out of 69196) do not get any citation on Twitter. From the 45 categories of the mathematical domain only 29 categories were mention on Twitter that is 64.44% of the total categories.

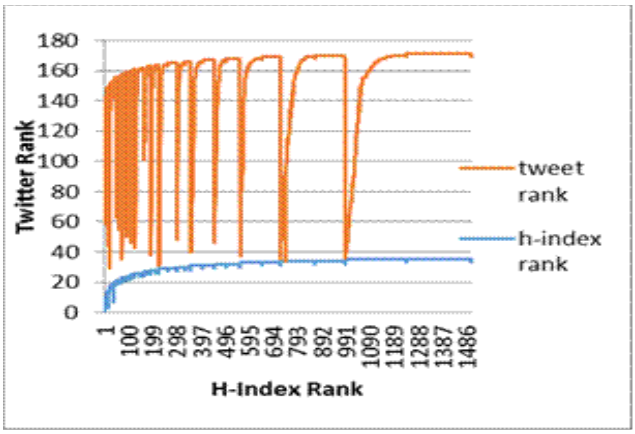

Fig. 6 Rank correlation between H-Index and Twitter that is round 0.23091

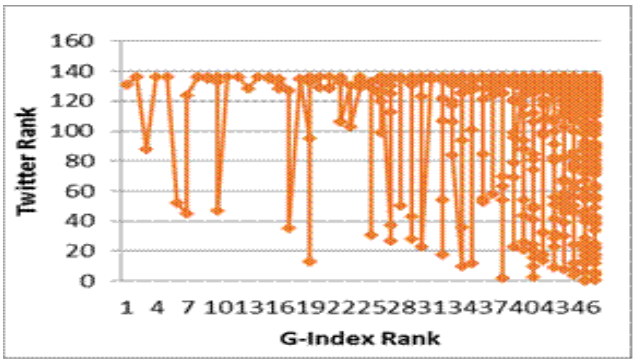

Fig. 7 Rank co-relation between G-Index and Altmetrics that is round 0.025936.

Round 34% (16 out of 45) categories do not get any citation over Twitter. Figure 2 shows the results of categories mentioned frequently on Twitter. With 7.009% from the total tweeted Categories, "Mathematical Logic and Foundations" were the most frequent tweeted category; scored 176 tweets of different papers that are 3.62% of the total dataset and from the tweeted 1021 papers it covers 17.23% of the tweeted data. "Algebra geometry" remains second with 102 tweet score. Likewise, form dataset of 57534 authors only 16% of them were tweeted at least once on the Twitter. Figure 3 shows the top twenty authors that get higher citations on the Twitter.

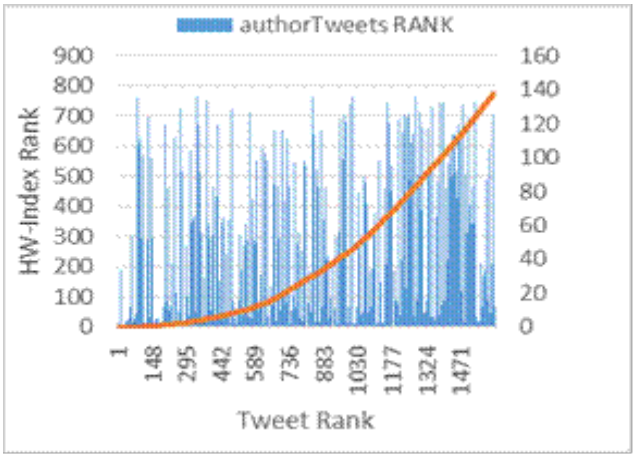

Fig. 8 Rank co-relation between Twitter rank and HW-Index that is round 0.067626.

