JOURNAL OF SOCIAL COMPUTING
ISSN 2688-5255 01/05 pp287−302
Volume 3, Number 4, December 2022
DOI: 10.23919/JSC.2022.0018

Characterizing and Understanding Development of Social Computing Through DBLP: A Data-Driven Analysis

Jiaqi Wu, Bodian Ye, Qingyuan Gong, Atte Oksanen, Cong Li, Jingjing Qu, Felicia F. Tian, Xiang Li, and Yang Chen*

Abstract: During the past decades, the term "social computing" has become a promising interdisciplinary area in the intersection of computer science and social science. In this work, we conduct a data-driven study to understand the development of social computing using the data collected from Digital Bibliography and Library Project (DBLP), a representative computer science bibliography website. We have observed a series of trends in the development of social computing, including the evolution of the number of publications, popular keywords, top venues, international collaborations, and research topics. Our findings will be helpful for researchers and practitioners working in relevant fields.

Key words: social computing; Digital Bibliography and Library Project (DBLP); bibliometric; evolution; visualization

1 Introduction

First brought out in 1994, the term "social computing" has become more and more popular within the intersection between computer science and social science. It was first defined by Schuler as any type of computing application in which software serves as a focus for a social relation[1]. In his opinion, social computing could be any software application for social relationship purposes. After that in 1999, researchers like Dryer et al.[2] illustrated that social computing is the interplay between person, social behaviors, and interactions with computing technologies. At that moment, social computing formed its essential concept: a novel subject lies in the intersection between computer science and social science[3].

During recent years, the trend of interdisciplinary research enriched the social computing discipline. In the meantime, new concerns on privacy[4], crowdsourcing[5], hate and harrassment[6], and gender issues[7] are also added to the popularity and complexity of this emerging subject. Consequently, it is of great importance for researchers and practitioners to observe and analyze research trends of social computing and its changing kernel. Our work also tries to find answers to the following two questions.

- **Q1**: How do we define social computing at this stage of time? What characteristics allow us to identify a study as social computing research?
- **Q2**: What are the characteristics and trends of the development of social computing research?

• Jiaqi Wu, Bodian Ye, Qingyuan Gong, and Yang Chen are with the Shanghai Key Lab of Intelligent Information Processing, the School of Computer Science, Fudan University, Shanghai 200438, China. E-mail: {wujq18, gongqingyuan, chenyang}@fudan.edu.cn; bdye22@m.fudan.edu.cn.
• Atte Oksanen is with the Faculty of Social Sciences, Tampere University, Tampere 33100, Finland. E-mail: atte.oksanen@tuni.fi.
• Cong Li is with the School of Information Science and Technology, Fudan University, Shanghai 200438, China. E-mail: cong_li@fudan.edu.cn.
• Jingjing Qu is with Shanghai Artificial Intelligence Laboratory, Shanghai 200232, China. E-mail: qujingjing@pjlab.org.cn.
• Felicia F. Tian is with the School of Social Development and Public Policy, Fudan University, Shanghai 200433, China. E-mail: ftian@fudan.edu.cn.
• Xiang Li is with the Institute of Complex Networks and Intelligent Systems, Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, China. E-mail: lix2021@tongji.edu.cn.
* To whom correspondence should be addressed.
Manuscript received: 2022-10-27; revised: 2023-01-05; accepted: 2023-01-12

© The author(s) 2022. The articles published in this open access journal are distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/).

In this paper we aim to provide a new angle of examining the development of social computing by conducting a data-driven analysis. By referring to Digital Bibliography and Library Project (DBLP), we build a new dataset of social computing-related literature since 1994. To ensure both coverage and accuracy of the dataset, we adopt a two-phase matching algorithm to extract social computing-related literature from DBLP. By analyzing the dataset from different aspects, we have made the following three key findings.

First, we find the increasing publication diversity in venues and international academic cooperation. Our findings show that the inclusion in research topics and themes has been growing continuously.

Second, analytic results show changing disciplinary roles. Earlier social computing research centered on classic social science issues with computer science metrics as tools, while evolution of social computing discipline has shaped computer science technologies as strong component in the core of the subject.

Last but not least, sharp increase in research volume has driven social computing research into a steady stage in which researches are relatively saturated. More and more challenging topics are brought to desk requiring researchers to dig deeper for more innovative designs and more insightful results.

Our findings provide insights for researchers and practitioners in social computing-related fields. No matter for junior people or well established experts, our study provides an informative view of the development of social computing and demonstrates the evolution of this field. Moreover, we have explored the discipline from different angles, such as authors' countries, publication venues, keywords, and the collaboration network.

2 Related Work

In this section, we briefly introduce some related studies that inspired our research or provided us with a theoretical basis. We describe early work for social computing research analytic and discipline interaction research in Sections 2.1 and 2.2, which paved ways and inspired our work. In Sections 2.3 and 2.4, we introduce the structural hole theory and existing algorithms for keyword generation.

2.1 Analytics of social computing research

Lots of works have been done to analyze social computing researches through quantitative analysis[8–11]. Authoritative publications are essential sources to learn the overall pattern of a subject in a long time scope. Some influential works[12–15] used bibliometric analysis, i.e., using statistical methods to see the feature and pattern behind the number of documents, year of publication, venues, authors, keywords, co-authorship, and citation.

