The Privacy Research Community in Computing and Information Technology

Charles Perez*, Karina Sokolova*

*PSB Paris School of Business, Chair D^3 Digital, Data, Design,
Paris, France
email:{c.perez, k.sokolova}@psbedu.paris

***Abstract*—The technologies of information and communications are part of our day to day activities. From computers to smartphones and with the success of social media and the Internet of Things (IoT), we are now surrounded and fully part of a digital society that produces a big amount of data. In this context, privacy is raising importance in computing and information technology. In this article, we propose a study of the privacy research community. We examine over 13,646 articles published on privacy during the last ten years. We focus our analysis on co-authorship and identify the dynamics and key researchers of this domain.**

Keywords*–*Privacy, Information Technology, Bibliometry, Computer science, Survey, Co-authorship graph

I. INTRODUCTION

In European Union, privacy is considered to be a fundamental human right. During the last decades, with the rapid evolution of technologies, personalized services and big data, the interests in privacy rapidly grows and the privacy research community seems to expand. With the European legislation evolution and an introduction of "Privacy by Design" (PbD) notions applicable to all information systems [1], privacy research starts to englobe information system research and computer science: researchers currently work on bridging the gap between legal notions and information systems engineers to propose adapted solutions for modern systems design and evaluation [2]. Modern technologies, such as World Wide Web, mobile systems [3][4], Internet of Things (IoT) [5], data treatment and sharing [6] strongly impacts privacy research field and community. With the popularization of digital social networks and sharing services, user behavior regarding privacy evolves: privacy research field expands with the notions of user education, visibility and transparency [7][8]. Privacy becomes a large multidisciplinary research field treating legal and technological aspects, privacy models [9], design patterns [10], Privacy Enhancing Technologies (PET) [11][12][13], effective user interface [14], and much more. However, in our knowledge, no bibliometric research has been yet conducted on the study of the privacy research field community.

In this paper, we investigate the computing-related privacy research field by exploring the evolution of the community and co-authorship over the last 10 years. Our research is based on a set of 13,646 articles collected from the well-known Association for Computing Machinery (ACM) digital library in October 2016. We provide a set of statistics on this particular field and apply social network analysis techniques to better understand the evolution of this research community and identify the most relevant contributors.

The article is organized as follows. Section 2 presents the methodology used for data collection, general data analysis metrics and the co-authorship analysis. Section 3 highlights the obtained results: general dataset metrics and interpretation as well as co-authorship graph analysis results. We also compare the obtained results with the state-of-the-art works on other research communities. Section 4 presents related works and section 5 concludes this article.

II. METHODOLOGY

A. Data collection and preprocessing

We collected 13,646 publications from the ACM digital library [15] in October 2016 using the import.io software and a crawler setup for this purpose. All articles were published between years 2006 and 2016 and are a collection of conference proceedings and journal articles published by the ACM digital library and partner publishers. All of the collected articles mention 'privacy' either in the title, keywords or the abstracts. The following features were collected about the articles: the title, the abstract, the list of authors, the list of keywords, the number of downloads (6 weeks, 12 weeks and overall), the number of citations, the publisher and the publishing date (month and year).

When working with human generated data and scientific articles in particular, a few preprocessing steps are required for preprocessing. For example, it is a common observation that the same author is mentioned using different strings in different articles: "Heather Ritcher Lipford" and "Heather Richter Lipford", "Renè Mayrhofer" and "Rene Mayrhofer", "Alvaro A. Cardenas" and "Alvaro Cardenas" are observed in our dataset. To homogenize the authors' names, we first performed the following preprocessing steps using Regular Expressions: 1-Remove dots, 2-Replace dashes by simple spaces, 3-Remove all diacritical marks (e.g., 'è' becomes 'e'), 4-Remove titles and honorifics.

From the preprocessed authors list, we calculate the Levenshtein distance [16] between author names to measure potential misspellings that could not be detected by the four aforementioned techniques. Note that the Levenshtein distance between two strings is measured as the minimum number of basic operations (i.e., deletions, insertions or substitutions) needed to transform the first string into the second string. We found that most of errors (90%) can be detected by matching authors that have a Levenshtein distance below 2. A manual investigation of potential matches was performed to avoid any abusive match. Asian names were often found to be false positives due to the names' shortness (e.g., "Yun Zhang" and "Jun Zhang"). After

 ISBN: 978-1-61208-534-0

resolving the entity disambiguation problem with authors, we finally obtained a total of 16,766 distinct authors for the 13,646 articles.

