Expanding Scientific Community Reach Based on Web Access Data

Vagner Figueredo de Santana[(⊠)] and Leandro Marega Ferreira Otani

IBM Research, IBM, Tutóia St., 1157 São Paulo, SP, Brazil
{vagsant,lmarega}@br.ibm.com

Abstract. Knowing the main characteristics of a scientific community, how it reaches all stakeholders, and understanding how individuals engage around a subject is needed in order to support decision makers to plan strategies to maintain and nurture the community. This work presents a new way of interpreting the reach of a scientific community by incorporating Web access data to the co-author network commonly considered. The case presented involves the Brazilian Human-Computer Interaction (HCI) community and the access to the website of the XV Brazilian Symposium on Human Factors in Computer Systems, the main HCI conference in Brazil. The proposed method is grounded on Organizational Semiotics and differs from the existing works because it considers a wider population than the conference authors. Inspired by the Organizational Onion, it considers three different levels of connection: Informal, Formal, and Technical. In the presented case, the reach commonly used (i.e., author-author network) counts on 257 authors while the total of people orbiting the event involved 5,432 unique visitors, in other words, the co-author network represents approximately 5% of the population orbiting the event. The presented method shows that data originated from Web accesses support a different way of representing a scientific community reach, including multiple segments that are commonly not considered as the target-audience, resulting in a more inclusive approach in the sense of considering the plurality of people orbiting an event, mediated or not by a computer. Our contribution shows a data informed approach of expanding the scientific community reach in order to characterize people orbiting the conference.

Keywords: Organizational semiotics · Log analysis · Social network · Author network

1 Introduction

Bibliometrics is a discipline that involves multiple methods to describe characteristics and identify patterns from published documents. Norton [11] defines Bibliometrics as the set of the various methods of measurement applied to artifacts of human communication forms. In Computer Science, much of the work performed in Bibliometrics involves the use of DBLP Computer Science Bibliography[1], which is an online reference for bibliographic information on major Computer Science publications. In addition, works

---

[1] http://dblp.uni-trier.de/.

involving the analysis of DBLP, including analyzes of community and relationship among people, uses co-author (or author-author) network. In these works, authors are connected when authored a paper, book, chapter, among other types of documents. Examples of works based on DBLP performing such analysis are [3, 10, 16].

In the literature, when the term 'scientific community' takes place, either it is used broadly as anyone doing research or, in more data-oriented approaches, it is restricted to the ones publishing in certain conferences or journals. According to Barbosa and Souza [2], Human-Computer Interaction (HCI) researchers themselves should have a deep understanding of the context in which the object of their investigation is placed. Thus, we are proposing an extension to the co-author network to represent the people orbiting a given scientific subject (e.g., conference, discipline), highlighting characteristics often not covered in co-author network analysis.

In this work, we present a new way of interpreting the reach of a scientific community by incorporating Web access data. The Web access considered was captured using Google Analytics, which is one of the many analytics existing tools, and people that liked the conference page in the Facebook[2]. The case presented involves the characterization of the Brazilian Human-Computer Interaction community based on the access to the website of the XV Brazilian Symposium on Human Factors in Computer Systems (aka IHC), the main HCI conference in Brazil.

In the last years, the conference counted on more than 200 attendees. Since it very beginning it is promoted by Brazilian Computer Society. Since 2006 it has been organized in cooperation with SIGCHI (Special Interest Group in Computer Human Interaction) and ACM (Association of Computer Machinery)[3]. The most frequent keywords of papers published in the Symposium are the following: (1) Semiotic Engineering, (2) Human-Computer Interaction, (3) Accessibility, (4) Usability Evaluation, and (5) Interface [5]. They represent the focus of the Brazilian HCI Community and how researchers are engaged to HCI topics and methods.

The main contribution of this work is a data informed approach of expanding the scientific community reach in order to represent from co-author network to people orbiting the conference. The next section presents the related works, Sect. 3 describes the theoretical background, Sect. 4 details the proposed method, Sect. 5 presents the results and, finally, Sect. 6 concludes.

2 Related Work

Next we present works related to the characterization of the Brazilian HCI community, focusing on Bibliometrics involving the IHC conference.

Henry et al. [8] present an analysis of the main HCI conferences (i.e., ACM SIGCHI Conference on Human Factors in Computing Systems, User Interface Software and Technology, Advanced Visual Interfaces and IEEE Symposium on Information Visualization). The work also has a focus on visualization of information related to

---

[2] www.facebook.com/2016IHC/.

[3] www.sbc.org.br/ihc.

published works. This section will focus on presenting details of Brazilian HCI community. For a broader review of works involving the ACM SIGCHI conference and the international HCI community, refer to [1].

Barbosa et al. [1] explored the metadata of **340 full papers** published in the first **14 editions** of IHC. The work details the authorship profile of the Brazilian HCI community and how it changed over time, including the evolution of co-author network and characterization of the community.

Gasparini et al. [4] present a descriptive analysis of the first **11 editions** of IHC, focusing on the visualization of all editions of the event by the time the work was published. The data set analyzed involved **236 full papers** published and **398 authors**. The characterization had shown authors that play a central role in the community, main research themes, and geographical distribution of authors, among other information.

Considering co-author network studies, Gasparini et al. [5] analyzed the co-author network regarding **12 editions** of IHC. The data set considered as composed of conference **full papers**. In the study authors present the researchers that are central do the co-author network, connection trends, institutions, and acting fields of IHC authors (e.g. computer science, design, engineering, psychology, etc.).

Beyond the co-author network, Gasparini et al. [6] listed the **authors with more than one full paper** in the IHC, inspected the Lattes Curriculum Vitae[4] of each one of **105 authors**, gathering data about institution, advisor, and title. In addition, authors applied an online questionnaire to the ones that have changed institution, state, or region (**36 authors**). The questionnaire counted on **21 participants**. The work shows that migration occurred changed the collaborations among authors, fostering emerging HCI research groups and diffusing HCI in Brazil.

