How Healthy are ACM SIGWEB Sponsored Conferences?

Swati Agarwal
IIIT Delhi, India. Email: swatia@iiitd.ac.in

Nitish Mittal
NSIT, India. Email: nitishmittal94@gmail.com

Ashish Sureka
ABB Corporate Research, India. Email: ashish.sureka@in.abb.com

ACM SIGWEB sponsors and supports several conferences covering a wide range of topics. CIKM, HT, WEBSCI and WSDM are four of the many sponsored conferences by ACM SIGWEB. CIKM, HT, WEBSCI and WSDM are reputed conferences, being held for several years and has a large community of contributing authors and PC members. In this article, we present a study on the health of these four conferences based on several factors and metrics such as stability, openness, inbreeding, representativeness, sustainability, prestige and workload. Studying the health of a conferences provides a reflection and historical overview in-terms of its performance which can be used by the conference community to bring improvements and ensure that the conference is meeting its desired objectives. We conduct statistical analysis and information visualization on the conference data downloaded from DBLP, ACM Digital Library web-pages and conference websites. Our analysis reveals that overall all the four conferences are showing good health indicators. We do observe variances in metrics values across conferences and within conferences across years.

1. RESEARCH MOTIVATION AND AIM

The ACM Special Interest Group on Hypertext and the Web (ACM SIGWEB)[1] sponsors seven annual conferences covering a wide range of topics such as hypertext & hypermedia, web science, digital libraries, document engineering, knowledge & information management, information retrieval, web search, data mining, knowledge discovery from databases, social media, user adaptation & personalization and user modeling. The seven ACM SIGWEB sponsored conferences are: HT (Hypertext and Social Media), JCDL (Joint Conference on Digital Library), DOCENG (Symposium on Document Engineering), WEBSCI (Web Science Conference), CIKM (Conference on Information and Knowledge Management), WSDM (Web Search and Data Mining) and UMAP (User Modelling, Adaptation and Personalization). All the seven ACM SIGWEB Conferences are prestigious and pre-

[1] http://www.sigweb.org/

mier conferences in the area of Web and provides an international forum for researchers and practitioners to exchange ideas, interact with each-other and present latest research results.

The ACM SIGWEB sponsored conferences have been successfully running over so many years and have become an important academic event for researchers and practitioners working in the area of Web. We believe that historical analysis and reflection of the recent past of the conference is important for the SIGWEB community to further improve the quality and impact of the conference. We believe that a bibliometric based analysis can be done to understand how the conferences have evolved over so many years and is it meeting its desired objectives. We select four out of seven conferences for analysis as they are relatively larger and broad than the remaining three conferences. In this paper, we study CIKM, WEBSCI, HT and WSDM as their topical coverage is broad. The topical coverage of JCDL, DocEng and UMAP is narrow and focused in comparison to CIKM, WEBSCI, HT and WSDM. For example, CIKM invites papers on the broad area of information and knowledge management whereas the focus of DocEng is primarily on document engineering which is a sub-area within information and knowledge management. Our objective is to divide our study into parts: present our analysis of the broad focused conferences in this paper and present our study on the narrow topical coverage conferences in our future paper. Our definition of narrow focused conference does not mean less in terms of participation or impact and rather narrow in terms of its topical coverage with respect to the broader topical coverage conferences.

We conduct an analysis of CIKM, HT, WEBSCI and WSDM. DBLP[2] is popular on-line reference for open bibliographic information on computer science journals and proceedings. DBLP releases a snapshot of its bibliography database. We use the meta-data available from DBLP for our analysis. We extract the author(s) and paper titles for all the conferences in our dataset from DBLP dump and we extract information about the PC members and acceptance rate from the conference website and ACM Digital Library web pages.

Vasilescu et al. study the health of software engineering conferences with respect to community stability, openness to new authors, inbreeding, representativeness of the PC with respect to the authors community, availability of PC candidates, and scientific prestige [Vasilescu et al. 2014]. They propose several metrics which are indicators of the health of a conference and analyze nine conferences over a period of more than ten years [Vasilescu et al. 2014].

Our motivation is to study SIGWEB conferences from the perspective of the metrics defined by Vasilescu et al. The specific research aim of the work presented in this paper is to replicate the work by Vasilescu et al. [Vasilescu et al. 2014] on CIKM, HT, WEBSCI and WSDM conferences and assess the health of these conferences with respect to factors such as author and PC turnover representing community stability, inbreeding and openness to new authors as well as representativeness of the PC with respect to the authors community. Our aim is determine if the conferences have a balanced PC turnover, have high openness to new authors and have high, moderate or low inbreeding.

[2] `http://dblp.uni-trier.de/`

2. RELATED WORK AND RESEARCH CONTRIBUTIONS

In this section, we present closely related work to the study presented in this paper and list the novel research contributions of our work in context to existing work. We first present two focused bibliometric analysis based studies conducted on ACM SIGWEB conferences and then discuss few research studies which are on bibliometric analysis of conferences in various computer science domains. Agarwal et al. present a bibliometric analysis of seven ACM SIGWEB sponsored conferences (HT, JCDL, DOCENG, WEBSCI, CIKM, WSDM, UMAP) [Agarwal et al. 2016a]. They analyze the DBLP data of the seven conferences since their beginning until September 17, 2015 [Agarwal et al. 2016a]. Agarwal et al. present a bibliometric analysis of scientific publications records of nine SIGWEB cooperating conferences (ASONAM, COMPUTE, IWCMC, RecSys. RuleML, W4A, WebIST, WebMedia and OpenSym) [Agarwal et al. 2016b]. They use the DBLP bibliographical database and perform an exploratory and scientometric analysis on publications, authors and conference database of 10 years (2006-2015) of SIGWEB cooperating conferences [Agarwal et al. 2016b].

The work by Vasilescu et al. is very closely related to our work [Vasilescu et al. 2014]. We replicate their study which was conducted on software engineering conferences to four ACM SIGWEB conferences [Vasilescu et al. 2014]. Agarwal et al. present a study on gender gap, imbalance and women participation in computer science research by conducting experiments on DBLP bibliographical database and analyzing several years of publication dataset across various domains of computer science research [Agarwal et al. 2016].

Sharma et al. conduct scientific paper publication mining and scientometric and bibliometric analysis of nine years of ISEC (India Software Engineering Conference) publications and programs [Sharma and Sureka 2016]. Sharma et al. present insights from a bibliometric analysis and scientific paper publication mining of 551 papers in Requirements Engineering (RE) series of conference (11 years from 2005 to 2015) [Sharma et al. 2016]. Kumar et al. conduct a bibliometric and scientific publication mining based study to how the Asia-Pacific Software Engineering Conference (APSEC) has evolved over a period of six years (year 2010 to 2015) [Kumar et al. 2016].

