Unified Author Ranking based on Integrated Publication and Venue Rank

Ammad Usmani and Ali Daud
Department of Computer Science and Software Engineering, International Islamic University, Pakistan

Abstract: *Authors' ranking can be used to determine authenticity of authors in particular domain. Several different methods for author ranking focusing on number of publications and number of citations are proposed. In this paper, we propose ranking algorithms for publications, conferences, journals and respective authors. In publication ranking, both incoming and outgoing citations are considered. In case a publication is published in a well-reputed venue (conference or journal) then it is expected to have a high number of citations. Resultantly, due importance is given to venues and their scores are computed from popularity of their publications. Both publications' ranking and venue scores are used to rank authors, where authors having published in well reputed venues would have added benefits. We used multiple features to rank publications and venue effectively. These scores are then further used for ranking authors, instead of just using the number of citations for author ranking. Results of comparative study show a significant improvement in author ranking due to the inclusion of proposed features.*

Keywords: *Publication, venue, ranking, author ranking, pagerank.*

Received July 21, 2014; accepted February 10, 2015

1. Introduction

The ranking methods developed for ranking internet pages [3, 8, 21] can also be applied to research publications where visits/hits are replaced by the citations an author receives as a result of her/his work [4]. Every year different researchers publish a large number of papers to present the latest research. It is often difficult for new researchers to find out quality work in their field. These academic ranking not only help the newcomers to find meaningful work but also provide a measure to quantify seasoned researchers.

The idea behind ranking of authors is to establish the reputation of individuals producing certain contents. It is assumed that an author having well reputation provides valuable contents as compared to a low repute author. Now, the question is how to rank the author? One common method in this regard is to count the total number of citations of Gupta *et al.* [12]. A significant amount of work has been done for devising methodologies for ranking Li *et al.* [15, 16, 19].

While ranking publications, (incoming) citations play an important role towards high ranking, the quality of a paper can also be evaluated through its referenced papers, which were referred either for recognition of work or for improvements [4, 7, 11, 16, 23, 25]. Existing methods either used number of citations or extended pagerank algorithm for author ranking but they did not consider importance of venues in which papers are published.

Significance of author ranking is dependent on publications rank and the venues in which her/his papers are published. In this work we propose the algorithms for ranking publications, venues and authors. In publication ranking we modified pagerank [3] algorithm by adding outgoing links of papers and for ranking venues we proposed Venue Rank (VR) algorithms that based on our publication ranking. Finally, in computation of author rank, we unified both publications and venues rank.

The rest of the paper is organized as follows: Section 2 provides literature review and section 3 is about our proposed algorithms Integrated Publication Rank (IPR), *VR* and Unified Author Rank (UAR). Section 4 describes experimental setup and results and discussions. Finally, section 5 concludes this work and provides future directions.

2. Literature Review

Ranking methods can be classified in page ranking, publication ranking and author ranking. The journey of ranking algorithms started in 1998 when Brin and Page proposed pagerank [3] algorithm. pagerank is a proprietary method for measuring "importance" of a webpage [20] based on link structure of the web [12]. They applied citation analysis for web search by taking incoming links as citations to the web page. With the passage of time many researchers made enhancements in pagerank algorithm and used their algorithms for ranking authors.

Variations of pagerank algorithm has also been used in multi-document summarization for enhancing the summarizer performance [13]. Yu *et al.* [24] proposed timed pagerank which adds the temporal factor to the pagerank to pay attention to the newly published high quality papers into the search result [1]. Emphasis of this algorithm is to promote new pages [22]. A decaying rate is defined so that the latest links would get the highest weights [10].