Considering citations of works published in IHC editions, Gasparini et al. [7] studied whether and how the IHC publications cite publications from IHC itself. The data set considered in the work involves **340 full papers** published in the first **14 editions** of the IHC, summing up to **7,350 authors**. The work describes a citation profile of the event and point to the growth of the community. However, it also presents that research produced by Brazilian HCI community should widen its visibility by its peers.

The existing works consider a restrict scope of the HCI Brazilian community when relying only on full papers or full paper authors information. Hence, people that act in the subject but are not present in the conference proceedings, participating, presenting, in multiple levels of engagement (e.g., demos, posters, and workshops), were not considered in previous analysis. It is also worth mentioning that these "soft connections" regarding people orbiting the conference were hard or even impossible to obtain in early editions of IHC. However, in this paper we present an approach that combines additional data to the co-author network to better characterize the reach of the conference.

According to Souza et al. [13], due to the large size of Brazil, IHC has been the main venue for HCI researchers to meet and discuss their works. However, for the same reason, people that orbit the IHC and are impacted by decisions taken considering only full paper authors. The proposed method differs from the existing ones because it

---

[4] http://lattes.cnpq.br/.

considers a wider population than the full paper authors, which is a group of people commonly considered in reports and studies to characterize scientific community or to build co-author networks.

3   Theoretical Background

Following the tradition of HCI in Brazil of being inspired and expanding the boundaries of Semiotics [13], this work is grounded on Organizational Semiotics (OS). OS is a discipline that deals with information and information systems in such a way it takes into account from technical to human and social aspects. OS supports the understanding on how individuals behave, norms governing behaviors, characteristics and functions of signs used [8, 14, 15]. According to Liu [9], "*an organization is a social system in with people behave in an organized manner by conforming to a certain system of norms.*" Moreover, an organization can be seen as an information system where people use signs with some purpose [9]. In this sense, organizations count on regular norms and processes that can be automated and supported by computer-based systems. However, the introduction of computer-based systems is just part of the solution, since informal and formal layers encompass this technical aspect. In OS, this view of organization is referred as the Organization Onion (Fig. 1).

The informal layer takes into account cultural aspects, values, habits and behavior of each individual member of the organization. In the informal layer the intentions are understood and emerging protocols and patterns are taken into account. The formal layer is within the context of informal layer. In the formal layer the literate culture plays the central role; rules are created to replace intentions. People involved in the formal layer function as sign-token transmitters in this organization. The technical layer is
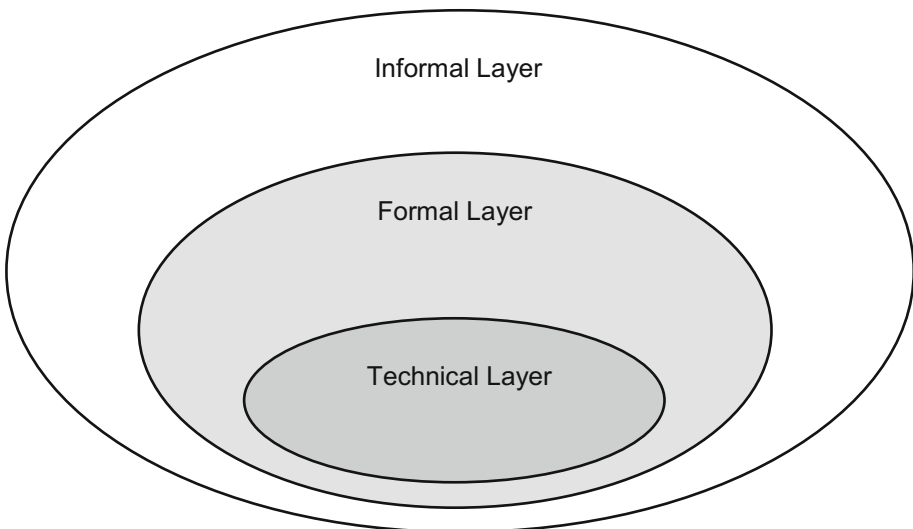
Fig. 1. Organizational Onion.

within the context of formal layer. The technical layer encompasses the system programmed according to rules and processes of the said organization. In the technical layer routine and repetitive tasks of well-defined work processes are coded. In sum, a technical system presupposes a formal system, which, in turn, relies on an informal system [9].

The Organizational Onion provides a rich framework to understand organizations and, specially, to consider the multiple aspects related to information systems and **how human perform tasks in these systems**, from well-defined tasks, reaching rules, behavior and intentions emerging from interactions among humans.

The rationale of applying the Organizational Onion to study the reach of a scientific community resides in the fact that, by studying the connections between authors (co-author network), only the technical information is considered; rules and processes are well-defined to be there. The role of sign-token transmitters is not considered, i.e., communications emerging when people interact among each. In this regard, the formal layer is considered. Finally, traces of interactions that are not mapped are taken into account by Web access data analysis.

4 Proposed Method

The method places accesses of people orbiting the subject in the informal layer, people formally engaged in the subject are placed in the formal layer, and, finally, people that publish at the conference are placed in the technical layer. Accordingly, based on the Organizational Onion, we propose the placement of connections and people around a scientific subject according to the three layers, thus, representing the reach of a community: informal, formal, and technical (Fig. 2).

- **Informal:** the set of people with some interest in the subject orbiting the subject in an informal way, mainly by the use of Web access data and Online Social Networks (OSN) data.
- **Formal:** the set of people with interest in the event, mainly represented by people attending to the event, having a formal connection with the subject.
- **Technical:** the set of people that strongly engaged in the subject, represented by people that published a work at the conference (i.e., co-author network).

Considering the works in the literature regarding the co-author network, the people and relationships are mapped to a graph. This mapping represents people as nodes (or vertices) and edges (or links) connect these people. In a co-author network, each author is represented as node and two nodes are connected if they published a work together. Hence, the technical layer is mapped according to the co-author network.

The formal layer considers people that participated in the event without publishing papers, representing a different engagement and sign-token transmission. The informal layer is represented by people orbiting the event, with no formal connection with the event, mapped by considering Web access and OSN connections.