B. Dataset general statistics

1) Global measures: From the obtained dataset, we compute a set of metrics to have a vision of the broadness of the research field and its overall weight and interest. Given the total amount of contributors denoted $\mathcal{A}$ and the total quantity of contributions denoted $\mathcal{N}$, we compute the average number of contributions per author $\langle \mathcal{N}_a \rangle$ and the average number of authors per contribution $\langle \mathcal{A}_n \rangle$.

Since we analyze the last ten years of the field, we want to highlight the evolution of the field over time. For this purpose, we have computed a set of dynamic metrics, such as the number of articles per year y (denoted $\mathcal{N}_y$), the number of authors per year (denoted $\mathcal{A}_y$), the distribution of authors regarding their number of contributions and the authors publishing lifetime.

Concerning the distribution of authors regarding their number of contributions, we verify if our dataset respects the Lotka's law [17]. This law states that the number of researchers publishing exactly X contributions is a fraction of the number of authors publishing only one. This fraction is expressed by the equation 1.

$$Y = \frac{C}{X^k} \tag{1}$$

Where X is the number of publications, Y the relative frequency of authors with X publications, and k and C are constants depending on the specific field. It is admitted that the k parameter for bibliometrics is generally about 2.

The authors publishing lifetime L_a for a given author is a duration (measured in years) when the authors considered to be part of the research field. The lifetime L_a of an author in the research field is defined by equation 2.

$$L_a = 1 + (t_{out}(a) - t_{in}(a)) \tag{2}$$

Where $t_{in}(a)$ is the year of its first contribution in the domain (arrival time when the author is considered to be the new author) and $t_{out}(a)$ the year of its last contribution (leaving time). We also measure the average authors lifetime denoted $\langle L_a \rangle$.

C. Co-authorship analysis using graph theory

We propose to analyse the collaborations between authors by creating a co-authorship graph. In a first part, the graph is analyzed at the broad scale to obtain a general vision of the research field. In a second part, we deeply investigate the position of authors in the graph and characterize the importance of contributors using centrality metrics.

1) Graph construction and general metrics: The undirected weighted co-authorship graph is denoted $G(N, E)$ where each author is a node $n \epsilon N$ of the graph and each edge between two nodes is created when two authors are found to be co-authors of the same article. The edge is weighted by the number of the authors' collaborations: more the authors collaborated, the higher is the edge weight. In order to build the graph using the raw data harvested by import.io, we converted a list of authors of each article into a set of edges between all co-authors of the article using Talend Open Studio 'Extract, Transform and Load' (ETL) software.

First, we calculate density, clustering coefficient, degree distribution and average path length of the co-authorship graph to capture some general statistics of the privacy research field. The density of the graph reveals the probability that two given researchers of the privacy field collaborate together. It is measured as $d = {}^{2|E|}/_{|N||N-1|}$ where $|E|$ is the number of observed edges and $|N|$ the total number of nodes.

The local clustering coefficient reveals how likely the co-authors of a given author are also co-authoring papers together. It is measured as shown in equation 3.

$$C_i = \frac{2L_i}{k_i(k_i - 1)} \tag{3}$$

Where L_i represents the number of links between the k_i neighbors of node i. The local coefficient of clustering equals 0 if neighbors are not collaborating at all and 1 if all neighbors are collaborating with each other (they form a complete graph). The average clustering coefficient captures the general vision of how co-authors of a same author tend to collaborate at the general scope of the graph/research field. Average path length, diameter, degree distribution and other metrics are applied to the dataset but not discussed in this paper. For more information on social network analysis metrics, the reader can refer to [18]. All of the general metrics are compared to other fields in order to highlight specificities of the privacy research community.

In order to investigate the communities of researchers, we apply the modularity based clustering algorithm [19]. Note that all graph visualizations of this paper are performed using the Gephi visualisation software [20].