Pagerank algorithm was applied by Chen *et al.* [4] on limited publication's citation network of American Physical Society (APS) journal to find scientific Gems. Liu *et al.* [16] proposed the co-authorship link weight in pagerank algorithm to form an author rank. This weight describes how strongly the authors are related to each other. It assesses the impact of an individual author [8]. Their approach is domain specific and it doesn't take advantage of different numbers of citations between Fiala *et al.* [9]. Ding *et al.* [7] also addressed author ranking in co-citation networks. Their main focus was to highlight the importance of different values of damping factors to model the freedom of citing any papers by authors. Effective and unbiased ranking of scientific literature thorough mutual reinforcement was also performed [14]. Inventor status was discovered and its effect on knowledge diffusion was studied in nanotechnology literature [17]. Yan and Ding [23] implemented pagerank algorithm on co-authorship network and assigned more weights to authors who have more citations. All the existing methods ignored outgoing citations and venues importance which is considered by us in this paper.

3. Unified Author Ranking

In our proposed method we have considered out-going links as a weight for ranking publications. In case an author references other's publication then he/she should get credit. Generally a good researcher refers high-quality publications and a lot of effort is evolved in literature survey so we have added this factor in our algorithm of ranking publications. Furthermore, we have considered venues importance in ranking of authors because credibility of publication is directly associated with the venue where it is published. When an author publishes their paper in a well reputed venue then they should get extra credit. While ranking authors their publication rank and venue rank are normalized.

We have performed author ranking in three steps. In first step, we propose *IPR* in which publications ranks are computed in iterative way then used the publication's ranking to rank the venues using *VR*. Finally we ranked authors using publications and venues scores using *UAR*.

3.1. Integrated Publication Rank

IPR is an extension of pagerank algorithm [3]. In which we have computed publications rank iteratively. We added out-going links of papers to this algorithm. *IPR* is defined as follows:

$$IPR(m) = (1-d) + d\left(\sum_{p_i \in M(p_i)} \frac{IPR(p_i)}{O(p_i)} + 1/\left(\sum_{p_j \in M(p_j)} \frac{IPR(p_j)}{I(p_j)}\right)\right) \quad (1)$$

Where $M(p_i)$ and $M(p_j)$ are the incoming and outgoing links of paper m respectively, $OM(p_i)$ are the number of outgoing links of paper p_i, $IM(p_j)$ are the number of incoming links of paper p_j and d is a damping factor. The *IPR* algorithm is as follows:

Algorithm 1: IntegeratedPublicationRank(PN).

```
#Set damping factor value
d=0.05
# For each publication in publication network
# initially set its Integerated Publication Rank to 1.0
foreach (P in PN)
{
    IPR[P][0]=1.0
}
for(i=1 to 40)
{
    foreach(P in PN)
    {
        # Initilize Inlink and OutlinkWeights
        IW=OW = 0
        IL=inlink_publications(P)
        OL=outlink_publications(P)
        # foreach inlinked publication IP in IL
        foreach (IP in IL)
        {
            OC=number_of_outlinks(IP)
            IW+=IPR[IP][i-1] / OC
        }
        # foreach outlinked publication OP in OL
        foreach (OP in OL)
        {
            IC=number_of_inlinks(OP)
            OW+=IPR[OP][i – 1] / IC
        }
        OW=1 / OW
        IPR[P][i]=(1- d)+d*(IW+OW)
    }
}
retrun IPR
```

3.2. Venue Rank

In this method we have calculated contribution of venue within entire network with respect to its *IPR*s citations and *IPR*. Therefore, sum of *VR* ratios within the network will be equal to one. In fact, we want to obtain actual share of venue over the whole network. Following equations defines *VR*:

$$TC(N) = \sum_{m \in P_i} CC(m) \quad (2)$$

Where $TC(N)$ is the total citations in the network N, P_i are the papers exist in the entire network N and $CC(P_i)$ are the citations count of papers P_i.

$$TPR(N) = \sum_{m \in P_i} IPR(m) \quad (3)$$

Where $TPR(N)$ is sum of all publication ranks in the network N, P_i are the papers exist in the entire network N and $IPR(P_i)$ are the IPR of papers P_i. Finally, we get VR as follows:

$$VR(v) = 1/2 \left(\frac{\sum_{m \in P_i} CC(m)}{TC(N)} + \frac{\sum_{m \in P_i} IPR(m)}{TPR(N)} \right) \quad (4)$$