Sakr et al. study the program committees of four top-tier and prestigious database conferences (SIGMOD, VLDB, ICDE, EDBT) over a period of 10 years (20012010) [Sakr and Alomari 2012]. Elshawi et al. study the computer systems research community by analyzing the research publications and the program committee memberships of three top-tier and prestigious computer systems conferences (EuroSys, SOSP and OSDI) over the period between 2006 and 2014 [Elshawi and Sakr 2016]. Bergamaschi et al. analyze three conferences of interest to the SIGWEB community (WWW, Hypertext & Hypermedia and JCDL) for a period of one decade [Bergamaschi et al. 2012]. Bartneck et al. present a reflection on the first five years of the International Conference on Human Robot Interaction (HRI) [Bartneck 2010]. The work presented in this paper makes the following novel and unique contributions:

First study on assessing the health of four ACM SIGWEB sponsored conferences (CIKM, HT, WEBSCI and WSDM) with respect to several factors and metrics defined by Vasilescu et al. [Vasilescu et al. 2014] such as author and PC turnover representing community stability, inbreeding and openness to new authors as well as representativeness

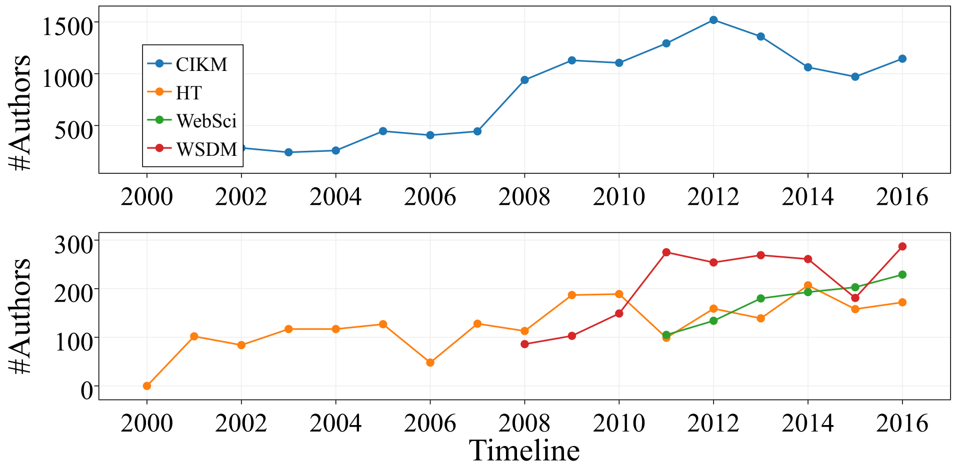

Fig. 1. **Basic Metrics Computations:** Demonstrates the Variation in Number of Distinct Authors for Conference C in Year Y.

of the PC with respect to the authors community

3. BASIC METRICS

Vasilescu et al. define two basic metrics: $\#A(c, y)$ and $\#C(c, y)[\#PCmem]$ [Vasilescu et al. 2014]. $\#A(c, y)$ denotes the number of distinct authors for conference c in year y. $\#A(c, y)$ represents the participation level and number of authors who belong to the conference community. $\#C(c, y)[\#PCmem]$ denotes the number of PC members for conference c in year y. $\#A(c, y)$ represents the size of the PC community and not the author community. $\#A(c, y)$ and $\#C(c, y)[\#PCmem]$ combined indicates how large the conference community is in-terms of the distinct number of authors and PC members.

3.1 #A(c,y)

Figure 1 shows the variations in number of distinct authors for all the four conferences in our dataset from the year 2000 to 2016. The first data point for conferences such as WEBSCI and WSDM is 2011 and 2008 respectively as started in these years.

We observe that $\#A(CIKM, 2000)$ is 173 and $\#A(CIKM, 2016)$ is 1145 resulting in a 6.61 times increase from 2000 to 2016. The highest value of $\#A(CIKM, y)$ is 1519 for the year 2012, The $\#A(CIKM, y)$ value is more than 1000 for the year 2009, 2010, 2011, 2012, 2013, 2014 and 2016. $\#A(CIKM, 2015)$ is 971 which is the only year from 2009 to 2016 in which the number of distinct authors is less than 1000. As shown in Figure 1, the biggest jump in the number of distinct authors is from the year 2007 to 2009. The $\#A(CIKM, 2009)$ value is 1129 whereas the $\#A(CIKM, 2007)$ value is 444.

We observe that the number of distinct authors for HT varies from a minimum of 48 ($\#A(HT, 2006)$) to a maximum of 207 ($\#A(HT, 2014)$). The $\#A(HT, 2001)$ value is 102 and $\#A(HT, 2016)$ is 172 resulting in a 1.68 times increase from 2001 to 2016. We

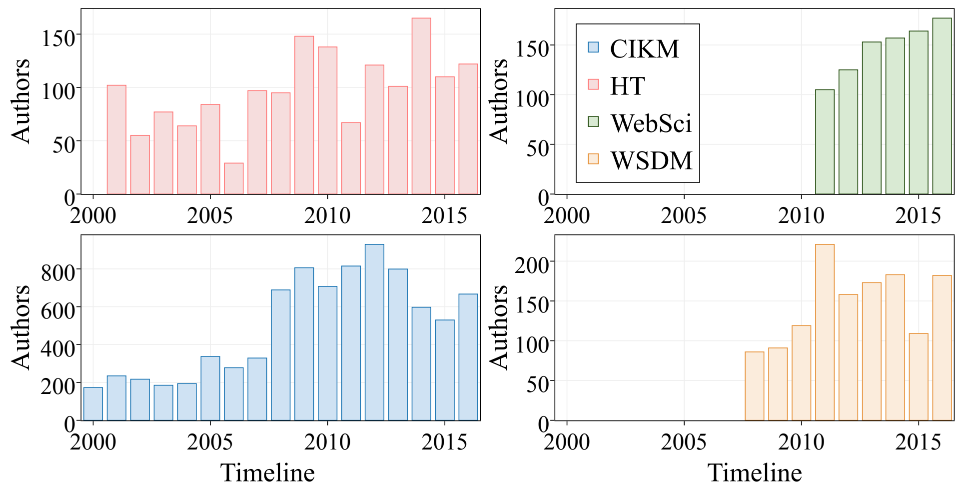

Fig. 2. Number of New Authors Publishing Articles in SIGWEB Conferences Every Year.

observe that the number of distinct authors for HT has always been more than 100 from 2007 to 2016 except for the year 2011 when the $\#A(HT, 2011)$ value is 99. As shown in Figure 1, the biggest jump in the number of authors is from the year 2006 to 2007 and the biggest decline is from the year 2010 to 2011.

We observe that $\#A(WEBSCI, 2011)$ is 105 and $\#A(WEBSCI, 2016)$ is 229 resulting in a 2.18 times increase from 2011 to 2016. The number of distinct authors for WEBSCI varies from a minimum of 105 in the year 2011 to a maximum of 229 in the year 2016. As shown in Figure 1, the $\#A(WEBSCI, y)$ value for WEBSCI has monotonically increased since its beginning until the year 2016. The biggest increase in the $\#A(WEBSCI, y)$ value is from the year 2012 to 2013 when 46 distinct authors got added.

We observe that $\#A(WSDM, 2008)$ is 86 and $\#A(WEBSCI, 2016)$ is 287 resulting in a 3.33 times increase from 2008 to 2016. The $\#A(WSDM, y)$ value ranges from a minimum of 86 in the year 2008 to a maximum of 287 in the year 2016. The number of distinct authors for WSDM is more than 250 for the years 2011 to 2014 and 2016. As shown in Figure 1, the biggest jump in the $\#A(WSDM, y)$ value is from the year 2010 to 2011 when 126 distinct authors got added. The biggest decline in $\#A(WSDM, y)$ is from the year 2014 to 2015 in which the number of distinct authors decreased from 261 to 181. While Figure 1 shows the variation in number of distinct authors for conference c in year y, Figure 2 shows the trend on number of new authors publishing articles in SIGWEB conferences every year.