2) Algorithms to identify key authors: We apply several centrality algorithms to identify the key users of the graph. Centrality defines the importance of a node depending on its position in the graph [21]. We applied the following measure of importance: degree, weighted degree, betweenness centrality, closeness centrality and PageRank [22]. Degree measures the number of distinct collaborators (number of edges). Weighted degree highlights the number of collaborations (sum of edges weights). Betweenness centrality measures how intermediate a given node is in the graph. It is based on appearance of a node in the shortest paths between any couple of nodes in the graph. It can be interpreted as with a kind of diversity in collaborations. PageRank is built on the hypothesis that important authors are authors whose collaborators are also important (high number of links). It is an iterative process.

III. Results

A. Dataset general statistics

The collected dataset contains 13,646 conference proceedings and journal articles from the ACM digital library having 16,766 distinct authors. Most of the collected articles were published by ACM (11,587 articles) and a minority was published by partner publishers, such as IEEE Press (256 publications), IEEE Computer Society (176 publications), Australian Computer Society Inc. (116 publications) and Consortium for Computing Sciences in Colleges (102 publications).

Fig. 1 highlights the number of the articles published every year by the privacy community. The blue line highlights the number of publications observed in our dataset; the orange dashed line is a linear trend-line. We observe that the field gained regularly in popularity and particularly between years 2007 and 2009 probably due to smartphones that cause many privacy and security concerns. We observe linear augmentation in yearly publications: the privacy field gains in the average 202 articles by year over the last ten years ($R^2 > 0.93$).

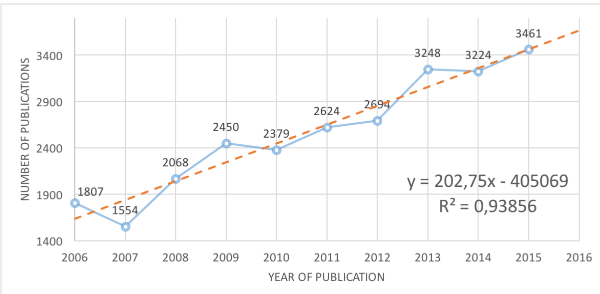

Figure 1. Evolution and trends in the number of published articles by year over the last ten years

Fig. 2 shows the distribution of authors regarding the number of articles they published (blue dots). A trend line is displayed as dashed yellow line. We observe that the decrease in publications by authors in the privacy community follows the Lotka's law with parameters $C = 28,424$ and exponent $k \approx 3$. We note that the exponent k is closer to 3 than to 2 as expected in well known bibliometrics datasets [17]. According to [23], this observation reveals that the privacy research field is a particularly productive community that is overestimated by the Lotka's law with exponent $k = 2$ and instead tends to follow the cube relationship ($Y = C/X^3$).

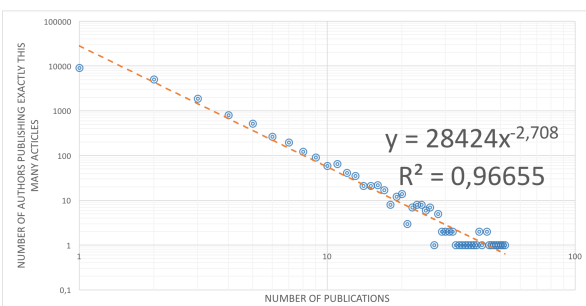

Figure 2. Research production in privacy field (logarithmic scale)

Fig. 3 shows the evolution of the authors in privacy community over the last 10 years. Fig. 3 shows the total number of unique authors by year, new authors and leaving authors. The new authors are the authors that publish their first article on the studied research topic at a given year. The authors are considered to be leaving the community if no article were published by the author after the given year. We observed that 15% of researchers have a 2 years delay between two publications, therefore we do not consider leaving authors for the years 2014 - 2015 as they could still publish in the near future and probably are still active in the community (e.g., working/reviewing/waiting decisions of papers). Even if the majority of authors are new authors, we observe that the community of republishing authors grows linearly: each year, the community gains authors and regularly keeps some of the new authors.

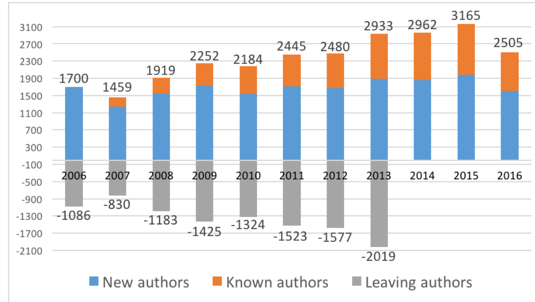

Figure 3. New, republishing and leaving authors by year.