Average tweets of an author are calculated by simple mean that is 1.305. From 1618 tweeted authors 766 score 1 that is 47.34% of the total tweeted authors. While 13.53 get two tweets. 97.18% authors do not get any citation over Twitter. Of the top twenty authors who have highest H-Index value, only ten get mentioned on the micro blogging platform that is 0.979% of the total tweeted data set With around -0.4 correlation coefficient values, no correlation exists between the first twenty authors with higher H-Index value and the Altmetrics data (Figure 4).

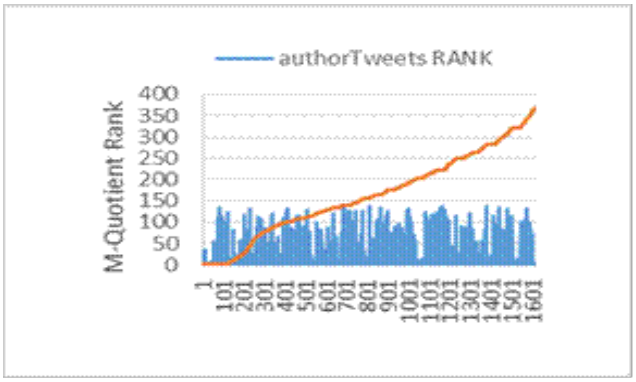

Fig. 9 Rank correlation between twitter and M-quotient that is round 0.008069.

Same is the case with the G-Index value; from top twenty authors with higher G-Index value, only eight authors work was tweeted at least once on the Twitter that is 0.783% of the total tweeted dataset. As far as correlation is concerned, there is negative but very low correlation (-0.3). Comparing the result of H-Index and G-Index of first twenty authors; G-Index shows some promise as compared to H-Index and

Altmetrics data (0.0203) that indicates no relationship. On the other side, with H-Index correlation coefficient with Altmetrics data of complete authors' data set comes with no correlation (0.0185). But omitting the non cited data on Twitter gives only about one thousand results. Figure 5 shows the correlation between rank value of Altmetrics and H-Index while figure 6 shows the correlation between Altmetrics and G-Index values.

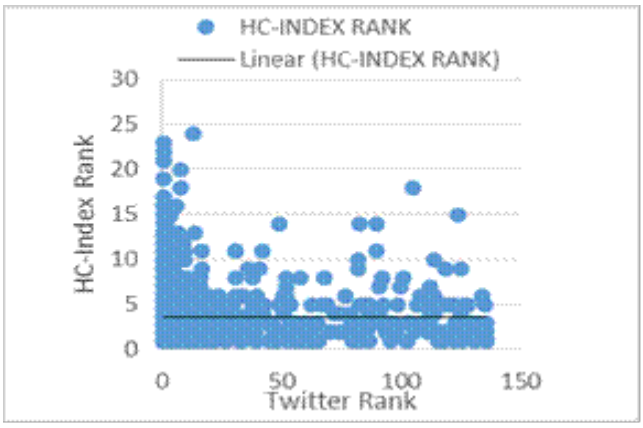

Fig. 10 Rank co-relation between twitter and HC-Index that is -0.00411

Figures 8, 9, and 10 shows the co-relation of HW,M-quotient and HC with Twitter data respectively. With the co-relation figure of 0.067626 HW- Index shows very low correlation with the twitter citation but still the trend line shows the positive behaviour of the co-relation. In case with M- Quotient 0.008069 shows less correlation with the twitter data as compared with the H-Index value. Rank correlation with the HC-Index does not show positive intent with value -0.00411.

5. Conclusion

This study covers entire continuum of the mathematical domain which provides reasonable data for the evaluation of Altmetrics. We classify distinctive relationships between Altmetrics and bibliographic citations, which can be used as a roadmap to evaluate twitting behaviour of people in other fields of study. The outcome of less than 2% of documents mentioned on the Twitter shows a very low coverage of mathematical documents on Twitter, which most probably can be due to scholarly focus on the traditional sources of impact measurement techniques. However, we were able to demonstrate that some categories are more popular than others. Less correlation between indices and Altmetrics shows that Altmetrics and indices measurement techniques are faraway from each other which means; Altmetrics cover other type of impact that is not comparable with the traditional citation system and therefore bibliometric indices should not be considered alone as being representative of an author's goodness, but altmetrics should also be taken into consideration to have a complete picture of the author's impact, that is his impact in author as well as non-author community.