Many studies were done in the 2010s. Wang et al.[12] focused on the research cooperation mode of social computing. After retrieving data from ISI Web of Science database by keywords from web and experts, they identified key researchers and institutions from four aspects: productivity, influence, collaboration, and transmission. Also, they used citation data to conclude the influence of paper. Lee and Chen[13] used a citation network to derive the intellectual structure of social computing. They identified core research themes in social computing and their relationships. Diversity of academic field affiliations is also assessed by applying the Herfindal-Hirschman Index. Li and Joshi[14] collected literature from four databases (EBSCO, IEEE Xplore, ACM Digital Library, and INFORMS) by querying "social computing" in abstracts or keywords to analyze the definition and research themes of the discipline, using Latent Semantic Analysis (LSA)[15]. Although above works have explored data-driven approaches on social computing, a more systematic analysis from a broader scope and a more comprehensive dataset are needed.

2.2 Interdisciplinary subject interaction

Research on interdisciplinary subject interaction[16, 17] can cast light on the overall logic of discipline development. In 2019, Frank et al.[17] studied the evolution citation graph of Artificial Intelligence (AI) research. They analyzed the association between various academic fields and AI research through the referencing relationship of papers published in each academic field. In this way, researchers are able to reflect on the growth of AI research from a whole new perspective and have insights on how AI's role has changed in relevant fields. However, interdisciplinary subject analysis is still an underexplored area.

2.3 Structural hole theory in social graph analysis

One of the most useful ways for social graph analysis is through the structural hole theory[18]. Considering a social graph, nodes form different communities according to the social connectivities. Nodes acting as bridges or intermediaries between different communities are known as Structural Hole spanners (SH spanners). By using metrics to identify SH spanners, we can easily find important nodes occupying critical bridge positions. In this paper, we use the effective size metric to qualify important countries, research topics, and researchers in collaboration graphs.

For node i in a social graph, redundancy is understood as an investment of time and energy in a relationship with another node q, with whom node j is strongly connected. The effective size of a node i is computed as the non-redundancy of all connected nodes j of i:

$$e(i)=\sum_{j\in N(i)}\left(1-\sum_{q}p_{iq}m_{jq}\right),\quad q\neq i,j \tag{1}$$

$$\text{Redundancy } = p_{iq}m_{jq} \tag{2}$$

where $N(i)$ is the set of neighbors of i, each q is a node different from i and j in the ego network, and p_{iq} is the mutual weight of the edge linking i and q from the matrix of network ties. The m_{jq} is the mutual weight of j and q divided by the largest weight between j and any of j's neighbors.

For the case of undirected social graphs, Borgatti[19] proposed a simplified formula to compute effective size:

$$e(i)=n-2t/n \tag{3}$$

where t is the number of ties in the ego network (not including ties to ego) and n is the number of nodes (excluding ego).

2.4 Keyword analysis

Keyword analysis for a certain subject is one of the most essential parts in our research. Keywords bring convenience for researchers to seize key points for a certain paper at the first glance, as well as attract researchers interested in the corresponding subjects. Chen et al.[20] exploited the keywords in funded projects to analyze research trends, which helps them identify the focus of researchers in China from 2011 to 2015. Madani and Weberi[21] extracted their keywords from the abstracts for exploring evolution of patent mining and patent analysis. Li and Joshi[14] used natural language processing techniques to measure the strength of research themes in order to assess how different research themes change overtime.

3 Constructing Social Computing Literature Dataset

3.1 Overview

We generated a social computing literature dataset from DBLP raw data[22]. Figure 1 shows that we establish a two-phase workflow for data collection. Our goal is to identify a set of papers related to social computing. In the first phase of filtering, we only look at the title of each paper collected by the DBLP XML dataset. By using a set of keywords shown in Table 1, we obtain a subset of papers which have a high chance to be related to social computing. Afterwards, in the second phase of filtering, we crawl the abstracts of these papers and then use another set of keywords shown in Table 2 to finally determine the social computing-related paper entries.

3.2 Detailed process of dataset construction

In this part, we will introduce how we build our research dataset.

3.2.1 DBLP XML dataset extraction

DBLP metadata[22] are stored in XML form. It can be seen as a tree structure file which includes eight different literature tags. We concentrate on two literature tags, i.e., "articles" and "inproceedings", which represent articles in a journal/magazine and a paper in a conference or workshop proceedings, respectively. In order to process these metadata conveniently, we build our XML extractor. For every paper entry, we extract its title, publish date, publisher, and the paper URL by using Python XML package. Finally, we transform the XML record to a pure text record for better later processing efficiency.

3.2.2 First phase keyword filtering

In this step, we obtain the aforementioned first phase dataset by keywords shown in Table 1. Apart from "social", we also conclude 121 keywords that are closely related to social computing according to our statistical investigation and suggestions offered by the editorial board members of the *Journal of Social Computing*. Ultimately, we acquire 185 482 paper entries using our keywords to match their titles. Please note that it is tough to determine whether a paper is related to social computing by looking at the paper title