Fig. 4 represents the number of authors according to their community publishing lifetime L. The publishing lifetime represents the number of years the author were observed in the community. The upper score indicates the number of incoming authors while the lower score indicates the number of leaving authors. We observe that most of the privacy authors (86,5%) had only a one year lifetime ($L = 1$), it means that the majority of the new authors are also leaving authors. Those authors could be researchers from different fields having short collaborations with the privacy community researchers. Fig. 4 depicts the number of researchers that stayed in the community more than a year. The minority of the authors leave the community in two years and approximately the same amount of authors were observed in the community during the full study period. Most of the privacy researchers contributes in the community for a period of 5 years ($\langle L_a \rangle = 5$).

We propose to have a closer look on the publication lifetime of the top 5 authors chosen by their lifetime and the number of published articles (Fig. 5). All the five authors are a part of the community from at least 2006 and are active publishers till 2016. Elisa Bertino (grey line) is the most productive author publishing in average 8 articles by year (publishing peak in 2009 with 18 published articles). Lorrie Faith Cranor (yellow line) and Ahmad Reza Sadeghi (orange line) in average publish 5 articles per year. Ninghui Li (blue line), Adam Lee (cyan line) and Ting Yu (green line) have an average publishing of 4 articles per year.

B. Co-authorship analysis

1) Large scale observations: Fig. 6 shows a global view of the obtained graph $G(N, E)$ composed of 16,766 nodes and 21,113 edges. One can observe a giant component that illustrates that most of privacy community has a core research community of collaborators and a set of isolated small connected

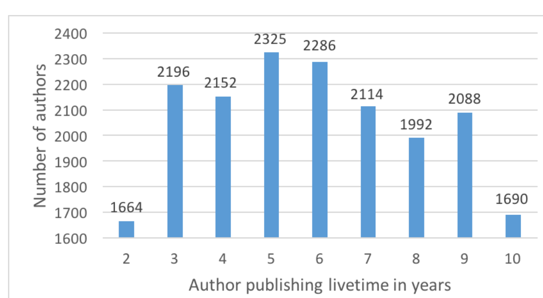

Figure 4. Number of authors according to their publishing lifetime L

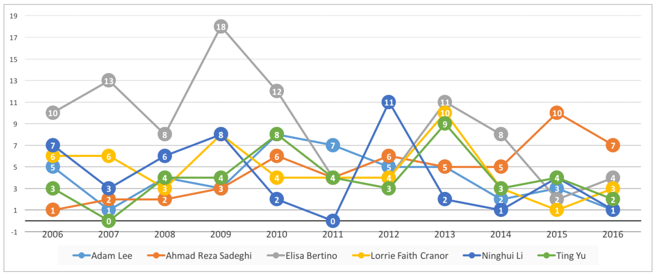

Figure 5. The publishing activity of the top 5 privacy authors for the last ten years

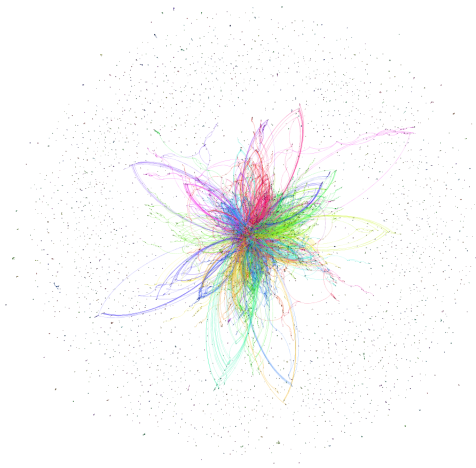

Figure 6. Privacy community co-authorship graph.

components (isolated collaborations). The giant component consists of 6,959 authors (41.5% of the total author number) and 11,373 collaborations (53,8% of total collaborations). An interactive graph of a subset of the privacy research community is available online at [24]. The density of the observed network is measured as ≈0.00015, which is relatively low but not surprising. This number illustrates that there is approximately 0.015% of chance that a random couple of researchers of the community has ever co-authored an article.We also observe a high number of connected components not belonging to the giant component (about 2,632). Most of these components are of size 2, 3 and 4 (small dots on the figure).