The data set considered combines access data to the conference website in conjunction with the data coming from conference proceedings and participation reports prepared by session chairs during the conference. When author name conflict (such as
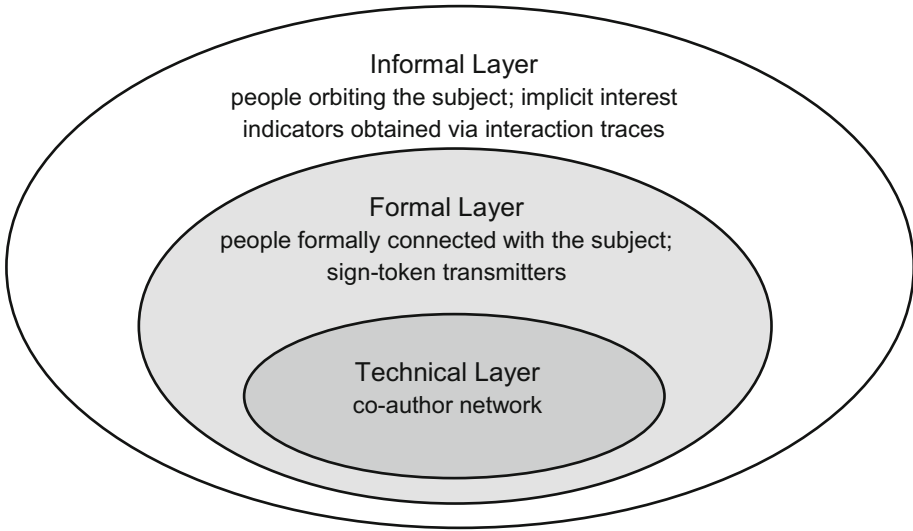
Fig. 2. Organizational Onion applied in the context of scientific community reach.

name similarity and/or abbreviation), last names were compared and, on of further conflicts, authors' CVs were checked. When conflicts persisted, the most recently mapped name was discarded.

For the analysis regarding co-author network, IHC and OSN participants, the data set consists of 276 Facebook users, 105 conference works, and a total amount of 120 authors that presented, 32 authors that did not present but participated in the event, 105 authors that were not present and 61 people that participated in the event without any publication in the conference. The works were distributed as follows: 34 full papers, 24 short papers, 2 workshops, 3 short courses, 6 works presented during the workshop of thesis and dissertation, 8 works presented during the workshop on HCI education works, 6 reports part of the evaluation competition, 4 works of the HCI in practice track, and 18 posters/demos.

The tool used was Google Analytics and the data set considered involves 10 months of Web access, from the publication/announcement of the website to December 21st. Google Analytics was chosen because it is the most popular analytics tool and it is being used by 54.5% of the websites in the whole Web [17].

In order to use the Google Analytics, it is required the insertion of a line of code in all Web pages of the website to analyze. This line of code refers to the JavaScript library responsible for logging Web data access. Google Analytics offers multiple features as history analysis, real time access view, demographic analysis, visitors interests, navigation flow, among others. However, in specific occasions it is not possible to capture data related to user profile due to privacy or technologic characteristics of users' device. Demographic data and topics of interest, for instance, are not identified by Google Analytics service, instead, Google partners as DoubleClick use mobile apps and cookies to obtain this data. In addition, the use of Google Analytics

was considered over other solutions based on Web server logs considering the short-comings presented in [12].

The data capture started from the moment the website was published aiming at constant evaluation of the website usage and characterization of the access. The analysis was postponed until the access started to decay, representing the decrease of interest a couple of weeks after the conference. Moreover, in the process of data cleaning, sessions with duration of 10 s or less where removed from the data set, as these accesses usually come from Web crawlers or developers debugging the website.

The data considered from OSN refers to the likes given by people to the event page on Facebook. Event participants were not included in the likes counting. The goal is to map also people interacting with the event only via OSN, given that people are using more and more such pages instead of navigating through the website (see the trend of searches regarding last editions of the event, even when the number of participants remains stable over the last IHCs[5]).

5 Results

Next we present the results from the characterization of people involved in the three layers, from informal, to formal, and then technical. Moreover, we discuss characteristics of different layers. The focus is on characterizing the informal layer and contrasting with basic characteristics of formal and technical layers.

5.1 Informal Layer

The Web data access considered in the analysis was captured during 10 months. It involves 25,519 sessions, 141,079 page views (mean number of pages per session 5.53), 10,904 unique visitors, 57.8% of returning visitors, and bounce rate of 12.02%. Figure 3 shows the number of daily sessions in the period versus mean session time. Before disclosing the website to the public, the website counts on low number of sessions and high duration time compared to the rest of the sessions, showing the activity of developers and the organizing committee reviewing the website content. Thus, to remove this effect from the data set, the period considered in the analysis is from February 22nd to October 18th. Hence, to select the proper data, the Google Analytics segment feature was used. An additional exclusion condition for sessions with duration of 10 s or less. The referred period is related to 11,461 sessions, 106,711 page views (mean of 9.31 pages per session), 5,432 users, mean visit time of 6 min 28 s and bounce rate of 0.01%. It is worth noting the reduction of the bounce rate from 12.02% (the whole data set) to 0.01% (after filtering), which highlights the role played by this segment feature. Thus, for the next analysis informing user details, the data set considered is related to this segment of the data set.

---

[5] https://www.google.com/trends/explore?date=all&q=ihc%202016,ihc%202015,ihc%202014,ihc%202013.

Fig. 3. Daily sessions; the peak is related to the period in which the conference occurred.

In Fig. 3 it is possible to verify that the number of accesses is reduced in the weekends and that the accesses increase in the first days of the week (Monday, Tuesday). On the other hand, the mean session time is greater in the weekends. In Fig. 4 it is possible to see that the mean session time decreases as time passes. This behavior is expected given that the website counts on 57.8% of returning visitors. Next, we present the characterization of accesses according to the following: demographics, topics of interest, geography, engagement and flow of users, technology, and device.