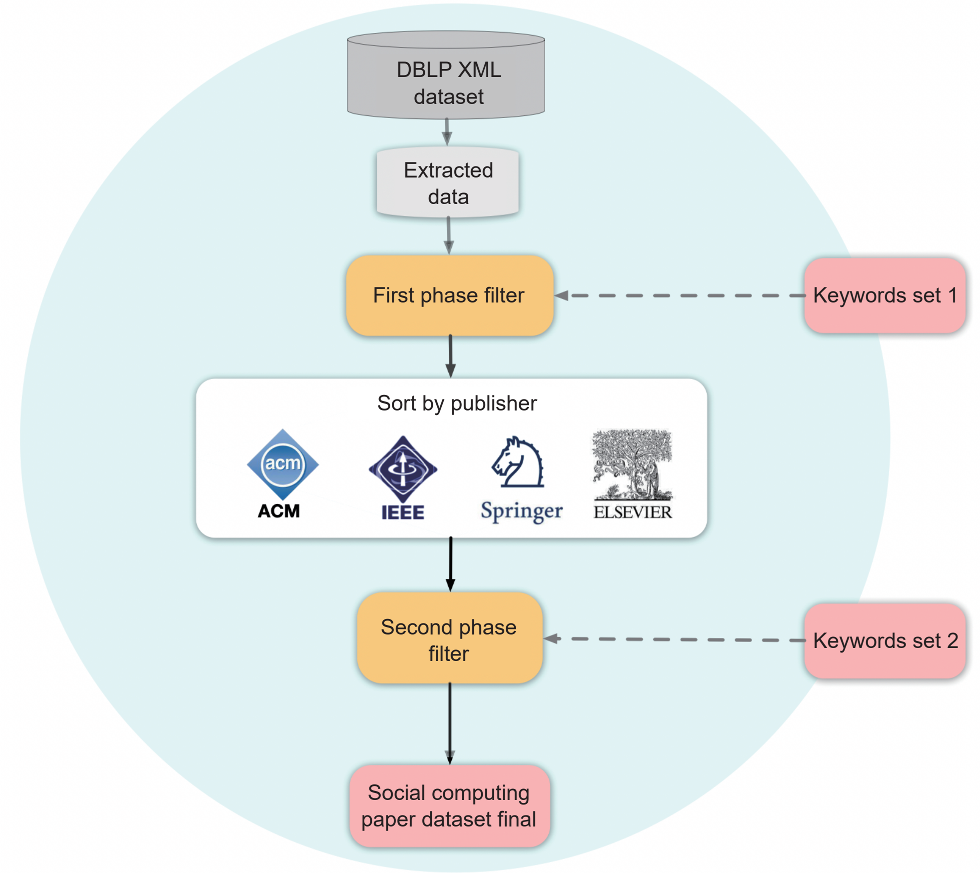


Fig. 1 Workflow of our social computing literature dataset construction. After the first phase keyword filtering, we gain 185 482 valid paper entries. After the second phase of filtering, we get 38 684 valid paper entries for our final dataset.

Table 1 First phase keywords.

Keyword set
"social", "soci", "soc-", "facebook", "youtube", "whatsapp", "instagram", "wechat", "tiktok", "qq", "weibo", "telegram", "snapchat", "kuaishou", "qzone", "twitter", "reddit", "quora", "renren", "orkut", "pinterest", "flickr", "google+", "skout", "yelp", "tinder", "momo", "acquistion", "soci", "gender", "governance", "innovativeness", "cooperate", "marketing", "poverty", "import export equality", "linkedin", "foursquare", "myspace", "employment", "economic", "rural", "income", "pessimism", "selfish", "altruistic", "Institutional", "religion", "immigration", "emmigrant", "expatriate", "satisfaction", "emotion", "psychology", "anthropology", "humanity", "demography", "crowdwisdom", "ontology", "globalization", "crowdsourcing", "population", "residence", "regional", "job", "family", "cultural", "culture", "finance", "capital", "race", "wealth", "humanity", "democra", "children", "adult", "women", "gerontology", "educational", "history", "linguistic", "ethnic", "occupation", "profession", "healthcare", "legal", "environmental", "marriage", "divorce", "personality", "investment", "partnership", "political", "rationality", "occupational", "ideology", "constitutional", "corporate", "elite", "leadership", "liber", "banking", "money", "feminization", "homophily", "psychiatry", "exceptionality", "entrepreneur", "stratification", "lifestyle", "organizational behavior", "special need", "education technology", "urban computing", "internationalization", "internationalisation", "business management", "health care", "user behavior", "knowledge mining", "rational choice", and "collective intelligence"

Table 2 Second phase keywords.

Keyword set
"social networks", "social network", "social media", "community detection", "twitter", "political issues", "social governance", "ethnographic", "qualitative methodologies", "big data analysis", "computer-supported cooperat", "collaborative innovation network", "social network analysis", "social behavior", "individual behavior", "organizational behavior", "social economics", "digital marketing", "web mining", "textual mining", "knowledge mining", "computational social ontology", "machine learning", "ai", "artificial intelligence", "data mining", "social computing application", "social computing", "complex social systems", "system dynamics", "human-computer interaction", "computer-assisted mediation", "HCI", "human-centric computing", "collective intelligence", "crowdsourcing", "computational social science", and "computational linguistics"

only. Therefore, in this phase, we use a more comprehensive keyword set.

3.2.3 Crawling and integration

After getting our first phase data, we need to gain their corresponding information including author, author affiliation, abstract, and keywords, but the DBLP dataset only provides links which help users redirect to their publishers' websites, so we need to recognize different publisher platforms to crawl these specific information respectively. We make use of the DOI information to extract the publisher platform of each paper entry. Now that we are able to count the paper entries whole first phase data according to their corresponding publisher platforms, and our result is shown in Table 3 (arXiv and CEUR Workshop Proceedings are excluded). As we can discover in Table 3, Association for Computing Machinery (ACM), Institute of Electrical and Electronics Engineers (IEEE), Springer, and Elsevier are the four most popular publishers, which cover about 60% of filtered paper entries. To narrow down our target to an operable number of high-quality publishers, in the second phase of paper entry filtering, we focus on these four publishers. Before we enter our next step, there is a brief introduction to these four publishers. ACM is an organization which is dedicated to providing cutting-edge discovery research in the field of computing and information technologies, while IEEE publishes technical literature in electrical engineering, electronics, and so on. For the remaining two publishers, both of them are equipped with literature that includes various disciplines and well-known authorities.