3.2 #C(c,y) [#PCmem]

Figure 3 displays the number of PC members for conference c in year y which is denoted as: (#C(c,y) [#PCmem]). We observe that the number of PC members for CIKM various from a minimum of 266 (2007) to a maximum of 630 (2016). The #PCmem value for CIKM has consistently increased in the past three years from 2014 to 2016. Figure 3

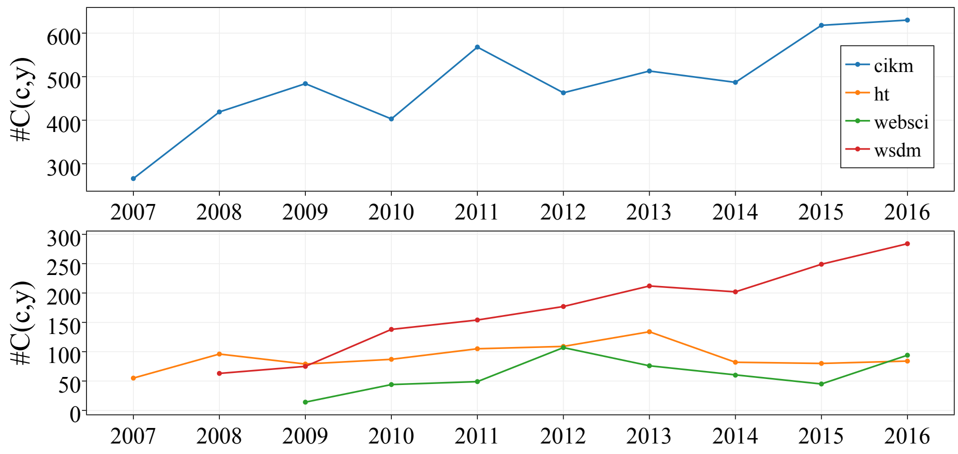

Fig. 3. **Basic Metrics Computations:** Number of PC members for conference c in year y. #C(c,y) [#PCmem]

reveals that except for the year 2007, the #PCmem value for CIKM has always been more than 400. We observe that the number of PC members for HT varies from a minimum of 55 in the year 2007 and maximum of 134 in the year 2013. The maximum decline in the #PCmem value for HT happened from the year 2013 to 2014 (134 to 82).

Figure 3 reveals that from year 2014 to 2016, the number of PC members for HT are consistently in the range of 80 to 85. Our analysis reveals that WEBSCI initially started as a small conference in which the number of PC members ranged from 14 to 49 from the year 2009 to 2011 and then gradually grew to a mid-size conference. Figure 3 shows that there was a major jump in the number of PC members for WEBSCI from the year 2011 to 2012 (49 to 107). Figure 3 reveals that except for a slight drop in one year (from 2013 to 2014), the #PCmem value for WSDM has increased consistently from the year 2008 to 2016. The number of PC members for WSDM has been above 200 from the year 2013 to 2016.

4. STABILITY

4.1 #$NA(c, y, n)$

#$NA(c, y, n)$ denotes the Number of New Authors for conference c in year y that were not author in years y - n to y - 1. As shown in Figure 4, we compute the #$NA(c, y, n)$ values for all the four conferences in our dataset and keeping the window size n equal to 4. Ability to attract new authors and author turnover is an important indicator of conference stability [Vasilescu et al. 2014]. A high author turnover shows that the conference is dynamic and able to attract new authors with respect to the previous editions of the conference. We observe that the stability metric #$NA(CIKM, y, 4)$ value for CIKM varies from a minimum of 173 in 2000 to a maximum of 968 in 2012. Even in the latest edition of CIKM (2016) with respect to this paper, the #$NA(CIKM, 2016, 4)$ value is 750.

The #$NA(CIKM, y, n)$ value is always above 570 from the year 2008 to 2016. HT is a relatively smaller conference than CIKM. The #$NA(HT, 2016, 4)$ value which is the

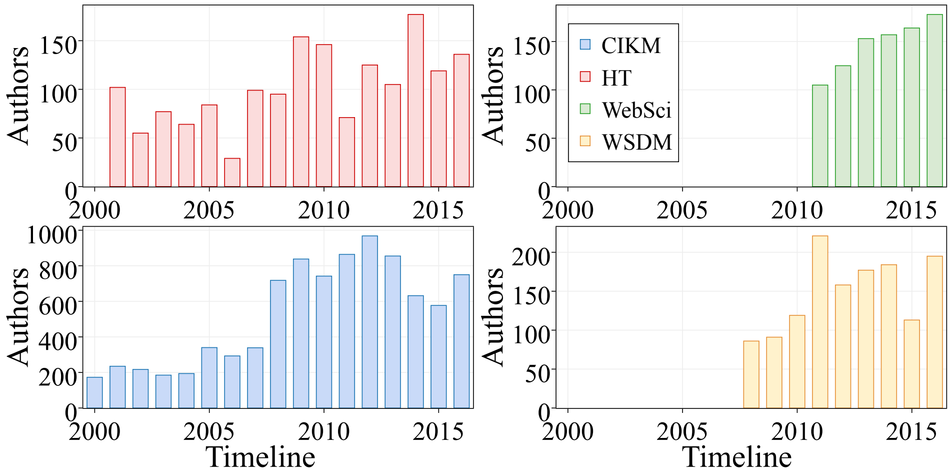

Fig. 4. **Stability Metrics Computation:** Number of New Authors for Conference C in Year Y that were not Author in Years $Y - N$ to $Y - 1$. N=4

most recent edition of the conference is 136 which shows that the conference is able to attract a substantial number of new authors with respect to previous four editions. The maximum value of $\#NA(HT, y, 4)$ is for the year 2009 (154) and 2010 (146). We observe that there is a jump from 2008 to 2009 (95 to 154) but major decline from 2010 to 2011 (146 to 71). The least number of author turnover happened from 2010 to 2011. The $\#NA(WEBSCI, y, 4)$ value for WEBSCI is always more than 100. Figure 4 reveals that the $\#NA(WEBSCI, y, 4)$ value has been monotonically increasing from 105 in the year 2011 to 178 in the year 2016. We observe that the maximum value of $\#NA(WSDM, y, 4)$ is for the year 2011 (221). Figure 4 shows that there is a significant jump in the metric $\#NA(WSDM, y, 4)$ value from the year 2015 to 2016 (113 to 195).

4.2 #NC(c, y, n) [#(real)newPCmem]

Figure 5 shows the trends on number of new PC members for conference c in year y that were not PC member in years y-n to y-1 (n=4). We observe that the #NC(CIKM, y, 4) value varies from a minimum of 171 (2010) to a maximum of 442 (2015). The graph in Figure 5 reveals that every year a large number of new PC members (with respect to past four years) are added in CIKM. We observe a major decline in the #NC(CIKM, y, 4) value from the year 2015 to 2016 (442 to 243). Figure 5 shows that the #NC(HT, y, 4) value varies from a minimum of 31 (2015) to a maximum of 82 (2012). In the year 2015 the #NC(HT, y, 4) value declined to 31 but then it increased in 2016 to 47.