Regarding demographics and interests, the group of people from 25 to 34 years has the greater participation in this layer (Fig. 5). Also, it was computed that 59.4% of the accesses are related to male and 40.6% are related to female. The distribution considering age and number of sessions throughout time (Figs. 6 and 7) shows peaks related to important dates of the event and the predominance of people from 25 to 34 years old in terms of activity. In addition, the 35 to 44 years old group accessed the website mainly during week days. No differences in the access behavior considering gender. Visitors interests involve technology (4.8%), movies (4%), and TV (3.7%)
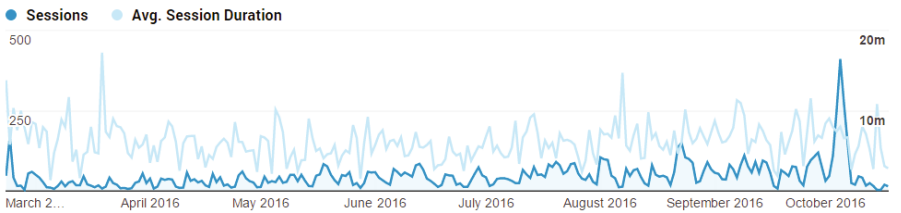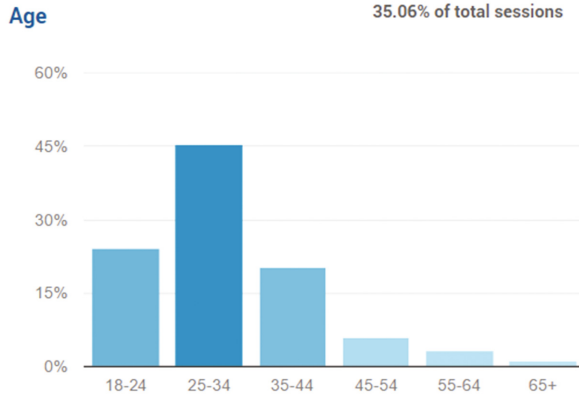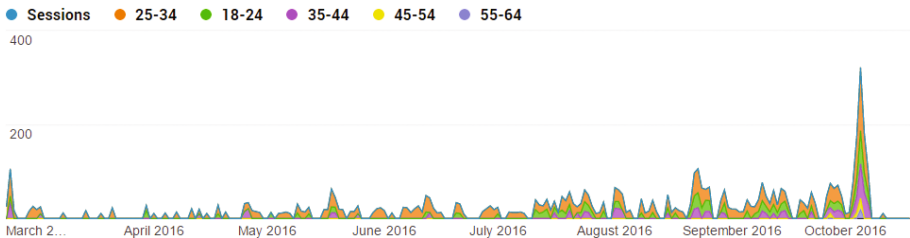

(a)
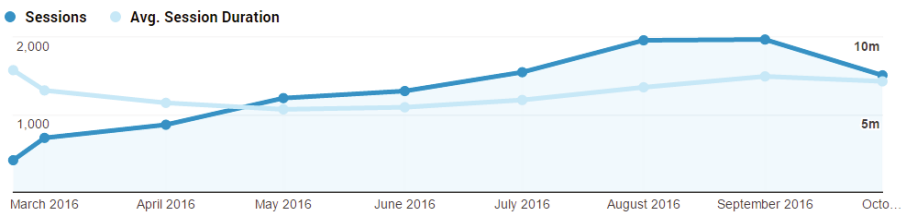
(b)

Fig. 4. Daily sessions versus mean time (a) and monthly sessions versus mean session time (b).
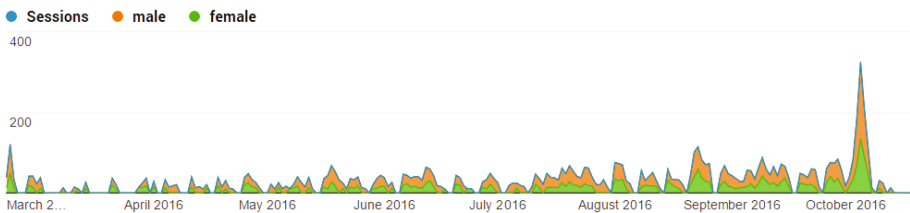
Fig. 5. Distribution of users by age.

Fig. 6. Distribution of sessions by age.

Fig. 7. Distribution of sessions by gender.

(Fig. 8). Considering market segments, education (5.15%), travel (3.47%), and consumer electronics (3.42%) are the top segments related to the website visitors (Fig. 9).

Considering geography, as expected, most of the access came from Brazil; 10,683 sessions (93.21%). In this group, 71.27% use Portuguese as default browser language and 20.68% use English as default browser language. Figure 10 shows the distribution of visits considering the Brazilian state that originated the accesses. From the 10,683 accesses containing information regarding the state of origin, 4,039 are attributed to the state of Sao Paulo (37.81%); this is expected since that Sao Paulo is the most populated
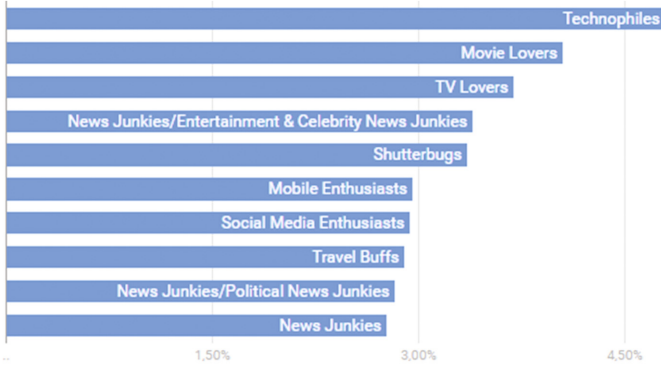
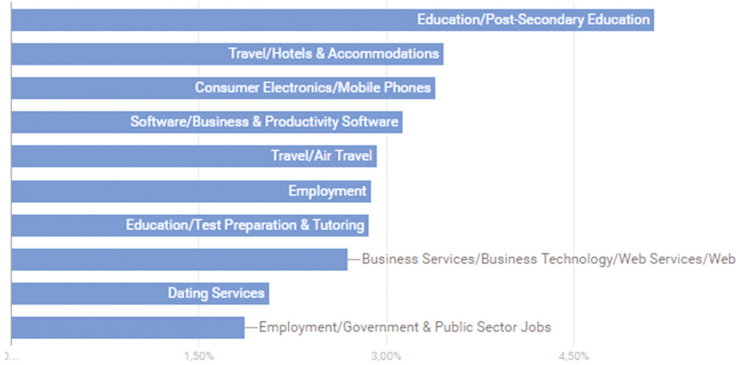**Fig. 8.** Distribution of visitors' affinities.