Table 3 First phase filtered dataset's top 15 publisher platforms.

Publisher	Percentage (%)
IEEE	21.2
Springer	18.1
ACM	11.2
Elsevier	10.0
AIS elibrary	2.4
MDPI	1.7
Wiley	1.6
IGI Global	1.6
IOS Press	1.5
Taylor and Francis Online	1.4
INDER Science	1.1
SAGE	0.7
Oxford Academic	0.6
Hindawi	0.6
ACL Anthology	0.6

We mainly make use of Python's Beautiful Soup and urllib packages to analyze HTML files in those platforms' web pages. Four crawlers were implemented for crawling data from each publisher platform. We focus on the information including author, author location, abstract, and keywords. The locations of authors' affiliations can help us conduct analysis on the country information of the literature, while the abstract helps us summarize the research topic or emphasis of the subject. During the period of crawling, some crawling attempts failed due to unstable network conditions and temporary service unavailability of the publishers. For such failures, we would try to re-crawl the paper entry after a period of time, and thus the crawling error rate could be reduced. Finally, we crawled 181 892 valid paper entries. The final error rate is roughly 1.5%. In addition, we need to clarify that ACM and IEEE platforms do not provide keywords, so such information is not included in our crawling result.

3.2.4 Second phase keyword filtering

For the paper entries crawled from the four representative publisher platforms, we are able to filter them through second phase keywords. We selected a set of social computing research keywords chosen by the editorial board members of the *Journal of Social Computing*. Meanwhile, we added a set of popular literature keywords in five important journals of social computing, i.e., *Journal of Social Computing*, *IEEE Transactions on Computational Social Systems*, *ACM Transactions on Social Computing*, *Springer Social Network Analysis and Mining*, and *Elsevier Social Networks*, into our second phase keyword set. Compared with first phase keywords which are wide, now we are preferable to select the keywords that are more precise. Ultimately, we totally select 37 keywords which are shown in Table 2. After the second phase of filtering, we obtain 38 684 valid paper entries.

4 Analysis and Findings

In this section, we illustrate our detailed analytics on our collected social computing research literature dataset and present a series of findings. In Section 4.1, we introduce results from the number of publications, venues, and authors. In Section 4.2, we show the key topics extracted from the dataset and the patterns of the evolving social computing research context. In Section 4.3, we bring out a cross-disciplinary analytic metric to see how the dual discipline feature of social computing changed over time.

4.1 Insights from publication information

After retrieving the final social computing research literature dataset※, we did a thorough analysis on the publication information of all the papers.

4.1.1 Numbers of publications

We first look at the number of publications. As shown in Fig. 2, the total number of social computing-related publications is constantly increasing, with a leap in growth rate in 2008. The growth rate became steady during 2009–2014 and dropped in 2015. The number of social computing-related literature has maintained a slower growth since 2015 till now. Based on the findings, we can define three stages of the development of social computing research, and within each stage the growing rate of literature data is relatively steady.

- Start-up: 1994–2008
- Rapid-growth: 2009–2014
- Steady: 2015–2021

We also look at the number of publications on all four different publishers (ACM, IEEE, Springer, and Elsevier) in our dataset. Results are shown in Fig. 3. We find that the fraction of the number of publications in each platform has been fluctuating since 2015. This corresponds to the beginning of a steady stage in social computing research.

4.1.2 Publication venues

Apart from the number of publications, publication venues are also important to reveal research trends in academia. The study of publication venue diversity helps us see the centralization and decentralization in social computing-related communities, and identifying key publication venues also helps researchers conduct more small-scale targeted analysis. We identify top-20 social computing-related publication venues in each of the three time stages based on their numbers of publications. Results are shown in Table 4.

Results show that at each time stage, there are various significant venues in social computing research. Venues that are most important in all time stages (identified by top-20 in all three stages) include *Hawaii International Conference on System Sciences* (*HICSS*), *Computer Human Interaction Conference* (*CHI*) *Extended Abstracts*, *the Web Conference* (*WWW*), *Expert Syst. Appl.* (*Expert Systems With Applications*), and *CHI*. We can also find that the top 20 venues in 1994–2021 are significantly influenced by publications in the steady time stage (2015–2021).

We also look at publication venue diversity of social computing research in the past 27 years. Publication venue diversity helps researchers conclude the centralization or decentralization trend of research works at different time and identify the formation of publication venue clusters. In this paper, we calculate the venue diversity score through our literature data at different time slots (1994–1997, 1998–2001, 2002–2005, 2006–2009, 2010–2013, 2014–2017, and 2018–2021). We use two classic mechanisms for venue diversity calculation: Gini-Index and Shannon diversity index. Gini-Index[23] is well-known to be used in income equity and economic diversity. It is calculated as the ratio of the area between the perfect equality line and the Lorenz curve divided by the total area under the perfect equality line. Taking values between 0 and

※The dataset is available upon request.