Where $VR(v)$ is the VR of v by taking average of citation percentage and rank percentage, P_i are the papers published in venue v, $CC(P_i)$ are the citations count of papers P_i and $IPR(P_i)$ are the Integrated publication rank of papers P_i. The VR algorithm is as follows:

Algorithm 2: VenueRank(v).

```
# Publication Network of venue
PN=publications_network(v)
TC=sum_of_citation_count(PN)
TPR=sum_of_integrated_publication_rank(PN)
# Initilize Venue Citation Counts
VCC=0
# Initilize Venue Integrated Publication Ranks
VIPR=0
# for each publication in venue
foreach (P in v)
{
   CC=citation_count(P)
   IPR=integrated_ publication_rank(P)
   VCC+=CC
   VIPR+=IPR
}
VR =1/2*(VCC/TC+VIPR/TPR)
return VR
```

3.3. Unified Author Rank

We have computed author ranks by using IPR and VR. In addition, score is normalized by equally distributed author's rank among co-authorship. UAR is defined as follows:

$$UAR(A) = \sum_{m \in P_i} \frac{VR(m)}{2} \times \left(\left(\frac{CC(m)}{NA(m)} \times \frac{1}{VCC(m)} \right) + \left(\frac{IPR(m)}{NA(m)} \times \frac{1}{VIPR(m)} \right) \right) \quad (5)$$

Where $UAR(A)$ is the UAR of author A, P_i are the papers published by author A, $CC(P_i)$ is the citation count of paper P_i, $VCC(P_i)$ is the citation count of venue in which paper P_i is published, $IPR(P_i)$ represents the IPR of paper P_i, $VIPR(P_i)$ is the sum of IPR of venue in which paper P_i is published, $NA(P_i)$ are the number of authors in paper P_i and $VR(P_i)$ is VR (computed in Equation 4) of venue in which paper P_i is published.

Algorithm 3: UnifiedAuthorRank(A).

```
UAR=0
# for each publication P of Authour
foreach (P in A)
{
   CC=publicaion_citations_count(P)
   IPR=publicaion_integrated_publication_rank(P)
   NA=number_of_authors(P)
   V=publication_venue(P)
   VCC=venue_citations_count(V)
   VIPR=venue_integrated_publication_rank(V)
   VR=VenueRank(V)
   UAR+=(VR/2)*(CC / NA / VCC + IPR / NA / VIPR)
}
return UAR
```

4. Experiments

4.1. Dataset

We have taken the dataset from Digital Bibliography and Library Project (DBLP) [6] database and CiteSeerX scientific literature digital library [5]. DBLP dataset contains title and year of publication, publication authors and venues including conferences or journals, in which papers are published while CiteSeerX contains citations of publications. Therefore, we have combined these two datasets by matching publication's titles after computing hash for exact matching of titles in both datasets. After performing all necessary preprocessing steps we got refined dataset that contains publication's title, publication's authors, and publication's year, incoming and outgoing links of papers and venue in which papers are published.

4.2. Baseline Method and Performance Measurement

Weighted Pagerank (PR_W) proposed by Yan and Ding [23] is taken as baseline method. No ground realities are available about ranking of authors by considering outgoing links of papers and venues ranking. Therefore, we implemented both baseline and our proposed method on the same dataset and set the same parameter settings for both methods so that the comparison is not biased.

For evaluation, we have used several measures for comparison between baseline and UAR methods. These measures consist of authors' citations, authors' outgoing links; number of papers published, number of co-authors in a paper, VR and citations rank. Here, citations rank is the sum of the worth of the papers cited an authors papers which is calculated through IPR. Although, baseline method only includes citations weight-age and the rest of the measures are not applicable but we will discuss the importance of these in results and discussion.