Figure 5 reveals that the #NC(WEBSCI, y, 4) value has shown both an upward and downward trend. The maximum #NC(WEBSCI, y, 4) value is 78 for the year 2012. There was a decline in the #NC(WEBSCI, y, 4) value for the year 2015 from 47 to 38 but then it picked-up to 60 in the year 2016. Figure 5 that from year 2013 to 2016, the #NC(WSDM, y, 4) value is greater than 160. We observe that except for the year 2012, the #NC(WEBSCI, y, 4) is always greater than 100 from 2010 to 2016.

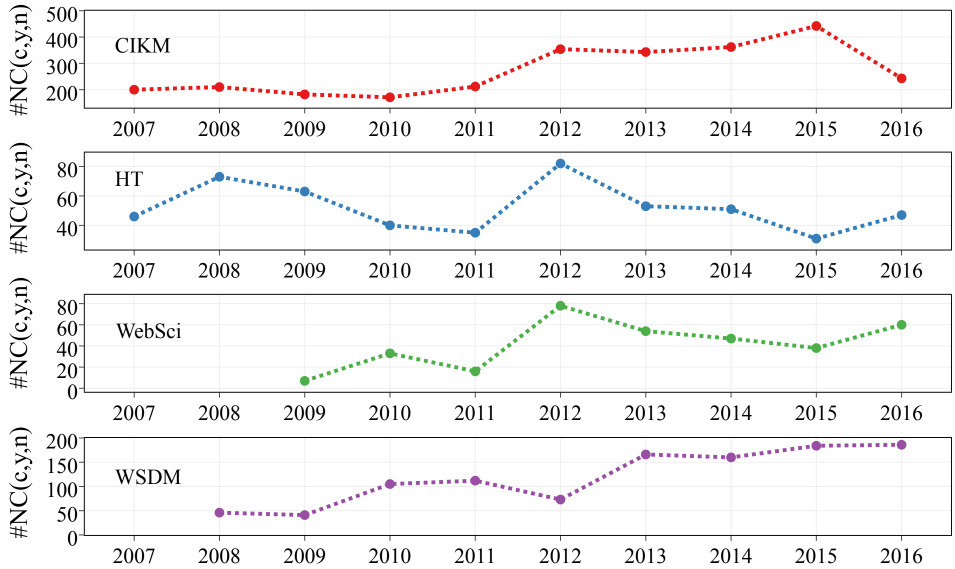

Fig. 5. **Stability Metrics Computation:** Number of New PC members for conference c in year y that were not PC member in years y-n to y-1. N=4

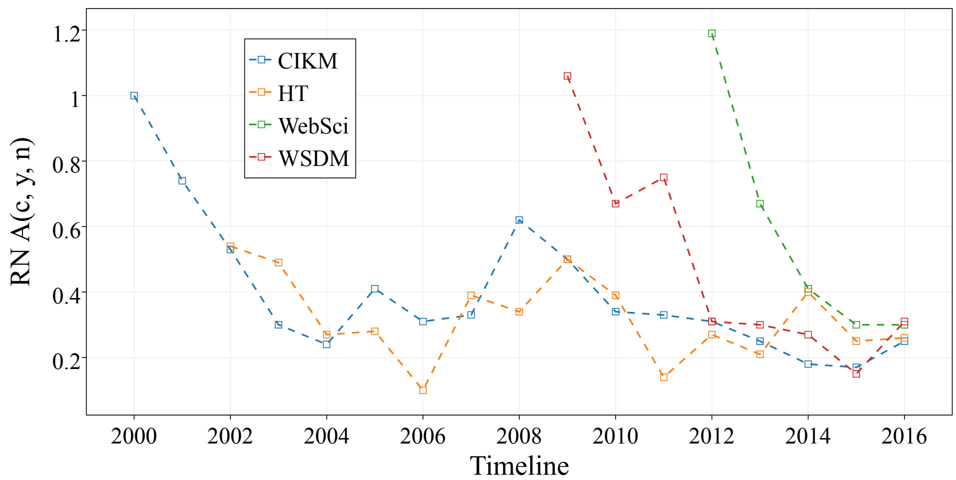

Fig. 6. **Stability Metrics:** Author turnover = Ratio of New Authors for conference c in year y w.r.t. years y-n to y-1

4.3 $RNA(c, y, n)$

$RNA(c, y, n)$ represents author turnover which is equal to the ratio of new authors for conference c in year y with respect to years y-n to y-1. Figure 6 displays the author turnover for all the four conferences in our dataset. Figure 6 reveals that the minimum author turnover for CIKM is 0.18 (2014) and 0.17 (2015). A value below 0.25 with a

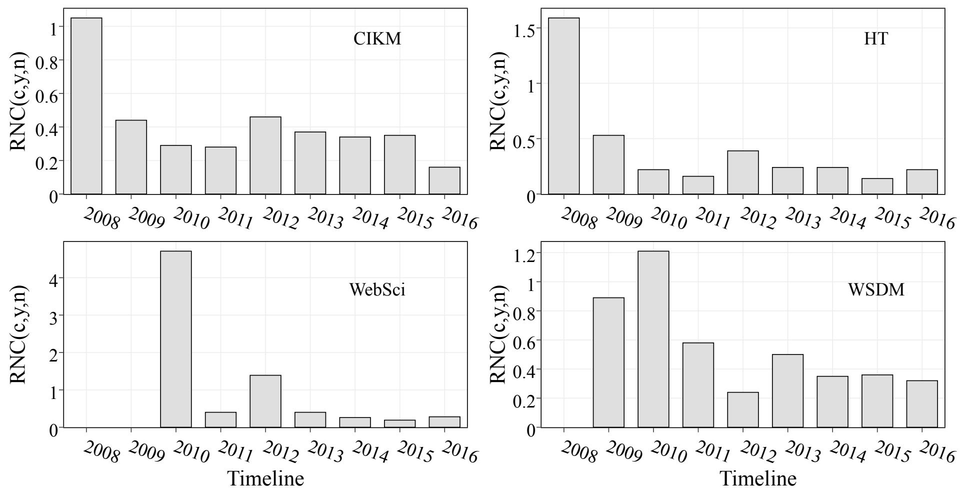

Fig. 7. **Stability Metrics:** PC turnover = Ratio of New programme Committee members for conference c in year y w.r.t. years y-n to y-1

window size of $n = 4$ is low. We found that from year 2013 to 2016, the author turnover is between 0.17 to 0.25 for CIKM. However, we observe a turnover of more than 0.5 for several years such as 2008 and 2009.

The $RNA(HT, y, 4)$ varies from a minimum of 0.10 (2006) to a maximum of 0.54 (2002). We observe that out of 15 years (2002 to 2016), the $RNA(HT, y, 4)$ value is below 0.25 for four years. For HT, the biggest decline in author turnover is from 0.39 to 0.14 (2010 to 2011). WEBSCI has a high author turnover every year. The minimum author turnover for WEBSCI is 0.30 . Figure 6 reveals a decline in the author turnover value for WEBSCI from 2014 to 2015 but it remained constant from 2015 to 2016. Figure 6 reveals that in general the author turnover for WSDM is high and above 0.25 except for one year (2015) when the $RNA(WSDM, y, 4)$ value is 0.15.