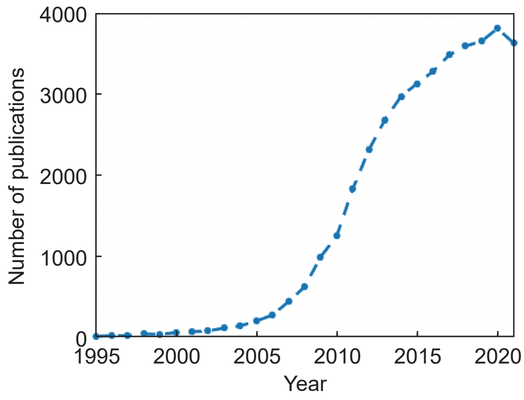


Fig. 2 Overall number of publications.

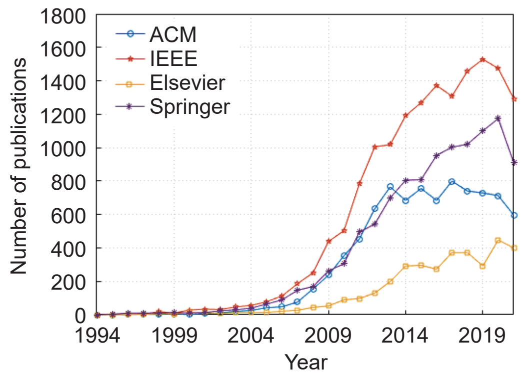


Fig. 3 Number of publications of each publisher.

Table 4 Top 20 publication venues in different time stages.

1994–2008	2009–2014	2015–2021	1994–2021
HICSS	*ASONAM*	*IEEE Access*	*IEEE Access*
IEEE Intell. Syst.	*Comput. Hum. Behav.*	*ASONAM*	*ASONAM*
Soc. Networks	*HICSS*	*IEEE BigData*	*HICSS*
CHI Extended Abstracts	*WWW (Companion Volume)*	*Soc. Netw. Anal. Min.*	*WWW (Companion Volume)*
SMC	*SocialCom/PASSAT*	*CHI*	*Soc. Netw. Anal. Min.*
Web Intelligence	*CIKM*	*WWW (Companion Volume)*	*Comput. Hum. Behav.*
IEEE Pervasive Comput.	*WWW*	*IEEE Trans. Comput. Soc. Syst.*	*CHI*
IEEE Softw.	*CHI*	*Telematics Informatics*	*IEEE BigData*
WWW	*CSCW*	*Multim. Tools Appl.*	*CIKM*
ISI	*Soc. Netw. Anal. Min.*	*Proc. ACM Hum. Comput. Interact.*	*Expert Syst. Appl.*
IEEE Trans. Engineering Management	*CHI Extended Abstracts*	*Expert Syst. Appl.*	*WWW*
Computer	*Expert Syst. Appl.*	*HICSS*	*IEEE Trans. Multim.*
Expert Syst. Appl.	*IEEE Trans. Multim.*	*IEEE Trans. Multim.*	*CHI Extended Abstracts*
CHI	*GLOBECOM*	*Int. J. Inf. Manag.*	*Multim. Tools Appl.*
KDD	*ACM Multimedia*	*IEEE Trans. Knowl. Data Eng.*	*Telematics Informatics*
PAKDD	*SBP*	*SMSociety*	*IEEE Trans. Comput. Soc. Syst.*
CSCW	*SocInfo*	*ICC*	*CSCW*
IEEE Internet Comput.	*SocialCom*	*Knowl. Based Syst.*	*GLOBECOM*
IEEE Trans. Syst. Man Cybern. Part C	*CSE*	*CIKM*	*Soc. Networks*
ICAIL	*INFOCOM*	*WWW*	*ICC*

1, the bigger the index, the more unequal the distribution of wealth. Gini-Index is also used in some researches for diversity analysis[17]. Shannon diversity index is based on Claude Shannon's formula for entropy and estimates species diversity[24]. Our results show in Fig. 4 that the venue diversity in social computing research is constantly growing according to both of these two metrics, indicating that the related fields and scope of influence of social computing discipline continue to expand.

4.1.3 Authors' countries

By extracting data from author affiliations, we can

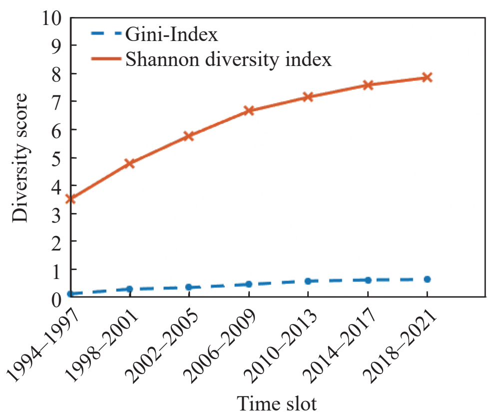


Fig. 4 Venue diversity.

mark country labels for each publication's authors. This information can reveal which country contributes more to this particular research area, and concluding the country information gives us the whole picture of the role different countries played in social computing research.