Fig. 9. Distribution of market segments associated to the visitors.

state in Brazil and hosted 2016's conference. Moreover, it is also possible to identify that Brazilian states that are hardly considered as part of the community, counted on accesses, for instance, Acre, Roraima, Rondonia, which originated (3, 5, and 6 visits, respectively). This highlights the importance of supporting HCI research in the North of Brazil.

Considering engagement, Fig. 11 shows the filtered sessions (0-10 s) and that the majority of visits are in the interval between 11 and 1800 s (30 min). In addition, the most common duration is in the interval of 1 to 3 min. Figure 12 shows the flow of users from the homepage of the conference's website. More than 6,200 accesses were originated at Google (56.36%). In addition, the pages that users reached most commonly from google were the homepage and the call for papers page. The next pages accessed involve important dates, venue, and program information. Figure 13 shows the origin of accesses and the following flow of users. Beyond accesses originated in Brazil, the website also counted on accesses coming from United States (410 visits), Canada (61 visits), Russia (46 visits), and United Kingdom (28 visits).
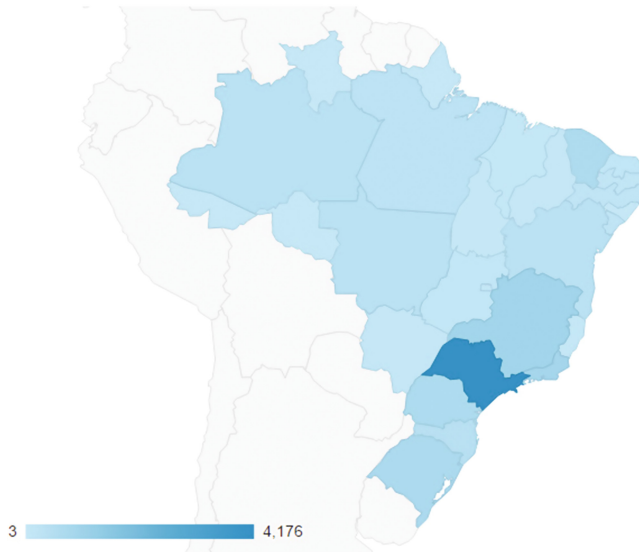
Fig. 10. Accesses by Brazilian states.

Session Duration	Sessions	Pageviews
0-10 seconds	0	0
11-30 seconds	2,430	11,837
31-60 seconds	1,793	10,698
61-180 seconds	2,618	20,356
181-600 seconds	2,359	23,416
601-1800 seconds	1,836	23,481
1801+ seconds	425	16,923

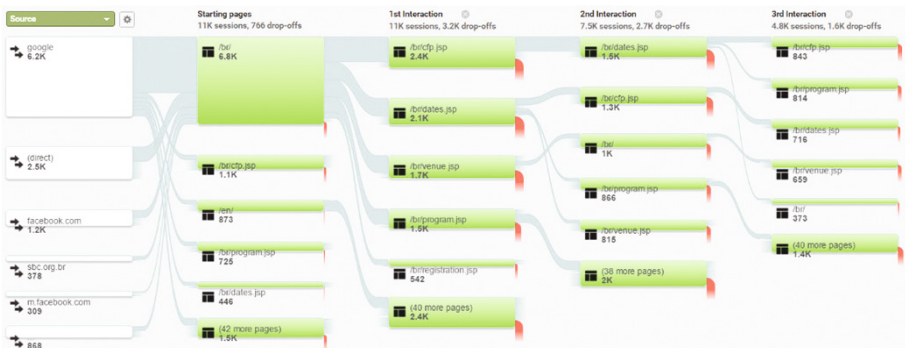**Fig. 11.** Duration of visits and visualizations of pages.

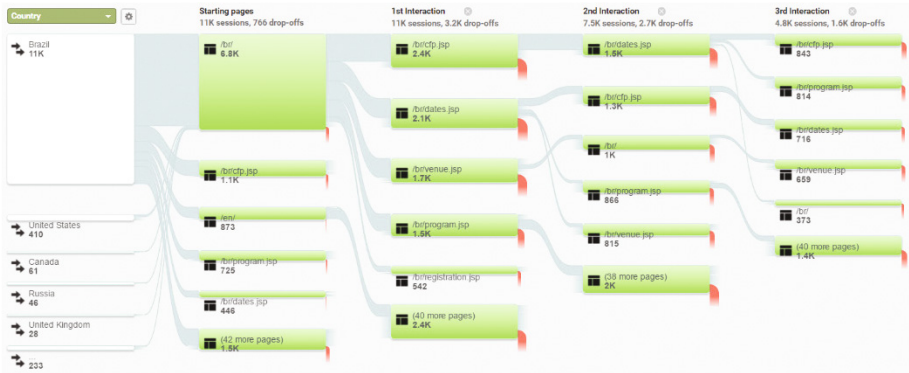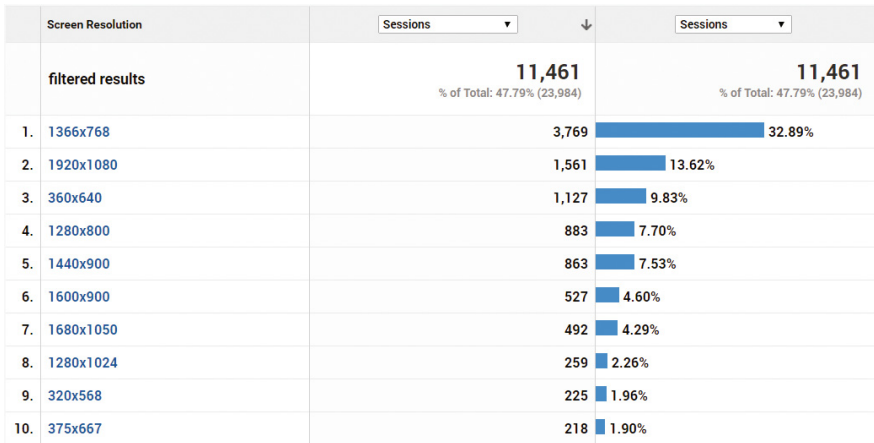**Fig. 12.** Flow of users across website pages.