4.4 RNC(c, y, n) [(real)newPCprop]

RNC(c, y, n) [(real)newPCprop] denotes PC turnover which is equal to the ratio of new programme Committee members for conference c in year y with respect to years y-n to y-1. Figure 7 displays the PC turnover for all the four conferences. We observe that RNC(CIKM, y, 4) for CIKM was between 0.34 to 0.46 from years 2012 to 2015 but declined in the year 2016 to 0.16. The RNC(HT, y, 4) value of HT ranges between 0.14 to 0.24 from the year 2013 to 2016. Figure 7 reveals that there was a decline in the RNC(HT, y, 4) value of HT from 2014 to 2015 (0.24 to 0.14) but then it picked-up to 0.22 in the year 2016. The RNC(WEBSCI, y, 4) value of WEBSCI shows both upward and downward trend. The RNC(WEBSCI, y, 4) value was 0.4 in the year 2011 and 2013. Figure 7 reveals a consecutive decline for two years after 2013 but then shows an increase in year 2016. We observe that the RNC(WSDM, y, 4) value for WSDM is in the range of 0.32 to 0.36 from 2014 to 2016. Figure 7 shows that the maximum jump in RNC(WSDM, y, 4) value is from 2012 to 2013.

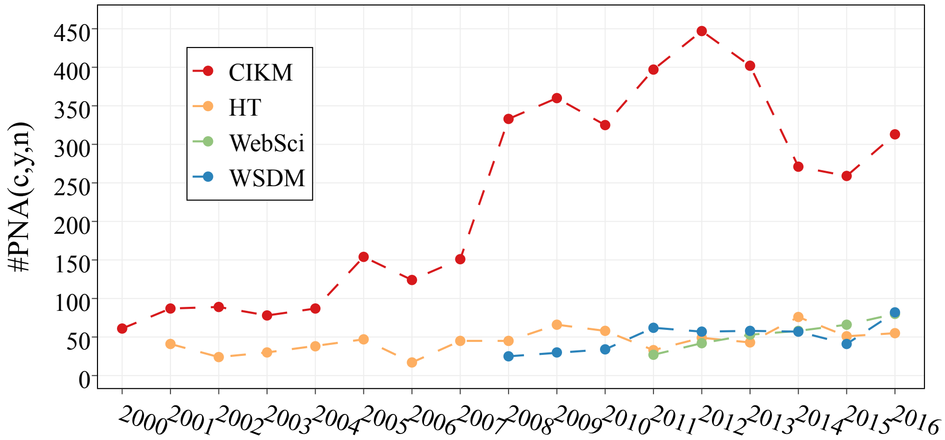

Fig. 8. **Openness Metrics Computation:** Number of Papers of conference c in year y by New Authors for which none of the co-authors has published at this conference in years y-n to y-1

5. OPENNESS

5.1 #PNA(c, y, n)

#PNA(c, y, n) denotes number of papers of conference c in year y by new authors for which none of the co-authors has published at this conference in years y-n to y-1. Figure 8 shows trends on the #PNA(c, y, n) metrics for the four conferences. We observe that the #PNA(CIKM, y, 4) value follows a declining trend from 2012 to 2016. There is a steady decline from 447 to 259 (from 2012 to 2016). The maximum jump in the #PNA(CIKM, y, 4) value is from 2007 to 2008 doubling from 151 to 333.

The #PNA(HT, y, 4) value for HT varies from a minimum of 17 (2006) to a maximum of 76 (2014). In comparison to 2014, the #PNA(HT, y, 4) value declined sharply from 76 to 51. Figure 8 reveals that the #PNA(WEBSCI, y, 4) value increases consistently from 27 in year 2011 to 66 in year 2015. We observe that the #PNA(WSDM, y, 4) value for WSDM remains in the range of 57 to 62 for four years (2011 to 2014) but declines sharply in 2015.

5.2 RPNA(c, y, n)

RPNA(c, y, n) denotes ratio of papers (by new authors) for conference c in year y for which none of the co-authors has published at this conference in years y-n to y-1. Figure 9 shows trends on the RPNA(c, y, n) metrics for the four conferences. The RPNA(CIKM, y, 4) value varies within a small range of 0.8 (2014) and 0.95 (2001). The RPNA(CIKM, 2016, 4) is 0.91 which is relatively higher than year 2006 to 2015. In comparison to CIKM, HT shows a much wider variance in terms of the RPNA(HT, y, 4) value. The minimum value is 0.62 (2002) and the maximum value is 0.92 (2014). In comparison to CIKM and HT, the variance in WEBSCI is less. The RPNA(WEBSCI, y, 4) value varies between the range of 0.89 and 0.96. Figure 9 reveals that the RPNA(WSDM, y, 4) declined from year 2014 (0.71) to 2015 (0.64) but then increased in year 2016 (0.88).

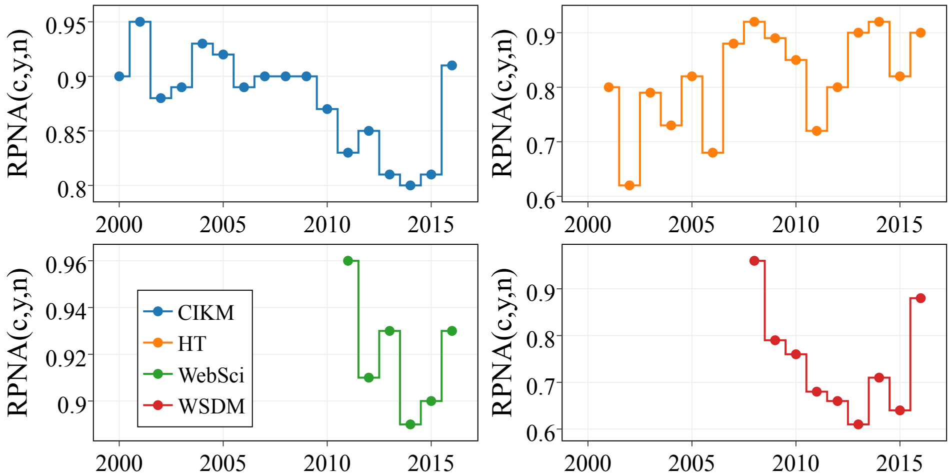

Fig. 9. **Openness Metrics Computation:** Ratio of Papers (by New Authors) for conference c in year y for which none of the co-authors has published at this conference in years y-n to y-1

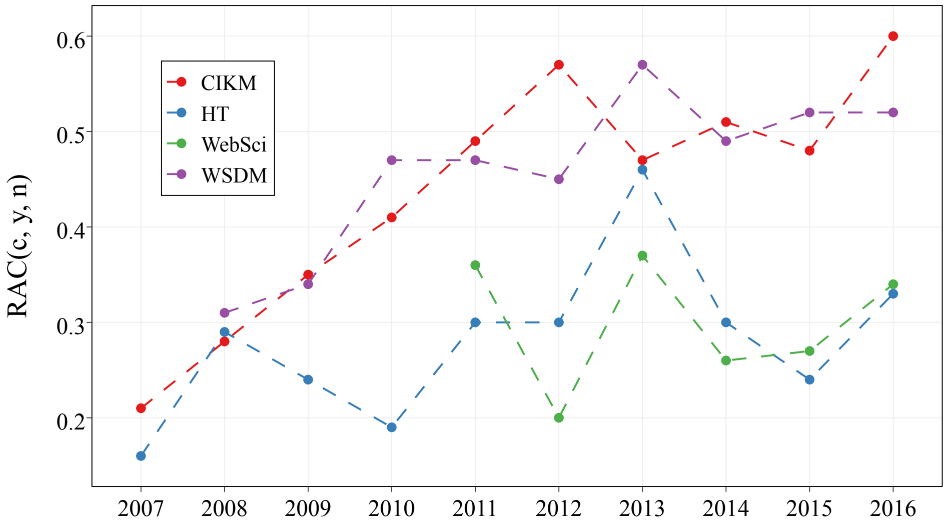

Fig. 10. **Inbreeding Metrics Computation:** Ratio of accepted papers for conference c in year y co-authored by programme committee members who served at least once during years y-n to y