Since our data are in high volume and with heterogeneous formats, it is important to apply a convenient way for extracting authors' country information. We adopt a 3-layer approach, illustrated in Fig. 5, and we first apply Python's Geotext library to automatically extract countries from author affiliations. Geotext uses a database to identify countries and is efficient for large-scale data. However, Geotext cannot work in certain circumstances where the location to be analyzed is not stored in its database. As in Refs. [25, 26], we use Google Map APIs to identify those extractions failed by Geotext. In detail, we use Google Map APIs to convert location strings to geographic coordinates, and then use Google Map APIs to obtain the country information from geographic coordinates. For those locations with multiple country results, we mark them as deviant and brought them to the final "manual labeling" stage. We identify those countries failed by Google Map API manually to compensate for the

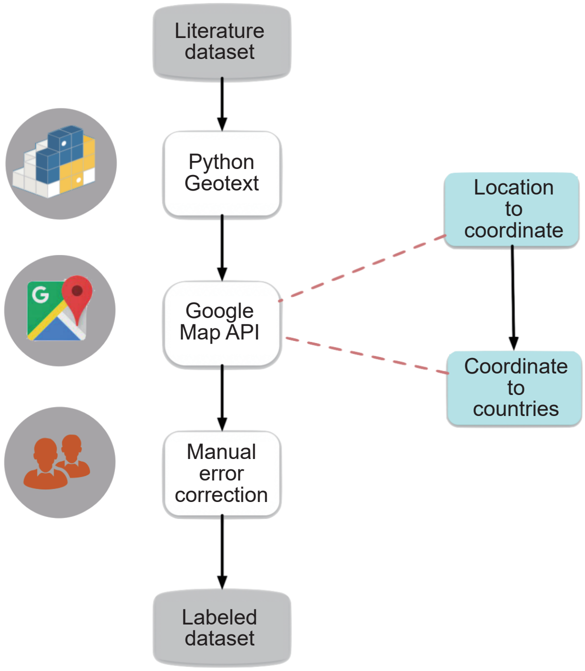


Fig. 5 Authors' country information extraction.

disadvantages of automatic extraction.

Considering the locations of first authors of all publications, we identify the top 10 countries according to the number of publications at every time stage shown in Table 5. In addition to the United States and China having a relatively fixed position, Germany, Italy, and Japan are also important to social computing research. It is notable that India took a giant leap and became the third largest country in social computing research in 2015–2021.

Considering co-authorship information of our literature data, it is possible for us to construct an evolving country collaboration network in Fig. 6.[¤] Each co-authorship with authors from different countries would add to an edge between the two corresponding countries. If the edge already exists, the weight of the edge will be increased by 1. It is important to notice that we only consider a co-occurrence between two countries once for each paper.

We have also extracted the closest country collaboration pairs from the whole network in Table 6. The research weight between two countries is defined by the edge weights between them in the network.

[¤]AT: Austria; BD: Bangladesh; BE: Belgium; CL: Chile; CY: Cyprus; CZ: Czech Republic; DZ: Algeria; EE: Estonia; FI: Finland; IL: Israel; LB: Lebanon; LK: Sri Lanka; MX: Mexico; MY: Malaysia; NL: Netherlands; NZ: New Zealand; PL: Poland; QA: Qatar; RO: Romania; RS: Serbia; RU: Russia; ZA: South Africa; SG: Singapore.

Table 5 Top 10 countries (number of publications) in different time stages.

Rank	1994–2008	2009–2014	2015–2021
1	US	US	US
2	CN	CN	CN
3	JP	DE	IN
4	DE	IT	IT
5	GB	JP	DE
6	IT	CA	JP
7	CA	ES	BR
8	AU	KR	AU
9	ES	IN	GB
10	FR	FR	ES

Note: US: United States; CN: China; JP: Japan; DE: Germany; GB: United Kingdom; IT: Italy; CA: Canada; AU: Australia; ES: Spain; FR: France; KR: Korea; IN: India; BR: Brazil.

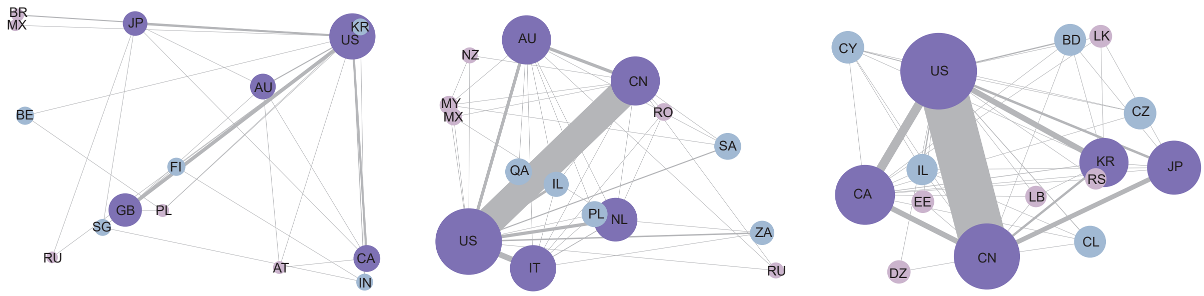
Most of the early research collaborations between countries/regions happen between the United States and European countries, and the pattern changed as India and China became more important in international research collaboration. The country pairs "(China, United States)", "(China, Australia)", and "(India, United States)" ranked the 1st, 3rd, and 7th among all pairs in 2015–2021. By further analyzing the entire country collaboration network, we take the average clustering coefficient into consideration, the results of the three stages are 0.35, 0.53, and 0.64, respectively. This trend hints at the increasing closeness of cooperation between countries from the perspective of the entire country collaboration network.