References

[1] Ron Daniel Jr., P 2014, Research Trends, viewed 1 September 2015, < HYPERLINK "http://www.researchtrends.com/issue-37-june-2014/predicting-citation-counts/" http://www.researchtrends.com/issue-37-june-2014/predicting-citation-counts/ >.

[2] Mcfedries, Paul. "Measuring the impact of altmetrics [Technically Speaking]." IEEE Spectrum 8, no. 49 (2012): 28.

[3] Galligan, Finbar, and Sharon Dyas-Correia. "Altmetrics: Rethinking the way we measure." Serials review 39, no. 1 (2013): 56-61.

[4] Viney, Ian. "Altmetrics: research council responds." Nature494, no. 7436 (2013): 176.

[5] Piwowar, Heather, and Jason Priem. "The power of altmetrics on a CV." Bulletin of the Association for Information Science and Technology 39, no. 4 (2013): 10-13.

[6] Boyd, Danah, Scott Golder, and Gilad Lotan. "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter." In System Sciences (HICSS), 2010 43rd Hawaii International Conference on, pp. 1-10. IEEE, 2010.

[7] Darling, Emily S., David Shiffman, Isabelle M. Côté, and Joshua A. Drew. "The role of Twitter in the life cycle of a scientific publication." arXiv preprint arXiv:1305.0435 (2013).Nitin Vijaykumar Swami1, "Li-Fi (LIGHT FIDELITY) – THE CHANGING SCENARIO OF WIRELESS COMMUNICATION," J. IJRET., eISSN: 2319-1163, pISSN: 2321-7308

[8] Sugimoto, Cassidy R., and Mike Thelwall. "Scholars on soap boxes: Science communication and dissemination in TED videos." Journal of the Association for Information Science and Technology 64, no. 4 (2013): 663-674.

[9] Symes, Mark D., and Leroy Cronin. "Decoupling hydrogen and oxygen evolution during electrolytic water splitting using an electron-coupled-proton buffer." Nature chemistry 5, no. 5 (2013): 403.

[10] Liu, Jean, and Euan Adie. "Five challenges in altmetrics: A toolmaker's perspective." Bulletin of the Association for Information Science and Technology 39, no. 4 (2013): 31-34.

[11] Thelwall, Mike, Stefanie Haustein, Vincent Larivière, and Cassidy R. Sugimoto. "Do altmetrics work? Twitter and ten other social web services." PloS one 8, no. 5 (2013): e64841.

[12] Dr. Friedrich Schwandt, TK 2015, Global social networks ranked by number of users 2015, viewed 1 September 2015, < HYPERLINK "http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/" http://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ >.http://pediaa.com/difference-between-wifi-and-lifi/

[13] Weller, Katrin, Evelyn Dröge, and Cornelius Puschmann. "Citation Analysis in Twitter: Approaches for Defining and Measuring Information Flows within Tweets during Scientific Conferences." In # MSM, pp. 1-12. 2011.