6. INBREEDING

6.1 RAC(c, y, n) [PCaccProp]

RAC(c, y, n) [PCaccProp] denotes ratio of accepted papers for conference c in year y coauthored by PC members who served at least once during years y-n to y. Inbreeding is an indicator of the fraction of papers co-authored by PC members. Figure 10 shows the

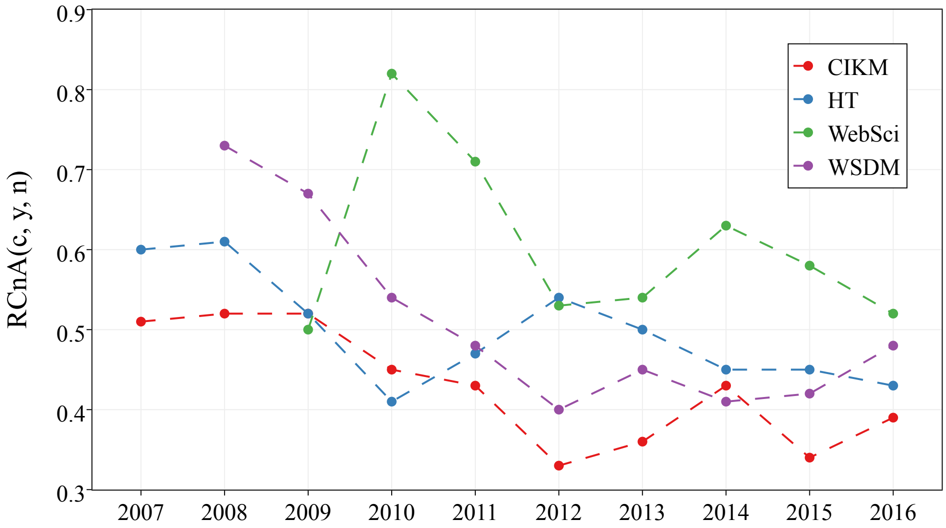

Fig. 11. **Representativeness Metrics Computation:** Ratio of PC members for conference c in year y that never have co-authored a paper at preceding instances of c between y-n and y-1

trends on inbreeding metrics for the four conferences. We observe that there are several years for which the inbreeding value is more than 0.5 for CIKM (2012, 2014, 2016). Figure 10 reveals a decline in the RAC(CIKM, y, 4) [PCaccProp] value after 2012 but then it increased again in 2016 to 0.6. The inbreeding value for HT varies from a minimum of 0.16 (2007) to a maximum of 0.46 (2013). Figure 10 reveals that there is only one year (2013) in which the RAC(HT, y, 4) [PCaccProp] is more than 0.4.

Figure 10 shows both an upward and downward trend in RAC(WEBSCI, y, 4) [PCaccProp]. We observe that the biggest jump in the inbreeding value for WEBSCI is from 0.2 to 0.37 (year 2012 to 2013). The RAC(WEBSCI, 2016, 4) value is 0.34. Figure 10 reveals that RAC(WSDM, y, 4) is stable and in the range of 0.45 to 0.57 from the year 2010 to 2016. The biggest decline in the RAC(WEBSCI, y, 4) value is from the year 2013 to 2014.

7. REPRESENTATIVENESS

7.1 RCnA(c, y, n)

RCnA(c, y, n) denotes the ratio of PC members for conference c in year y that never have co-authored a paper at preceding instances of c between y-n and y-1. Figure 11 displays the trends for the representativeness metrics for the four conferences. Vasilescu et al. define the representative metrics based on the belief that PC members should be representative of their respective communities, i.e., they should largely be established authors within those communities. However, not all PC members should be expected to have published at a conference before [Vasilescu et al. 2014]. We observe that the RCnA(CIKM, y, 4) value for CIKM varies from a minimum of 0.33 to a maximum of 0.52. The representativeness of CIKM is relatively higher in the initial years (about 0.5 from 2007 to 2009) in comparison to 2015 and 2016 (about 0.34 to 0.39). Figure 11 reveals that the representativeness of

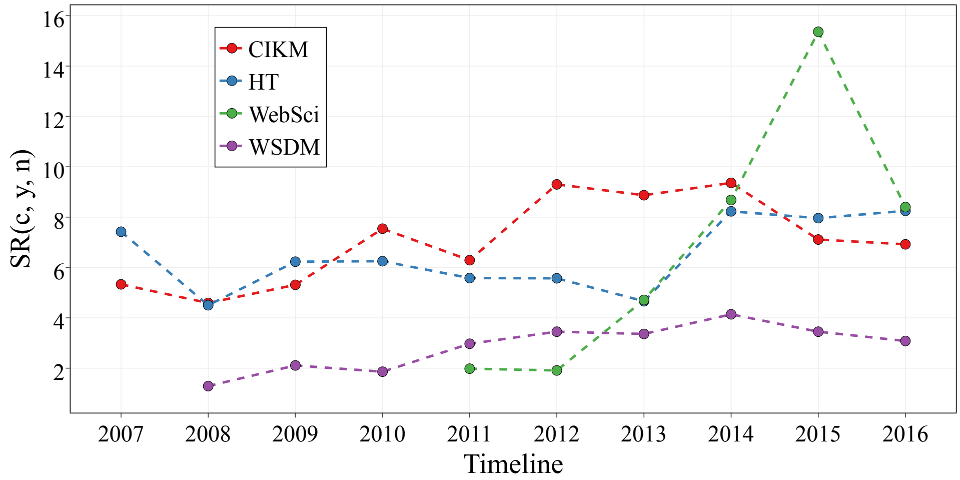

Fig. 12. **Sustainability Metrics:** Ratio between the number of core authors that have not served on the PC in years y-n to y and #C(c,y).

WEBSCI is consistent for the three most recent years (0.43 to 0.45 from 2014 to 2016).

In comparison to CIKM and HT, the variance in representativeness for WEBSCI is higher. We observe that the RCnA(WEBSCI, y, 4) value varies from a minimum of 0.5 to 0.82. Figure 11 reveals a declining trend in representativeness from year 2014 to 2016. We observe that the RCnA(WSDM, y, 4) is stable and does not vary much for the past six years of 2011 to 2016 (0.4 to 0.48).