By referring to the structural hole theory discussed in Section 2.3, we calculate the values of effective size metrics of different countries in each time stage. We are able to extract some core countries with important roles in the country collaboration network, shown in Table 7. We can see that core countries in social computing research are mostly countries in Europe and America, with China taking a giant leap and ranked the 2nd in the steady stage. Additionally, we also calculate the betweenness centrality of each country, we find that the United States obtains the highest value.

4.2 Key research topic analysis

4.2.1 Key research topic extraction

We divide the period from 1995 to 2021 into three stages and analyze the key research topic extracted from each stage, so as to further obtain the research focus of social computing in each stage. To gain more



(a) 1994–2008 (b) 2009–2014 (c) 2015–2021

Fig. 6 Country collaboration network in different time stages.

Table 6 Country pairs with the highest weights in the country collaboration network in different time stages.

Rank	1994–2008		2009–2014		2015–2021	
	Country pair	Weight	Country pair	Weight	Country pair	Weight
1	US, CN	26	US, CN	409	CN, US	1165
2	US, GB	23	US, GB	138	US, GB	294
3	US, DE	16	US, DE	107	CN, AU	293
4	US, CA	14	US, CA	97	CA, US	218
5	US, JP	13	IT, US	81	US, IT	217
6	US, IT	10	US, KR	79	CN, SG	193
7	US, KR	8	US, ES	67	US, IN	190
8	US, ES	8	CN, SG	60	US, AU	158
9	US, AU	7	US, SG	58	US, DE	148
10	US, FR	7	US, FR	55	KR, US	139

Table 7 Core countries and effective sizes in the country collaboration network in different time stages.

Rank	1994–2008		2009–2014		2015–2021	
	Core country	Effective size	Core country	Effective size	Core country	Effective size
1	US	31.59	US	65.44	US	81.83
2	GB	13.74	GB	30.35	CN	51.23
3	DE	11.67	ES	29.89	GB	43.43
4	FR	10.75	FR	29.14	DE	41.38
5	ES	8.71	CN	27.76	FR	39.43

representative key research topics, we choose Yake[27], a keyword extractor which can be applied to our unlabeled collection successfully and earn a good performance. The most prominent advantage of the method is its domain-independence, which also saves our time for training on a certain corpus. Finally we obtain a series of fine-grained key research topics in social computing.

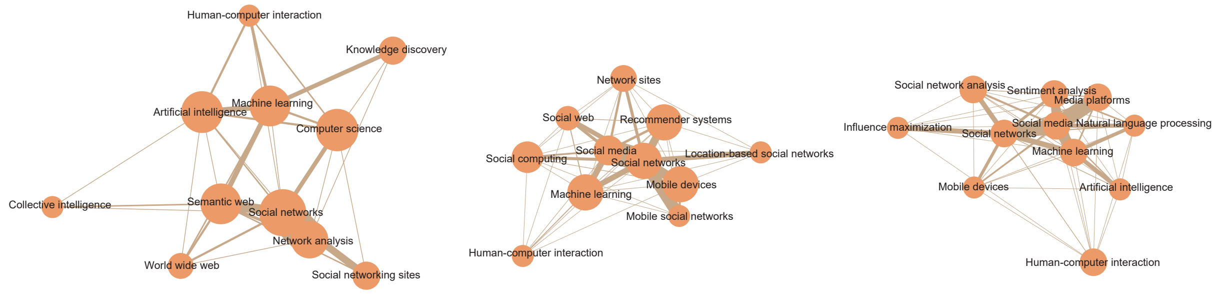
4.2.2 Key research topic network

According to the research topics we generated in the previous section, we construct our research topic networks (shown in Fig. 7 by seeking topic co-occurrence in abstracts. If two topics are shown in the same abstract, then an edge will be added between them in the network. If edge exists already, then edge weight will be increased. In the end we got three networks, each representing the topic relationship at its corresponding time stage.

Our research topic pairs with the highest weights generated by research topic networks are in Table 8. From these topic relationships, we can find insights about social computing research.

In the first stage, the research mostly focused on applying computational metrics to solve classic problems like social network analysis, knowledge discovery, and human-computer interaction.

In the second stage, as the number of social computing-related publications began to grow rapidly, researchers tended to give more attention to various online social media services. The explosive information and data



(a) 1994–2008 (b) 2009–2014 (c) 2015–2021

Fig. 7 Research topic network in different time stages.

Table 8 Research topic pairs with the highest weights in different time stages.

Rank	1994–2008		2009–2014		2015–2021	
	Topic pair	Weight	Topic pair	Weight	Topic pair	Weight
1	Social networks and network analysis	49	Social networks and social media	301	Social media and media platforms	835
2	Social networks and semantic web	14	Social networks and mobile devices	124	Social media and social networks	774
3	Machine learning and artificial intelligence	8	Social networks and recommender systems	102	Social media and machine learning	585
4	Semantic web and machine learning	7	Social media and machine learning	78	Social media and sentiment analysis	435
5	Social networks and computer science	6	Social networks and machine learning	73	Social networks and machine learning	288
6	Machine learning and knowledge discovery	6	Social networks and social web	60	Social networks and influence maximization	267
7	Human-computer interaction and machine learning	4	Social networks and network sites	39	Machine learning and artificial intelligence	226
8	Semantic web and network analysis	4	Social networks and social computing	39	Social media and natural language processing	217
9	Social networks and world wide web	3	Social media and social web	39	Machine learning and natural language processing	189
10	Human-computer interaction and artificial intelligence	3	Social media and mobile devices	37	Social networks and mobile devices	165
11	Semantic web and world wide web	3	Social media and recommender systems	32	Machine learning and sentiment analysis	154
12	Machine learning and computer science	3	Social media and network sites	30	Social networks and sentiment analysis	150
13	Artificial intelligence and computer science	3	Social media and social computing	23	Social media and mobile devices	89
14	Social networks and artificial intelligence	2	Human-computer interaction and machine learning	12	Machine learning and media platforms	87
15	Human-computer interaction and computer science	2	Social web and machine learning	12	Natural language processing and sentiment analysis	84

brought out new forms of interdisciplinary topics including recommendation systems and community detection, with emerging techniques like machine learning[28] and data mining[29]. Another important feature of research in this stage is the development of mobile devices that brings about new problems and ways of social interactions.