Fig. 13. Country of origin of accesses and flow of users.

Bearing in mind the technology used, Google Chrome, Mozilla Firefox, and Safari are the most popular Web browsers (Fig. 14). The most used screen resolutions are 1366 × 768, 1920 × 180, and 360 × 640 (Fig. 15). This last resolution is an interesting result, probably related to the use of mobile devices (Fig. 16), which occurred in 17.55% of the visits; tablet accesses represent 1.88% and desktop accesses represent 80.57% of the visits.
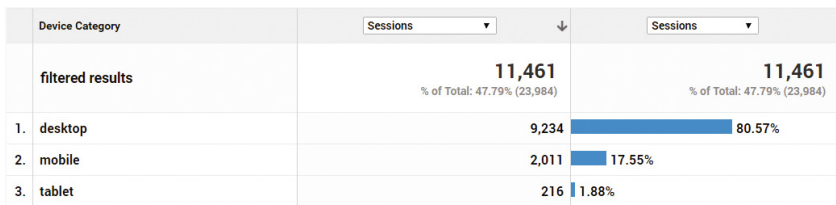
Figure 17 shows the view of the informal layer, omitting nodes related to unique visitors. In the image, it is possible to see how authors are positioned in the center, and how conference participants and people that followed information of the conference via Facebook distance from them. The graph has 6,026 nodes (i.e., the sum of authors, participants, unique visitors, and Facebook users), 382 edges, density of 2.10×10^{-05}, and 5831 connected components.

Fig. 14. Web browsers used to access the conference website.

Screen Resolution	Sessions ▼	↓	Sessions ▼
filtered results	**11,461** % of Total: 47.79% (23,984)		**11,461** % of Total: 47.79% (23,984)
1. 1366x768	3,769		32.89%
2. 1920x1080	1,561		13.62%
3. 360x640	1,127		9.83%
4. 1280x800	883		7.70%
5. 1440x900	863		7.53%
6. 1600x900	527		4.60%
7. 1680x1050	492		4.29%
8. 1280x1024	259		2.26%
9. 320x568	225		1.96%
10. 375x667	218		1.90%

Fig. 15. Screen resolutions used by visitors.

Device Category	Sessions ▼	↓	Sessions ▼
filtered results	**11,461** % of Total: 47.79% (23,984)		**11,461** % of Total: 47.79% (23,984)
1. desktop	9,234		80.57%
2. mobile	2,011		17.55%
3. tablet	216		1.88%

Fig. 16. Devices used by users to access the conference website.

Informal Layer
6026 nodes

Formal Layer
594 nodes

Technical Layer
257 nodes
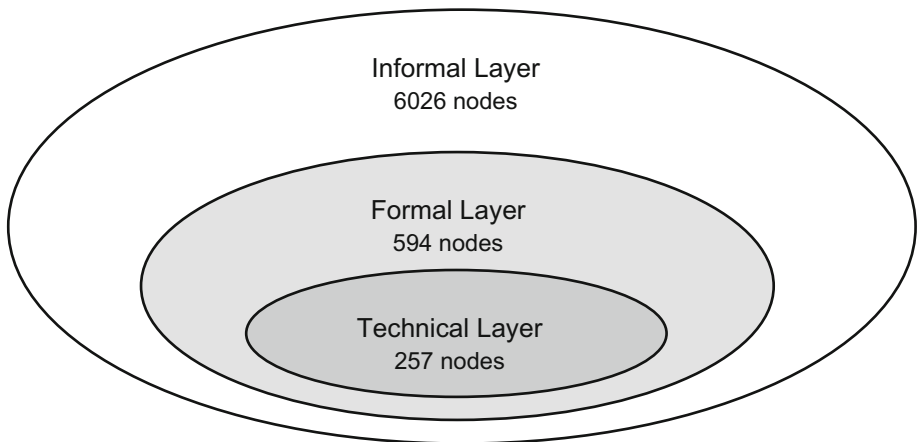
Fig. 17. Representation of the universe related to the IHC conference. Nodes represent authors, participants, authors that presented during the conference, and people that liked the conference page in the Facebook

5.2   Formal Layer

Figure 18 shows the view of the formal layer. In the image, it is possible to see how that 61 participants (10.27% of the network) orbit the co-author network, reinforcing the proposed approach. The graph has 594 nodes, 382 edges, density of 0.005, 399 components, and average degree of 1.286.