To get a better understanding of the specificities of the privacy research field, we propose to compare global characteristics of the graph to well-known co-authorship graphs of other research fields (i.e., Management, Physics and IT). The comparison features are displayed in Table I: number of authors, number of papers, average degree, main component size, main component percent, clustering coefficient, mean authors per paper, mean papers per author. We observe that Privacy research field is the unique sample that has a higher number of authors than the number of papers. This matches with the mean authors per paper feature that is significantly high in our dataset. This indicates that researchers of privacy research fields tend to publish with a higher amount of co-authors. This is certainly a proof of the particular dynamic of the field. However, we note that the average degree (number of collaborators by author) is relatively low compared to the mean authors per paper. This may reveal diversity in behavior between researchers (some having a very high number of collaborators versus some having a very few collaborators - see Fig. 7). The clustering coefficient is significantly high compared to the other research fields which reveals that if an author x publishes with author y and author z, it is very likely that y and z publishes also together. Note that this result is also partly due to the high number of co-authors per papers.

TABLE I. COMPARISON OF CO-AUTHORSHIP FEATURES FOR DISTINCT RESEARCH FIELDS

	Privacy	**Management**	**Physics**	**IT**
Number of authors $\mathcal{A}$	16,766	10,176	52,909	11,994
Number of papers $\mathcal{N}$	13,646	11,022	98,502	13,169
Average degree (collaborators per author)	2.5	2.43	9.7	3.59
Main component size	6,959	4,625	4,4337	6,396
Mean authors per paper $\langle \mathcal{A}_n \rangle$	3.32	1.88	2.530	2.22
Mean papers per author $\langle \mathcal{N}_a \rangle$	2.25	2.04	5.1	2.55
Main component %	41.5%	45.4%	85.4%	57.2%
Clustering coefficient	0.8	0.681	0.430	0.496
Reference	*Our study*	[25]	[26]	[26]

2) Authors characterization: We have applied the following centrality measures to the co-authorship graph: Degree, Weighted degree, Closeness, Betweenness and PageRank. Table II presents the top 5 authors according to the degree, weighted degree and PageRank. Three researchers (Elisa Bertino, Wei Wang, Adam Lee, Ahmad Reza Sadeghi) are highlighted in bold due to their apparition in top 5 of the three metrics. Table III shows top 5 authors according to the closeness and betweenness centralities where we observe 4 authors in common: Elisa Bertino, Michael Reiter, Ari Juels and Gene Tsudik (shown in bold).

The unweighted degree measures the total number of distinct collaborators for each author. The maximum degree of 68 is observed for Elisa Bertino; Ahmad Reza Sadeghi is observed to be the second most collaborative author having nearly half less collaborators (37 co-authors).

Fig. 7 shows the degree distribution representing the number of authors according to the number of unique collaborators they had during the last 10 years. A relevant ratio of authors (46%) collaborated with only 2 researchers and only a very small portion of authors collaborated with more that 15 researchers (0.5%). It is, however, more common to have a single collaborator (23.9%) than having three collaborators (16%). Considering only the giant component of the graph, most of authors have at least 3 collaborators and the average degree rises to 3.2.

It is interesting to note that authors having the highest

 ISBN: 978-1-61208-534-0

TABLE II. TOP 5 AUTHORS ACCORDING TO DEGREE, PAGERANK AND WEIGHTED DEGREE MEASURES

Rank	Author	**Degree**	Author	**Weighted degree**	Author	**PageRank**
1	**Elisa Bertino**	68	**Elisa Bertino**	138	**Elisa Bertino**	0.0023
2	**Ahmad Sadeghi**	37	**Adam Lee**	74	Wei Wang	0.0013
3	Wei Wang	31	Ting Yu	68	**Ahmad Sadeghi**	0.0012
4	**Adam Lee**	31	Li Xiong	66	**Adam Lee**	0.001
5	Ninghui Li	29	**Ahmad Sadeghi**	64	Mahesh Tripunitara	0.0009