8. SUSTAINABILITY

8.1 SR(c, y, n)

SR(c, y, n) denotes the ratio between the number of core authors that have not served on the PC in years y-n to y and #C(c, y). Vasilescu et al. define core author for a given conference as an author who frequently (co)authored papers published at that conference during the current or previous four editions [Vasilescu et al. 2014]. Figure 12 shows trends on the sustainability metrics for the four conferences. We observe that the SR(CIKM, y, 4) value for CIKM varies from a minimum of 4.59 to a maximum of 9.36. Figure 12 reveals a declining trend in the sustainability metric for CIKM from the year 2014 to 2016. Our analysis shows that the SR(HT, y, 4) value for HT varies from a minimum of 4.5 in 2008 to a maximum of 8.25 in 2016. Figure 12 reveals that the biggest jump in SR(HT, y, 4) is from the year 2013 to 2014 when the metric grows from 4.66 to 8.23.

In comparison to CIKM and HT, WEBSCI shows a much wider variance in terms of the SR(WEBSCI, y, 4) value. We observe that the SR(WEBSCI, y, 4) value varies from a minimum of 1.91 to a maximum of 15.36. Figure 12 reveals that in the initial years (2011 and 2012), the sustainability metric value was low and then it started increasing. The SR(WEBSCI, 2016, 4) value is 8.4. WSDM has the least variation with respect to the SR(WSDM, y, 4) metric. Figure 12 reveals that the SR(WSDM, y, 4) value varies from a

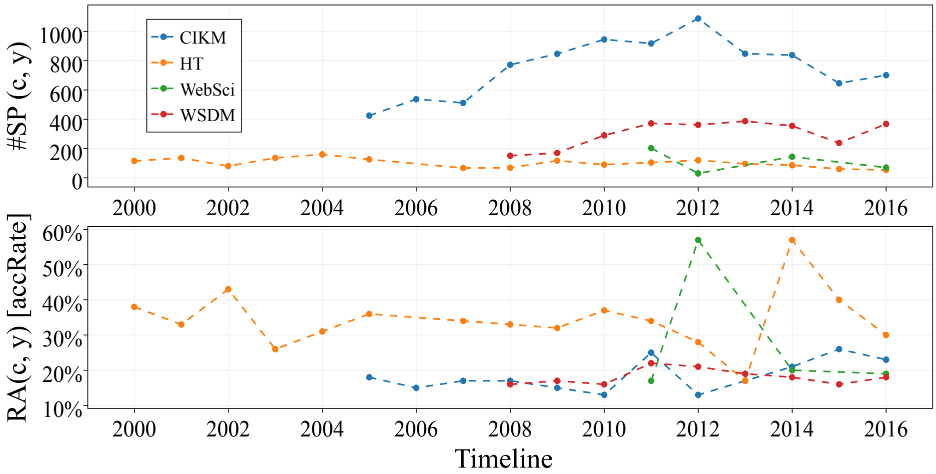

Fig. 13. **Prestige Metrics:** Number of Submitted Papers and Ratio of accepted papers for conference c in year y

	CIKM	HT	WebSci	WSDM
h5-index	42	24	25	58

Table I. **Prestige Metrics:** Conference Impact of conference c = Google Scholar h5-index for c between 2011 and 2015.

minimum of 1.29 to a maximum of 4.14. Figure 12 shows that there is a declining trend in SR(WSDM, y, 4) value from 2014 to 2016.

9. PRESTIGE

9.1 #$SP(c, y)$ [#$subm$]

Figure 13 displays the number of submitted papers for conference c in year y (#$SP(c, y)$ [#$subm$]) and the ratio of accepted papers for conference c in year y (RA(c, y) [accRate]). #SP(CIKM, y) denotes the number of submitted papers for conference c in year y. The minimum #SP(CIKM, y) value is 425 for the year 2005 and the maximum value is 1088 for the year 2012. The #SP(CIKM, y) value is more than 900 for the year 2010 to 2012 and then declines from the year 2013 to 2015.

Figure 13 reveals that the minimum number of submitted papers for HT is 54 (2016) and the maximum is 160 (2004). #SP(HT, y) value is more than 100 for the years 2000, 2001, 2003, 2004, 2005, 2009, 2011 and 2012. The #SP(WEBSCI, y) value varies from a minimum of 30 (2012) to a maximum of 203 (2011). The #SP(WSDM, y) value varies from a minimum of 151 (2008) to a maximum of 387 (2013). Figure 13 reveals that the #SP(WSDM, y) value steadily increases from 2008 to 2011 and then declines for one year (2012) and then again increases in 2013. The #SP(WSDM, y) value is more than 300 from 2011 to 2016.

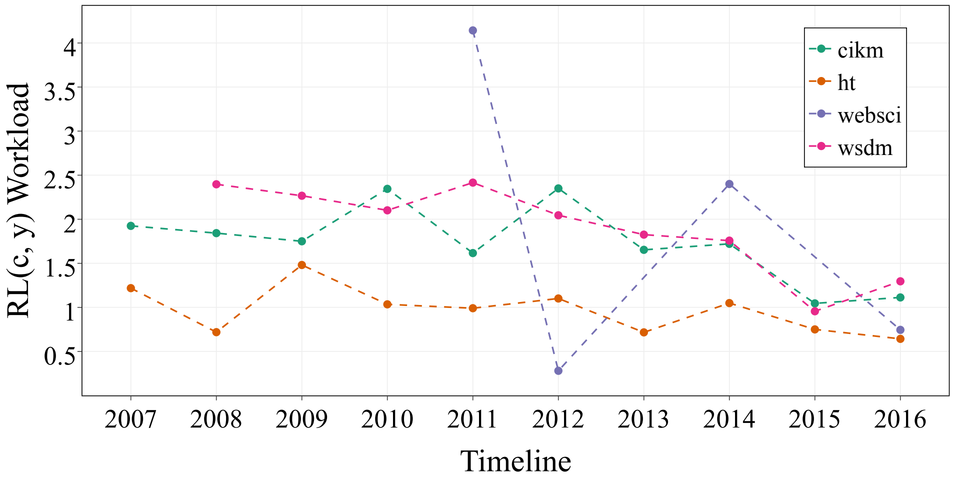

Fig. 14. **Workload Metrics:** Review Load for conference c in year y. RL(c,y) = #SP (c,y)/#C(c,y)

9.2 RA(c, y) [accRate]

RA(c, y) [accRate] denotes the ratio of accepted papers for conference c in year y. Figure 13 reveals that the RA(CIKM, y) [accRate] value for CIKM varies from a minimum of 13% to a maximum of 26%. In 2012, CIKM received 1088 papers and only 146 (13%) papers were accepted. The acceptance rate of HT is higher than CIKM. We observe that the RA(HT, y) [accRate] value for HT varies from a minimum of 17% in the year 2013 to a maximum of 57% in the year 2014. Figure 13 shows that the RA(WEBSCI, 2014) [accRate] value is 20% and the RA(WEBSCI, 2016) [accRate] value is 19%. We observe that WSDM is a highly selective conference. The acceptance rate of WSDM varies from a minimum of 16% to a maximum of 22%. Figure 13 reveals that from year 2011 to 2016, the number of papers submitted are between 230 and 390 and the acceptance rate is consistently around 20%.