5.3   Technical Layer

Figure 19 shows the view of the technical layer. In the image, it is possible to see how authors of the IHC 2016 are connected. Note how few authors central connect with multiple authors. These central authors are professors that supervise multiple students
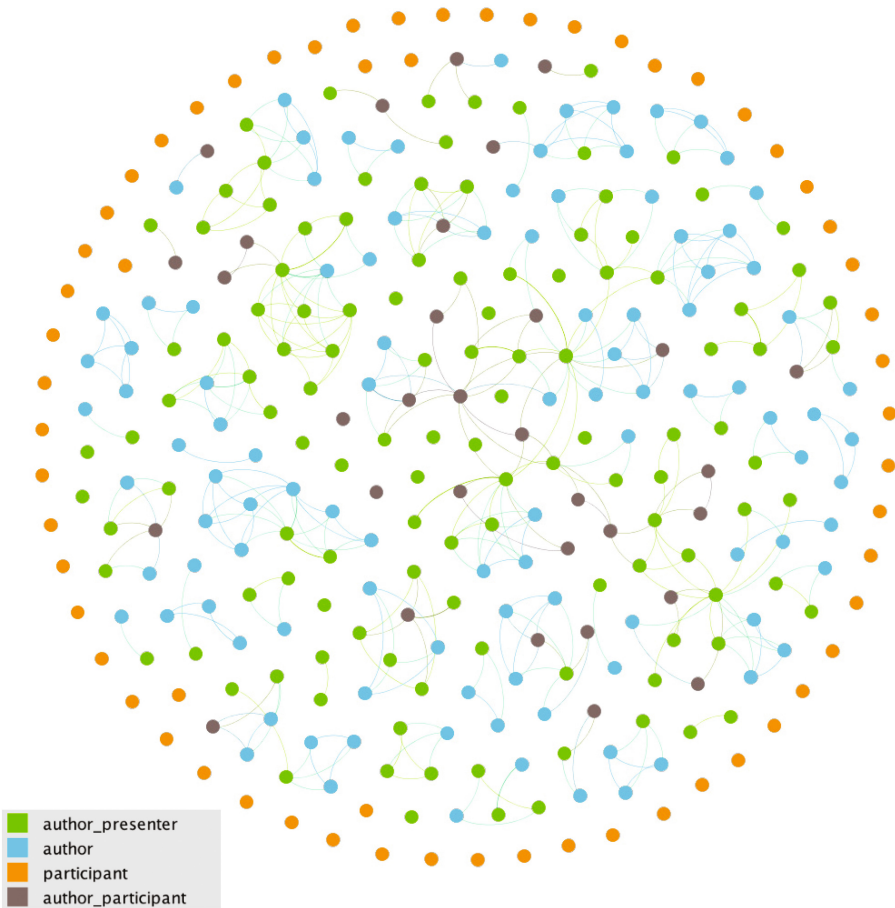
Fig. 18. Connections among authors, authors that participated in the conference, and the authors that presented during the conference.
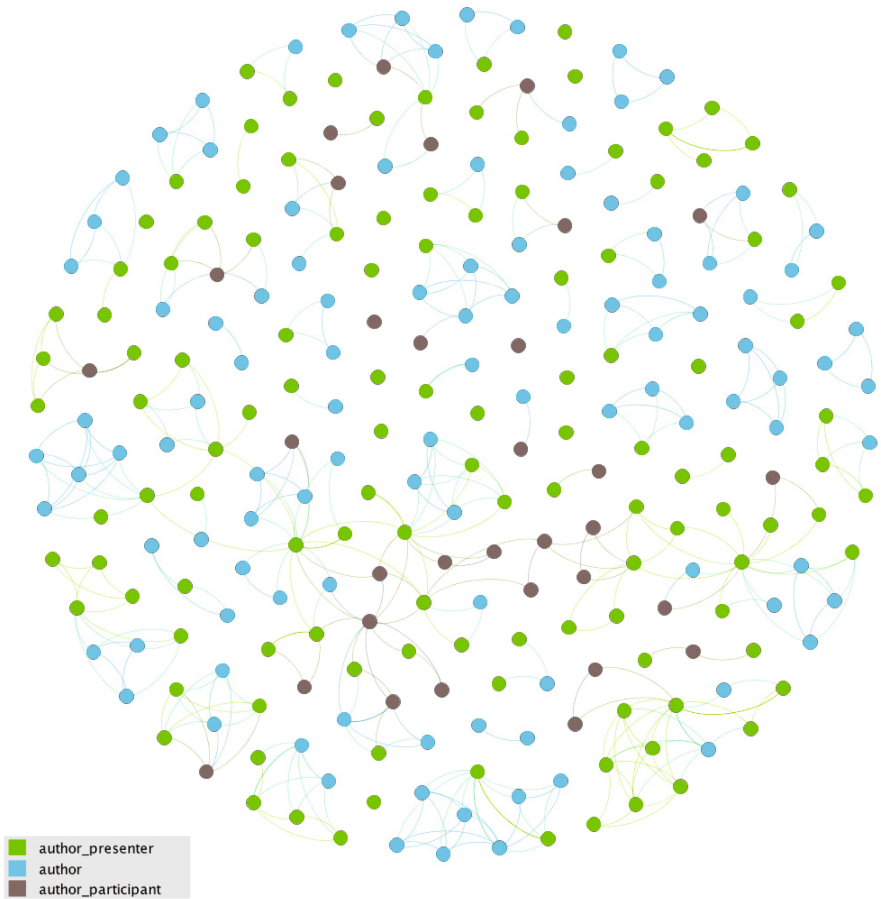
Fig. 19. Connections between authors, authors that participated in the conference, and the authors that presented during the conference.

in HCI. The graph has 257 nodes, 382 edges, density of 0.012, 62 connected components, and average degree of 2.973.

6   Discussion

The presented method shows that data originated from Web accesses support a different way of characterizing a community, considering people that are interested in the subject and that are touched somehow by the conference and actions taken having in mind only the authors. In sum, the method places accesses of people orbiting the subject in the Informal Layer, people formally engaged in the subject are placed in the Formal Layer, and, finally, people that are present in the conference proceedings are placed in the Technical Layer. The results show that a scientific community touches more people
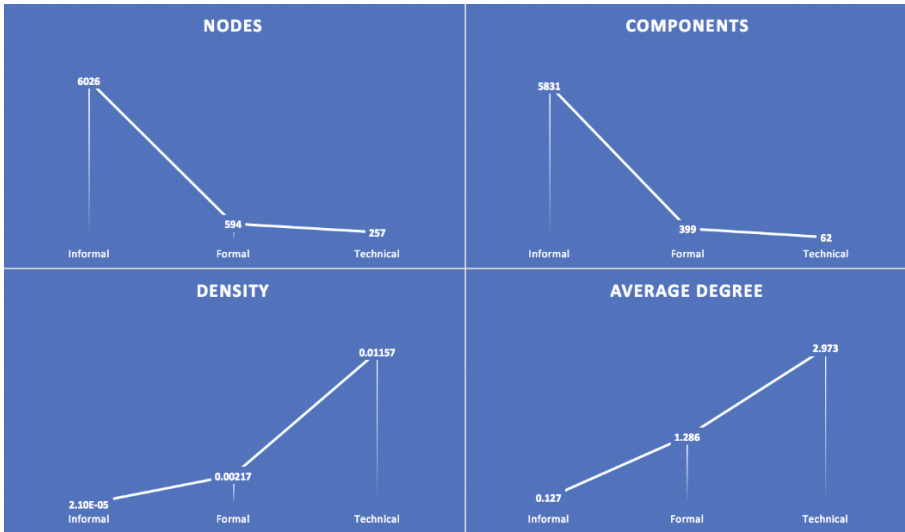
Fig. 20. Comparison between informal, formal, and technical layers regarding number of nodes in the network, connected components, network density, and node average degree.