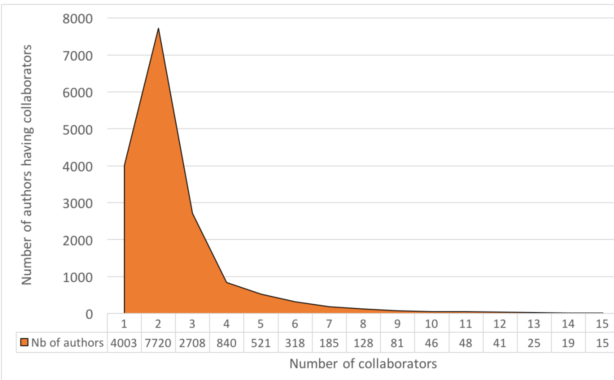

Figure 7. Authors and the number of unique collaborators (Degree distribution of the graph $G(N, E)$)

Figure 8. Author graph filtered by the node degree up to 25

number of collaborators do not collaborate with each other. Fig. 8 depicts the graph where nodes were filtered by degree to only keep track of authors having at least 25 unique collaborators. The size of the nodes shows the degree of the node before the filter is applied, authors names are proportional to the node size, the link shows the collaboration and the node colors indicated the modularity class (modularity class reflects the different clusters identified using the optimisation based algorithm). Elisa Bertino co-authored with only one researcher that has as many collaborators as herself (Ninghui Li with 29 unique co-authors). Ahmad Reza Sadephi, Wei Wang and Adam Lee both having 31 unique co-authors, Serge Edelman (27 collaborators) and Li Xiong (26 collaborators) never collaborate with each other.

The weighted degree represents the number of collaborations and not only the number of collaborators. Comparing to the degree ranking, we observe that Ting Yu and Li Xiong replaces Wei Wang and Ninghui Li at the top 5. Even if those authors have less unique collaborators (25 for Ting Yu and 26 for Li Xiong), their total quantity of collaborations is higher (68 for Ting Yu and 66 for Li Xiong): it means that they prefer long term collaborations with the same collaborators.

Fig. 9 represents the authors that collaborate the most with each other (more than 9 collaborations). The thickness of the edge corresponds to the edge weight and the node size

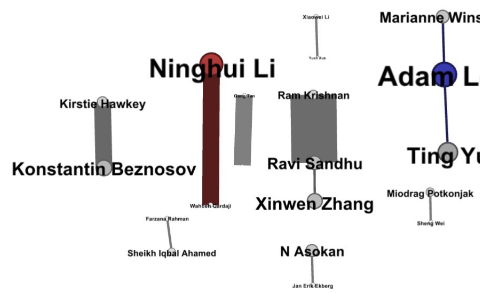

Figure 9. The most collaborative authors.

corresponds to the node degree before filtering. Ram Krishnan and Ravi Sandhu co-authored 12 articles that is the maximum value observed in our dataset. We observe that most of very strong collaborations occurs between couples of authors. Also we note that the privacy domain does not have any very strongly collaborative communities of size 3 or more.

PageRank measures the prestige of the author in the community. In our case, the top 4 authors are equivalent to the top 4 authors ranked according to unweighted degree (Table II). Mahesh Tripunitara replaces Ninghui Li on the fifth place: even if he has less collaborators and collaborations, he appears to be connected to more prestigious nodes.

Betweenness centrality measures how authors appear in between others. Regarding co-authorship, high betweenness can reveal an interdisciplinary researchers that are part of/in the middle between different communities (all belonging to a possible different clusters). Elisa Bertino and Ahmad Reza Sadeghi appear to be highly collaborative with a diversity of collaborators in terms of communities. We observe that Michael Reiter, Gene Tsudik and Ari Juels also appear in-between nodes probably due to a variety in their collaborators' interests (Table III).

Closeness centrality highlights the authors that are the closest to all other authors in the co-authorship network. That would highlight the authors that one would contact if one wants to relate to all/any other author of the network. We observe that the results are similar to the betweenness centrality: only Ahmad Reza Sadeghi is replaced by Peng Ning (Table III).