4.3. Parameter Settings

We have used 0.5 value of damping factor for our experiments. In Web pages ranking, 0.85 damping value gives better results while 0.5 should be used in ranking of publications [18], it means that average

citation link in citation network would be ½.

4.4. Results and Analysis

4.4.1. Comparative Study

Top 15 authors obtained from *UAR* are shown in Table 1. We have also included the variation of authors' position with respect to the baseline method PR_W in terms of position up, position down and position stable. Leslie Lamport appeared on top in *UAR* while Rakesh Agrawal achieved top position in baseline method PR_W. Leslie Lamport moved up to first position in *UAR* and Rakesh Agrawal lost his position and moved down to second position. Rakesh has higher number of citations, higher number of papers published and higher venue score than Leslie Lamport but has more number of co-authors, and for that reason he lost his top position. Randal E. Bryant is ranked at number 3 with *UAR* while he was ranked 60 with the PR_W. One can see that Randal has published 73 papers but his papers were published in high quality venues which gave him high *VR* score and ranked him high. Every parameter in our algorithm plays a significant role in ranking of authors. As far as other ranking algorithms, they mostly focused on single parameter like citations. Considering only a single parameter might not be appropriate for ranking authors.

Some authors' positions remain same in *UAR* method because they have higher citations and published their papers in well reputed venue. Among these authors Hector Garcia-Molina got fourth position in both PR_W and *UAR* methods. In Table 1 Hector Garcia-Molina has most number of co-authors but has a very good *VR* score and also have more citations and high *IPR* score for his publications. These factors lead him to get fourth position in both PR_W and *UAR*.

Table 1. Top 15 Authors of UAR and PR_W (baseline method).

Authors	Citations	Out Links	Papers Published	Co Authors	VR	Citation's Rank	UAR	Positions		Variation in Positions
								UAR	PR_W	
Leslie Lamport	3898	235	76	65	0.3954	4189.26	0.0009831	1	8	+7
Rakesh Agrawal	8094	850	113	302	0.8379	8109.46	0.0009243	2	1	-1
Randal E. Bryant	2343	366	73	108	0.4161	2134.23	0.0006296	3	60	+57
Hector Garcia-Molina	5049	1588	236	558	1.1485	4731.62	0.0005823	4	4	0
W. Bruce Croft	3588	1138	157	294	0.895	3931.02	0.0005692	5	14	+9
Ben Shneiderman	2334	902	171	304	0.9022	2523.53	0.000566	6	61	+55
J. Ross Quinlan	4025	51	22	18	0.0673	4030.27	0.0005454	7	10	+3
David G. Lowe	1924	117	19	26	0.0849	1708.67	0.0005376	8	95	+87
Thorsten Joachims	2388	367	48	93	0.2124	2135.63	0.0005369	9	59	+50
Adi Shamir	2091	133	85	110	0.3464	1667.78	0.0005117	10	75	+65
Christos Faloutsos	4328	1836	188	466	0.972	4595.03	0.0005107	11	6	-5
Robert Endre Tarjan	2808	424	128	239	0.7476	2800.23	0.0005057	12	28	+16
Jiawei Han	5230	2545	278	781	1.288	4089.93	0.0004896	13	3	-10
Ramakrishnan Srikant	3870	256	30	69	0.2505	3385.56	0.0004885	14	12	-2
Christos H. Papadimitriou	3324	564	167	264	0.7026	3200.07	0.0004774	15	18	-3

4.4.2. Author's Profile

Table 2 provides an introduction of top 6 authors ranked through *UAR* which depicts the competency of authors. As most of them are members of famous organizations and also have won several research awards.

Table 2. Top 6 authors profile.