9.3 CI(c)

CI(c) denotes the conference Impact of conference c which is equal to the SHINE h-index for c between 2000 and 2012. We do not have the data available for SHINE h-index and hence we compute the Google Scholar h5-index for all the four conferences between 2011 and 2015. Table I shows the h5-index of selected conferences fetched from Google Scholar Metrics. The h-5 index for the four conferences are: CIKM = 42, HT = 24, WEBSCI = 25 and WSDM = 58. CIKM is the largest conference and has a good Google Scholar h5-index but WSDM being a relatively smaller conference than CIKM has a much higher Google Scholar h5-index than CIKM.

10. WORKLOAD

10.1 RL(c, y) [revCoeff]

RL(c, y) [revCoeff] denotes review Load for conference c in year y. RL(c, y) = #SP(c, y)/#C(c, y). Figure 14 displays the trend of review load for conference c in year y. Workload is an indicator of the review load of the program committee members in terms of number of papers submitted and number of program committee members available to review the submitted papers. Each submitted papers gets reviewed by three program committee members (except under very rare cases where there can be two reviewers per paper). The workload metrics is based on the number of papers submitted and not the number of reviews required per paper. We observe that the RL(CIKM, y) [revCoeff] value varies from a minim of 1.05 to a maximum of 2.35. Figure 14 reveals that for the year 2014, the review load for CIKM was 1.72 and then it declined to 1.05 and 1.11 in the year 2015 and 2016 respectively.

Our analysis shows that the review load for HT varies from a minimum of 0.64 to a maximum of 1.48. Figure 14 reveals that there is a declining trend in the review load for HT from 2014 to 2016. We observe a wide variation in the review load for WSDM. Our analysis shows that the workload for WEBSCI varies from a minimum of 0.28 to a maximum of 4.14. The RL(WEBSCI, 2016) [revCoeff] value is 0.74. We observe that the workload for WSDM varies from a minimum of 0.96 in the year 2015 to a maximum of 2.42 in the year 2011. The RL(WSDM, 2015) [revCoeff] and RL(WSDM, 2016) [revCoeff] value is 0.96 and 1.30 respectively which is lower in comparison to the year 2007 to 2012 which is around 2.0.

11. CONCLUSION

The number of distinct authors for CIKM has shown the maximum increase in comparison to HT, WEBSCI and WSDM. WEBSCI shows a consistently increasing trend since its beginning until the year 2016. In comparison to the four conferences, HT has the minimum growth in terms of the number of distinct authors. Our analysis reveals that all the four conferences are healthy from the perspective of author turnover which is an indicator of conference stability. We observe that similar to the number of distinct authors, the number of new authors for WEBSCI is also monotonically increasing. We observe that even if there is a decline in the number of new authors for a conference for a particular year, the trend does not continue and the author turnover increases from the subsequent editions itself which is a positive indicator of stability. We observe a high variation in the number of PC members for large conferences like CIKM. The number of PC members for HT are about 20% - 25% of CIKM. WEBSCI initially started as a small conference in which the number of PC members ranged from 14 to 49 from the year 2009 to 2011 and then gradually grew to a mid-size conference.

From the stability metrics we observe that all the four conference are able to attract a substantial number of new authors with respect to previous four editions. Stability metrics also shows a healthy author and PC turnover. Our analysis of openness metrics shows that there are a good number of papers of papers in SIGWEB conferences by new authors for which none of the co-authors has published at the conference in the past four editions. We observe a consistent increase in the openness metric of WEBSCI in comparison to the other three conferences. We do observe both increasing and declining trend in the openness metric for the four conferences and observe that mid-size conference HT has a wide variance in comparison to large conference like CIKM. CIKM shows a high inbreeding for several years. The amount of inbreeding in WEBSCI is relatively less in comparison to the other conferences. We observe that there are several PC members who are established and core authors in the SIGWEB conferences in our dataset and also find several PC members who have not published in the last four edition of the conference. In comparison to CIKM and HT, the variance in representativeness for WEBSCI is higher. We observe that there are several core authors who frequently (co)author papers and publish at that conference. We observe a declining trend in the sustainability metric for CIKM in recent years from the year 2014 to 2016. In comparison to CIKM and HT, WEBSCI shows a much wider variance in terms of the sustainability metric value. CIKM receives highest number of papers in comparison to other conferences. The acceptance rate of HT is higher than CIKM. The h-5 index for the four conferences are: CIKM = 42, HT = 24, WEBSCI = 25 and WSDM = 58.

REFERENCES

AGARWAL, S., MITTAL, N., KATYAL, R., SUREKA, A., AND CORREA, D. 2016. Women in computer science research: What is the bibliography data telling us? *SIGCAS Comput. Soc. 46,* 1 (Mar.), 7–19.

AGARWAL, S., MITTAL, N., AND SUREKA, A. 2016a. A glance at seven acm sigweb series of conferences. *SIGWEB Newsl.* Summer (July), 5:1–5:10.

AGARWAL, S., MITTAL, N., AND SUREKA, A. 2016b. A scientometric analysis of 9 acm sigweb cooperating conferences. *SIGWEB Newsl.* Autumn (Nov.), 6:1–6:15.

BARTNECK, C. 2010. The end of the beginning: a reflection on the first five years of the hri conference. *Scientometrics 86,* 2, 487–504.

BERGAMASCHI, R. A., REZENDE, R. C., DE OLIVEIRA, H. P., AND KUMON, JR., A. 2012. A quantitative analysis of www, hypertext and jcdl conferences in the last decade. *SIGWEB Newsl.* Winter (Jan.), 5:1–5:8.

ELSHAWI, R. AND SAKR, S. 2016. International conferences on computer system: Analysis of eurosys, sosp, and osdi during 2006-2014. *Collnet Journal of Scientometrics and Information Management 10,* 1, 175–195.

KUMAR, L., SRIPADA, S., AND SUREKA, A. 2016. A bibliometric study of asia pacific software engineering conference from 2010 to 2015. *CoRR abs/1610.09635*.

SAKR, S. AND ALOMARI, M. 2012. A decade of database conferences: a look inside the program committees. *Scientometrics 91,* 1, 173–184.

SHARMA, R., AGGARWAL, P., AND SUREKA, A. 2016. Insights from mining eleven years of scholarly paper publications in requirements engineering (re) series of conferences. *SIGSOFT Softw. Eng. Notes 41,* 2 (May), 1–6.

SHARMA, R. AND SUREKA, A. 2016. A nine year story of the india software engineering conference from 2008 to 2016. *SIGSOFT Softw. Eng. Notes 41,* 5 (Nov.), 31–44.

VASILESCU, B., SEREBRENIK, A., MENS, T., VAN DEN BRAND, M. G., AND PEK, E. 2014. How healthy are software engineering conferences? *Science of Computer Programming 89*, 251–272.

Swati Agarwal is a senior PhD candidate in Information Management and Data Analytics group at IIIT-Delhi. Her research interests include Social Computing, Social Informatics, Text Mining and Analytics, Natural Language Processing and Information Visualization. Link to Swati's homepage: `www.swati-agarwal.in`

Ashish Sureka is currently a Principal Scientist at ABB Corporate Research Center (India) and an Adjunct Faculty Member at IIIT-Delhi. His research interests are in the area of Software Analytics, Social Media Analytics, Mining Software Repositories and Process Mining.

Nitish Mittal is a software engineering and a graduate from Computer Engineering Department at NSIT, Delhi. His research interests are in the area of Data Mining, Big Data Analytics, Computational Social Networks and Information Visualization.