IV. RELATED WORKS

Social network analyses field gained considerable attention in bibliometrics to measure the evolution of research fields using graphs. Multiple types of bibliographical data may be modeled and analyzed as a directed/undirected and weighted/unweighted graphs. Thus, co-citation graph may be

 ISBN: 978-1-61208-534-0

TABLE III. TOP 5 AUTHORS ACCORDING TO BETWEENNESS AND CLOSENESS CENTRALITIES

Rank	Closeness	Betweenness
1	**Michael Reiter**	**Elisa Bertino**
2	**Ari Juels**	Ahmad Sadeghi
3	**Elisa Bertino**	**Michael Reiter**
4	**Gene Tsudik**	**Gene Tsudik**
5	Peng Ning	**Ari Juels**

built linking articles or authors that cite/are cited by other articles or authors; co-authorship graph links two authors if they co-published at least one article together; keywords graph links keywords that appears together in the same article, etc.

Co-authorship network is frequently used to study scientific collaborations and highlight the key actors of the field. Newman [27] studied co-authors in biomedical research, physics and computer science between 1995 and 1999 and highlighted the similarities between those networks. The authors of [28] studied mathematics and neuroscience between 1991 and 1998. In [29], authors studied scientometrics research collaborations (1980-2012). The authors of [30] analyzed all publications of the ACM Special Interest Group on Management of Data (SIGMOD) conferences between 1975 and 2002. In [31], authors analyses co-publications of ACM and IEEE conferences (1994 and 2000). Resent studies analyzed co-authors in computer science [32], eParticipation [33], industrial ecology [34], front-end of innovation [35] and digital heritage [36] to cite a few.

V. CONCLUSION

In this paper, we investigated the privacy research field from a computing and information technology perspective. We collected and analysed 13,646 publications published by 16,766 authors. We characterized this community with general statistics but also by analysing the underlying co-authorship graph. We show that the privacy research field is growing (up to 200 contributions by year), productive and strongly collaborative (about 2.5 collaborations per author) having high average number of authors per paper (3.32). Authors contribute to the community for an average time of 5 years. Using the co-authorship graph, we identified a set of authors that are key players of the field: Elisa Bertino, Wei Wang, Adam Lee and Ahmad Reza Sadeghi. Despite a strong activity, we highlighted that key authors of these fields do not often collaborate together. This work can be extended by qualitative analyses of the top communities and by topics evolution analyses.

REFERENCES

[1] A. Cavoukian, “Privacy by design: The 7 foundational principles,” 2009.

[2] “Privacy and data protection by design – from policy to engineering,” European Union Agency for Network and Information Security (ENISA), Tech. Rep., December 2014.

[3] J. Häkkilä and J. Mäntyjärvi, “Developing design guidelines for context-aware mobile applications,” in Mobility ’06: Proceedings of the 3rd international conference on Mobile technology, applications & systems. ACM, Oct. 2006.

[4] E. Chin, A. P. Felt, V. Sekar, and D. Wagner, “Measuring user confidence in smartphone security and privacy,” in Proceedings of the Eighth Symposium on Usable Privacy and Security, ser. SOUPS ’12. New York, NY, USA: ACM, 2012, pp. 1–16.

[5] C. Perera, C. McCormick, A. K. Bandara, B. A. Price, and B. Nuseibeh, “Privacy-by-design framework for assessing internet of things applications and platforms,” CoRR, vol. abs/1609.04060, 2016.

[6] S. Yu, “Big privacy: Challenges and opportunities of privacy study in the age of big data,” IEEE Access, vol. 4, 2016, pp. 2751–2763.

[7] K. Sokolova, M. Lemercier, and J.-B. Boisseau, “Respecting user privacy in mobiles: privacy by design permission system for mobile applications,” International Journal On Advances in Security, vol. 7, no. 34, December 2014, pp. 110–120.

[8] F. Stutzman, R. Capra, and J. Thompson, “Factors mediating disclosure in social network sites,” Computers in Human Behavior, vol. 27, no. 1, Jan. 2011, pp. 590–598.

[9] K. Harris, “Privacy on the go,” California Department of Justice, Jan. 2013, pp. 1–27.

[10] M. Hafiz, “A collection of privacy design patterns,” in Proceedings of the 2006 Conference on Pattern Languages of Programs, ser. PLoP ’06. New York, NY, USA: ACM, 2006, pp. 1–13.

[11] G. W. van Blarkom, J. J. Borking, and J. G. E. Olk, “Handbook of privacy and privacy-enhancing technologies,” 2003.