than only its authors. In the presented case, the Technical layer counts on few hundreds of authors (the set usually considered to characterize a scientific community) while the Informal layer involves more than 6,000 people. Moreover, novel characteristics were also highlighted, allowing a better characterization of this wider population involved in a scientific conference. Even in the data access was possible to identify different levels of engagement, from people getting in touch with the subject to people that research the subject for decades. Figure 20 shows how different metrics change when different levels are considered. People orbiting the conference represent a considerable population. In addition, given that the edges are only related to the conference of 2016, the number of edges do no alter, thus, the number of connected components represents the same situation as the graph nodes chart. When considering density, it is interesting to see how sparse such network is; the technical layer, which is the most connected one counts on 1% of all possible connections. Moreover, regarding average degree, the technical layer points that in the technical layer that each author of IHC 2016 co-authored a paper with 2.973 other authors.

The proposed approach has some limitations. The presented analysis was performed considering only one year of the conference. An interesting approach involves analyzing the history of a conference, especially if the conference website stays at the same domain, which enables a long-term analysis. It's also valid to notice that, since Google Analytics define new website visitors based on cookies, the differentiation of a new visitor from a returning visitor connecting from a different browser or device is not possible. However, due to technology (i.e., Javascript is enabled in almost all modern browsers) and adoption characteristics (i.e., requiring authentication), this is the best method for identifying visitors at scale.

This work has an especial value in the context of accessibility, given that it includes people that are away of the venue. Knowing characteristics of the whole audience and expanding the reach of the scientific community might leverage policies that promote the participation of people that are interested in the subject, but struggle to enter the community. Future work involves the analysis of more than one year in order to evaluate trends and how people orbiting the subject became authors or participants over time.

Finally, our contribution shows a data informed approach of expanding the scientific community reach in order to foment strategies of bringing more and more people to the core of the community.

References

1. Barbosa, S.D.J., Silveira, M.S., Gasparini, I.: What Publications Metadata Tell Us about the Evolution of a Scientific Community: The Case of the Brazilian Human-Computer Interaction Conference Series. Scientometrics **110**(1), 275–300 (2016). doi:10.1007/s11192-016-2162-4

2. Barbosa, S.D.J., Souza, C.S.: Interacting with public policy: are HCI researchers an endangered species in Brazil? Interactions **18**(3), 69–71 (2011). http://dx.doi.org/10.1145/1962438.1962454

3. Elmacioglu, E., Lee, D.: On six degrees of separation in DBLP-DB and more. ACM SIGMOD Rec. **34**(2), 33–40 (2005)

4. Gasparini, I., Kimura, M.K., Pimenta, M.S.: Visualizando 15 anos de IHC. XII Simpósio Brasileiro sobre Fatores Humanos em Sistemas Computacionais (IHC) (2013)

5. Gasparini, I., da Cunha, L.F., Kimura, M.H., Pimenta, M.S.: Análise das redes de coautoria do simpósio brasileiro sobre fatores humanos em sistemas computacionais. In: Proceedings of the 13th Brazilian Symposium on Human Factors in Computing Systems, pp. 323–332. Sociedade Brasileira de Computação (2014)

6. Gasparini, I., Silveira, M.S., Barbosa, S.D.J.: Caminhos Migratórios da Comunidade Brasileira de IHC. In: Proceedings of the 14th Brazilian Symposium on Human Factors in Computing Systems (IHC 2015), pp. 242–251 (2015)

7. Gasparini, I., Barbosa, S.D. J., Silveira, M.S., Mendonça, F.C.: Auto(re)conhecimento: refletindo sobre a influência das publicações do IHC no próprio IHC. In: Proceedings of the 15th Brazilian Symposium on Human Factors in Computing Systems (IHC 2016) (2016)

8. Henry, N., Goodell, H., Elmqvist, N., Fekete, J.D.: 20 Years of four HCI conferences: a visual exploration. Int. J. Hum.-Comput. Interact. **23**(3), 239–285 (2007)

9. Liu, K.: Semiotics in Information Systems Engineering. Cambridge University Press, Cambridge (2000)

10. Liu, X., Bollen, J., Nelson, M.L., Van de Sompel, H.: Co-authorship networks in the digital library research community. Inf. Process. Manage. **41**(6), 1462–1480 (2005)

11. Norton, M.: Introductory Concepts in Information Science. Information Today Inc., Medford (2000)

12. Santana, V.F., Baranauskas, M.C.C.: WELFIT: A remote evaluation tool for identifying web usage patterns through client-side logging. Int. J. Hum. Comput. Stud. **76**, 40–49 (2015)

13. Souza, C.S., Baranauskas, M.C.C., Prates, R.O., Pimenta, M.S.: HCI in Brazil: lessons learned and new perspectives. In: Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems (IHC 2008), pp. 358–359. Sociedade Brasileira de Computação, Porto Alegre, Brazil (2008)

14. Stamper, R.K.: A semiotic theory of information and information systems/applied semiotics. In: Invited Papers for the ICL/University of Newcastle Seminar on 'Information', 6–10 September (1993)
15. Stamper, R.K.: Extending semiotics for the study of organisations. In: Proceedings of Conference on Semiotics and the Information Sciences (1998)
16. Sun, Y., Barber, R., Gupta, M., Aggarwal, C.C., Han, J.: Co-author relationship prediction in heterogeneous bibliographic networks. In: 2011 International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 121–128. IEEE (2011)
17. W3TECHS—Web Technology Surveys. Usage of traffic analysis tools for websites (2016). https://w3techs.com/technologies/overview/traffic_analysis/all