Name	Position	Organization	Awards
Leslie Lamport	Researcher	Microsoft Research in Mountain View, California	Dijkstra Prize (2000 and 2005), IEEE John von Neumann Medal (2008), ACM Turing Award (2013), initial developer of the document preparation system LaTeX
Rakesh Agrawal	Technical Fellow	Microsoft Research Labs	ACM-SIGMOD Edgar F. Codd Innovations Award, ACM-SIGMOD Test of Time Award, VLDB 10-Yr Most Influential Paper Award, and the Computerworld First Horizon Award
Randal E. Bryant	Dean	Carnegie Mellon University, USA	ACM Paris Kanellakis Theory and Practice Award, IEEE W.R.G. Baker Prize Paper Award, IEEE Emanuel R. Piore Award and Phil Kaufman Award by the EDA Consortium
Hector Garcia-Molina	Professor	Stanford University	1999 ACM SIGMOD Edgar F. Codd Innovations Award, Fellow of ACM
W. Bruce Croft	Professor	University of Massachusetts Amherst	ACM fellow, ASIS and T research award and Gerard Salton Award
Ben Shneiderman	Professor	University of Maryland, College Park	Member National Academy of Engineering, ACM Fellow, AAAS Fellow, IEEE Fellow, IEEE Visualization Career Award, SIGCHI LifeTime Achievement, Miles Conrad Award

4.4.3. Comparative Study in Terms of Publications and Average Citations

Average citations, average papers and average co-authorship of top 30 authors of *UAR* and baseline PR_W methods are shown in Figure 1. Average number of citations for both methods is divided by 20 to make clear comparison with average papers and co-authorship. Furthermore average values of these methods are rounded to 0 for clarity.

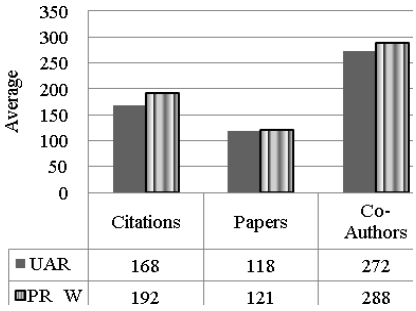

Figure 1. Comparison of authors on average of citations, papers and co-authorship.

Although, proposed *UAR* method is not only relying on number of publication and number of citations, several parameters in *UAR* including outgoing links of publications, venue rank and number of co-authors also affect author ranking. We noticed that in Figure 1, co-authors of baseline method (PR_W) is high than *UAR*, this may lead to increase the average citations of PR_W method.

To display the affect of co-authorship in average citations we have computed author rank through our proposed *UAR* method without distribution of equal credit among co-authors in same paper. Figure 2 shows this affect.

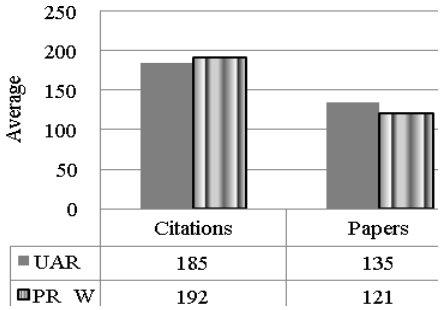

Figure 2. Average citations and papers of top 30 authors without including co-authors.

Since we are not focusing on citation count, average citations of PR_W (baseline) method are higher than *UAR* (proposed method). Furthermore, in Figure 2, it is clearly shown that by excluding distribution of equal credit among co-authors in *UAR*, average citations become high as compared to average citations in Figure 1.

4.4.4. Comparative Study on Specific Domain

We have also compared proposed method results with baseline method on domain specific dataset. For this purpose we have selected topic specific conferences related to data mining from ArnetMiner [2]. Authors' standings among *UAR* and PR_W methods are shown in Table 3.

Table 3. Author's position in UAR and PR_W method.