[12] G. D’Acquisto, J. Domingo-Ferrer, P. Kikiras, V. Torra, Y.-A. de Montjoye, and A. Bourka, “Privacy by design in big data: An overview of privacy enhancing technologies in the era of big data analytics,” arXiv.org, Dec. 2015.

[13] Y. Deswarte and C. Aguilar Melchor, “Current and future privacy enhancing technologies for the internet,” Annales Des Télécommunications, vol. 61, no. 3, 2006, pp. 399–417.

[14] L. F. Cranor, P. Guduru, and M. Arjula, “User interfaces for privacy agents,” ACM Transactions on Computer-Human Interaction, vol. 13, no. 2, Jun. 2006, pp. 135–178.

[15] A. for Computing Machinery. Acm digital library. [Online]. Available: http://dl.acm.org

[16] V. I. Levenshtein, “Binary Codes Capable of Correcting Deletions, Insertions and Reversals,” Soviet Physics Doklady, vol. 10, 1966, p. 707.

[17] A. J. Lotka, “The frequency distribution of scientific productivity,” 1926.

[18] R. Alhajj and J. Rokne, Encyclopedia of Social Network Analysis and Mining. Springer Publishing Company, Incorporated, 2014.

[19] M. Newman, “Fast algorithm for detecting community structure in networks,” Physical Review E, vol. 69, September 2003.

[20] M. Bastian, S. Heymann, and M. Jacomy. Gephi: An open source software for exploring and manipulating networks. [Online]. Available: http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154 (2009)

[21] L. C. Freeman, “Centrality in social networks conceptual clarification,” Social Networks, 1978, p. 215.

[22] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web.” no. 1999-66. Stanford InfoLab, November 1999, Technical Report, previous number = SIDL-WP-1999-0120.

[23] D. J. d. S. Price, Little science, big science. New York: Columbia Univ. Press, 1963.

[24] K. Sokolova and C. Perez. Privacy co-authorship graph. [Online]. Available: http://www.charlesperez.net/privacy

[25] J. F. Acedo, C. Barroso, C. Casanueva, and J. L. Galan, “Co-authorship in management and organizational studies: an empirical and network analysis,” Journal of Management Studies., vol. 43, 2006, pp. 957–983.

[26] M. Newman, “Scientific collaboration networks. i. network construction and fundamental results,” Rev. E, vol. 64, 2001, p. 2001.

[27] ——, Who Is the Best Connected Scientist: A Study of Scientific Coauthorship Networks, E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

[28] A. L. Barabasi, H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, “Evolution of the social network of scientific collaborations,” arXiv.org, Apr. 2001.

[29] M. Erfanmanesh, V. A. Rohani, and A. Abrizah, “Co-authorship network of scientometrics research collaboration,” Malaysian Journal of Library & Information Science, 2012.

[30] M. Nascimento, J. Sander, and J. Pound, “Analysis of SIGMOD’s co-

authorship graph," ACM SIGMOD Record, vol. 32, no. 3, Sep. 2003, pp. 8–10.

[31] X. Liu, J. Bollen, M. L. Nelson, and H. Van de Sompel, "Co-Authorship Networks in the Digital Library Research Community," arXiv.org, Feb. 2005.

[32] G. Cabanac, G. Hubert, and B. Milard, "Academic careers in Computer Science - continuance and transience of lifetime co-authorships." Scientometrics, 2015.

[33] E. Kaliva, D. Katsioulas, E. Tambouris, and K. Tarabanis, "Understanding researchers collaboration in eparticipation using social network analysis," Int. J. Electron. Gov. Res., vol. 11, no. 4, Oct. 2015, pp. 38–68.

[34] J. Kim and C. Perez, "Co-authorship network analysis in industrial ecology research community," Journal of Industrial Ecology, vol. 19, no. 2, 2015, pp. 222–235.

[35] G. H. S. Mendes and M. G. Oliveira, "Bibliometric analysis of the front-end of innovation," in 2015 Portland International Conference on Management of Engineering and Technology (PICMET). IEEE, Aug 2015, pp. 648–661.

[36] S. Münster and M. Ioannides, "A scientific community of digital heritage in time and space," in 2015 Digital Heritage, vol. 2, Sept 2015, pp. 267–274.