[14] Costas, Rodrigo, Zohreh Zahedi, and Paul Wouters. "Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective." Journal of the Association for Information Science and Technology 66, no. 10 (2015): 2003-2019.
[15] Eysenbach, Gunther. "Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact." Journal of medical Internet research 13, no. 4 (2011).
[16] Priem, Jason, and Kaitlin Light Costello. "How and why scholars cite on Twitter." Proceedings of the Association for Information Science and Technology 47, no. 1 (2010): 1-4.
[17] Bar-Ilan, Judit, Stefanie Haustein, Isabella Peters, Jason Priem, Hadas Shema, and Jens Terliesner. "Beyond citations: Scholars' visibility on the social Web." arXiv preprint arXiv:1205.5611 (2012).
[18] Thelwall, Mike, Liwen Vaughan, and Lennart Björneborn. "Webometrics." ARIST 39, no. 1 (2005): 81-135.
[19] Kaur, Jasleen, Diep Thi Hoang, Xiaoling Sun, Lino Possamai, Mohsen JafariAsbagh, Snehal Patil, and Filippo Menczer. "Scholarometer: A social framework for analyzing impact across disciplines." PloS one 7, no. 9 (2012): e43235.
[20] Haustein, Stefanie, Timothy D. Bowman, and Rodrigo Costas. "Interpreting" altmetrics": viewing acts on social media through the lens of citation and social theories." arXiv preprint arXiv:1502.05701 (2015).
[21] Costas, Rodrigo, Zohreh Zahedi, and Paul Wouters. "Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective." Journal of the Association for Information Science and Technology 66, no. 10 (2015): 2003-2019.
[22] Reher, Sabrina, and Stefanie Haustein. "Social bookmarking in STM: Putting services to the acid test." Online 34, no. 6 (2010): 34-42.
[23] Shuai, Xin, Alberto Pepe, and Johan Bollen. "How the scientific community reacts to newly submitted preprints: Article downloads, twitter mentions, and citations." PloS one 7, no. 11 (2012): e47523

Mr. Bilal Ahmed has done his Master of Science in Computer Science from Capital University of Science & TechnologyIslamabad.He is serving as visiting faculty member at PRESTON University Islamabad, Pakistan. His research interests are (Cloud Computing, Networks, and Machine Learning)

Mr. Ahthasham Sajid is working as Assistant Professor, UITEMS, Quetta, Pakistan. He is PhD Scholar in Computer Science from SZABIST, Islamabad, Pakistan. He has 10 years of teaching experience at various keypositions at university. His areas of Research Interest is Wireless Networks and Opportunistic Networks. He has more than 17 national and international publication.

Mr. Haque Nawaz Lashari is persuing his PhD in Computer Science from Shaheed Zulfikar Ali Bhutto Institue of Science and Technology, Karachi, He received his MS degree from Mohammad Ali Jinnah University Karachi from Network and Telecommunication in in 2010. He is working as Lecturer at Sindh Madressatul Islam University, Karachi. He has more than 17 research publications in national and international journals. His areas of research interests are wireless communication, network security, routing protocols, optimization algorithms and mobility management in mobile ad hoc networks

Dr. Malook earned his Ph.D (Information Technology) from International Islamic University Malayia. He has more than 14 years of diversified industry, teaching and research experience. He is an active researcher having over 30 technical articles published and presented in various ISI, Scopus, IEEE, HEC recognized journals and conferences. His research interests are SDN, communication & networks, M-commerce, network security, pervasive computing, social networking and information system.

Mr. Basit Hassan holds a Master of Information Systems from the University of Melbourne, Australia, and a Master of Science in Computer Science from Iqra University, Karachi. He has more than two years of experience as lecturer and also served as Software Engineer specializing in Enterprise Resource Planning Systems design and implementation. His areas of interest include Software Engineering, Data Modeling and System Engineering.

Dr Abdullah Maitlo assistant professor department of computer science shah abdul latif university khairpur. dr maitlo has done phd from university of central lancasaire UK. Currently he is working as assistant professor of cyber security and knowledge management. further more his interest of research includes encryption knowledge economy of social network security and cyber crime.he has worked at coventry university UK for one year.

Mr. Sadaquat Ali Ruk is working as HOD of Computer Science Department at Shah Abdul Latif University, Ghotki Campus. His of areas of research are both theoretical and experimental research. The main focused on data mining & NLP techniques. He is the founder of Geeks Academy which is a Digital Skills training institute. Previously he was the consultant to Government of Sindh for implementation of various automation projects for Works& Services Department.