Author	Positions		Variation in Positions with Respect to PR_W
	UAR	PR_W	
Jiawei Han	1	1	0
Mohammed Javeed Zaki	2	5	+3
Philip S. Yu	3	2	-1
Pedro Domingos	4	3	-1
Charu C. Aggarwal	5	15	+10
Rakesh Agrawal	6	4	-2
Alexander Tuzhilin	7	10	+3
Heikki Mannila	8	7	-1
Jian Pei	9	6	-3
Christos Faloutsos	10	9	-1

In Table 3, author Mohammed Javeed Zaki has secured second position in *UAR* method, while he is on fifth position in PR_W method. He has taken second position from author Philip S. Yu whose citations, publications and venue rank score is higher than Zaki but due to his large number of co-authorship he has lost his position in *UAR* method.

Average citations, average papers and average co-authorship of top 30 authors of *UAR* and baseline PR_W methods are shown in Figure 3. Average number of citations for both methods is divided by 20 to make clear comparison with average number of papers and co-authorship. Furthermore, average values of these methods are rounded to 0 for clarity. One can see that for both methods average number of citations are same even average number of papers and average number of co-authors are less for *UAR*, although the difference in negligible.

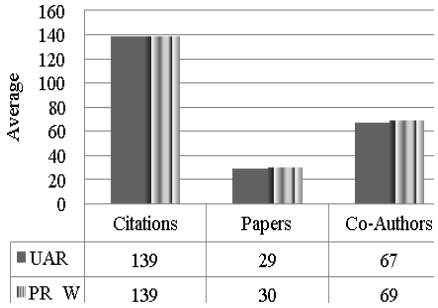

Figure 3. Comparison of authors on average number of: Citations, papers and co-authorship.

5. Conclusions

Like pagerank, publication ranking helps researchers to find quality publications more effectively. Discussion clearly shows the effectiveness of outgoing links and venues importance as these parameters have significant affect on author ranking. Multiple parameters consideration in this regard, such as, incoming and outgoing links of publications, venue importance, total number of co-authors is realistic. As a future work time factor for publications can also be incorporated for temporal rankings of authors.

References

[1] Ahmadi-Abkenari F. and Selamat A., "A Clickstream-based Focused Trend Parallel Web Crawler," *International Journal of Computer Applications*, vol. 9, no. 5, pp. 1-8, 2010.

[2] Arnetminer Search and Mining of Academic Social Networks, available at: http://arnetminer.org, last visited 2013.

[3] Brin S. and Page L., "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *in Proceedings of the 7th International World Wide Web Conference*, Australia, pp. 107-117, 1998.

[4] Chen P., Xie H., Maslov S. and Redner S., "Finding Scientific Gems with Google's Pagerank Algorithm," *Journal of Informetrics*, vol. 1, no. 1, pp. 8-15, 2007.

[5] Citeseer Scientific Literature Digital Library, available at: http://citeseer.ist.psu.edu/, last visited 2013.

[6] DBLP Computer Science Bibliography, available at: http://www.informatik.uni-trier.de/~ley/db/, last visited 2013.

[7] Ding Y., Yan E., Frazho A., and Caverlee J., "PageRank for Ranking Authors in Co-Citation Networks," *Journal of the American Society for Information Science and Technology*, vol. 60, no. 11, pp. 2229-2243, 2009.

[8] Dwork C., Kumar R., Naor M., and Sivakumar D., "Rank Aggregation Methods for the Web," *in Proceedings of the 10th International World Wide Web Conference*, Hong Kong, pp. 613-622, 2001.

[9] Fiala D., Rousselot F., and Ježek K., "PageRank for bibliographic networks," *Scientometrics*, vol. 76, no. 1, pp. 135-158, 2008.

[10] Gao B., Liu T., Liu Y., Wang T., Ma Z., and Li H., "Page Importance Computation Based on Markov Processes," *Information Retrieval*, vol. 14, no.5, pp. 488-514, 2011.

[11] Gollapalli S., Mitra P., and Giles C., "Ranking Authors in Digital Libraries," *in Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries*, Canada, pp. 251-254, 2011.

[12] Gupta S., Duhan N., and Bansal P., "A Comparative Study of Page Ranking Algorithms for Online Digital Libraries," *International Journal of Scientific and Engineering Research*, vol. 4, no. 4, pp.1225-1233, 2013.

[13] Hariharan S., Ramkumar T., and Srinivasa R., "Enhanced Graph Based Approach for Multi Document Summarization," *The International Arab Journal of Information Technology*, vol. 10, no. 4, pp. 334-341, 2013.

[14] Jiang X., Sun X., and Zhuge H., "Towards an Effective and Unbiased Ranking of Scientific Lliterature Through Mutual Reinforcement," *in Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, USA, pp. 714-723, 2012.

[15] Li E., Liao C., and Yen H., "Co-Authorship Networks and Research Impact: A Social Capital Perspective," *Research Policy*, vol. 42, no. 9, pp. 1515-1530, 2013.

[16] Liu X., Bollen J., Nelson M., and Sompel H., "Co-Authorship Networks in the Digital Library Research Community," *International Journal of Information Processing and Management*, vol. 41, no. 6, pp. 1462-1480, 2005.

[17] Liu X., Kaza S., Zhang P., and Chen H., "Determining Inventor Status and its Effect on Knowledge Diffusion: A Study on Nanotechnology Literature from China, Russia, and India," *Journal of the American Society for Information Science and Technology*, vol. 62, no. 6, pp. 1166-1176, 2011.

[18] Maslov S. and Redner S., "Promise and Pitfalls of Extending Google's PageRank Algorithm to Citation Networks," *The Journal of Neuroscience*, vol. 28, no. 44, pp. 11103-11105, 2008.

[19] Meng Q. and Kennedy P., "Discovering Influential Authors in Heterogeneous Academic Networks by a Co-Ranking Method," *in Proceedings of the 22nd ACM International Conference on Information and Knowledge Management*, San Francisco, USA, pp. 1029-1036, 2013.

[20] Modi H. and Narvekar M., "A Comparative Study of Various Page Ranking Algorithms," *International Journal of Engineering Research and Technology*, vol. 3, no. 1, pp. 3328-3333, 2014.

[21] Page L., Brin S., Motwani R. and Winograd T., "The PageRank Citation Ranking: Bringing Order to the Web," *Technical Report*, 1999.

[22] Pushpa S. and Easwarakumar K., "Exploring the Influence of Time-Sensitive and Unsupervised Learning of Topic-Specific Information in Citation Analysis," *European Journal of Scientific Research*, vol. 67, no. 3, pp. 474-485, 2012.

[23] Yan E. and Ding Y., "Discovering Author Impact: A PageRank Perspective," *International Journal of Information Processing and Management*, vol. 47, no. 1, pp. 125-134, 2011.

[24] Yu P., Li X. and Liu B., "Adding the Temporal Dimension to Search-A Case Study in Publication Search," *in Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence*, France, pp. 543-549, 2005.

[25] Zhang J. and Liu X., "Full-Text and Topic Based Author Rank and Enhanced Publication Ranking," *in Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries*, USA, pp. 393-394, 2013.

Ammad Usmani received MS degree in Computer Science from International Islamic University, Islamabad, Pakistan in 2014. He completed his thesis under the supervision of Dr. Ali Daud. He is currently working as a Database Administrator in National University of Sciences and Technology (NUST), Pakistan and he is also involved in teaching at NUST.

Ali Daud is working as Assistant Professor in the Department of Computer Science at International Islamic University, Islamabad. He obtained his PhD degree from Tsinghua University in 2010. He is head of Data Mining and Information Retrieval Group. He published about 23 papers in reputed international journals and conferences. He has taken part in many projects and PI of a project funded by higher education commission, Pakistan. His current research interests include: Text mining, social networks analysis and applications of probabilistic topic models. He has published 29 Journal and conference papers among some of them are in top tier.