In the third stage, the state-of-the-art computational technologies like deep learning and neural networks had become an inseparable part of social computing research[30, 31]. Pervasive online communities and digital information deeply integrated computing technologies in social science-related scenarios. New trends of research were applying large scale data (from large media platforms like Twitter)[32, 33] and context analysis techniques to study issues[34, 35] with social, economic, and political implications. Typical topics in this regard include sentiment analysis and influence maximization. In the new era, social computing became not only an interdisciplinary approach for problem solving, but also a broad discipline context where new concepts, styles, and problems are born.

4.3 Interdisciplinary study

To cast light on the interdiscipline nature of social computing, we design and implement an interdisciplinary study to find out how computer science and social science intersect over time in social computing research.

We first categorize each publication in our dataset to fit in specific pre-defined fields of study. Each field is pre-identified as either computer science-related or social science-related. The detailed categorization method is through querying field keywords in the abstract of each given publication.

Our definition for fields of study comes from two main sources. For computer science-related fields of study, we use the field classification methods and keywords from ACM Digital Library. For social science-related fields of study, we refer to the field classification methods and keywords from Moody's research paper and classification for sociology areas[36]. Detailed fields and corresponding keywords can be found in Tables 9 and 10.

Once we have successfully categorized each publication into specific fields, we construct interdisciplinary network graphs in the three time stages. The principle of graph construction is similar to the construction of research topic networks. If two fields could be identified in the same publication paper, then an edge will be added between them in the graph. If the edge already exists, the weight of the edge will be increased. In Fig. 8, the computer science-related fields and social science-related fields are shown in different colors, and blue edges represent cross-discipline interactions. The size of a node represents the effective size of the corresponding field within the interdisciplinary network. From Fig. 8 we can see that the cross-discipline field interactions have been constantly growing during the evolution of social computing research. What is more, as the field interactions continued to grow, the effective sizes of different discipline fields have been converging, matching the inseparable trend in Section 4.2.2.

Table 9 Computer science fields and keywords.

Field	Keyword
AI	Artificial intelligence, machine learning, computer vision, natural language processing, ai, nlp, ml
Algorithm and theory	Computational theory, algorithms, mathemat
Software and application	Application, software, app
Graph	Graph
Hardware	Hardware, electronic, robotic
System and architecture	Architecture, informational system, computational system, computer system, operating system
Network	Computer network, internet

Table 10 Social science fields and keywords.

Field	Keyword
Social policy	Polic, politic, welfare, poverty
Psychology	Psycholog
Sociology	Social probem, sociology
Education	Education
Health issues	Health, clinical, medic
Economics	Business, econom, market
Linguistics and art studies	Language, art, arts
History and theory	History, theory
Religion studies	Religion
Area and development studies	Rural community, urban community, urban development, rural development, community development,city community city development, demograph

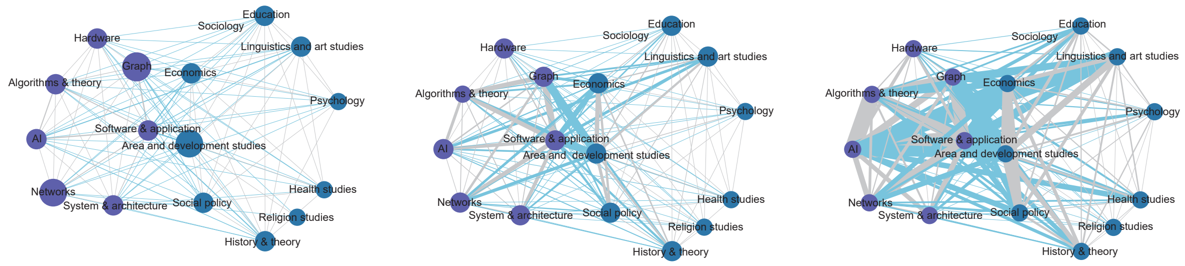
By analyzing cross-discipline pairs with the highest weights in Table 11, we are able to see how the two disciplines intersect with each other over time. Economics has been prone to intersections since social computing was brought out. Early typical computational topics in economics topics were software applications and graph or network analysis, which switched to AI systems in recent years. Also, as time goes by, health studies have become a new emerging topic by its intersection with software and AI technologies. As key components of humanities and culture, linguistics and arts met with the explosive development of AI/ML models and entered a new stage of development, in which human-AI collaboration systems have been developed significantly.

5 Limitation

In this section we discuss the limitations of our work.

5.1 Dataset construction

Some social computing-related literature is published in journals which are not covered by DBLP or the



(a) 1994–2008 (b) 2009–2014 (c) 2015–2021

Fig. 8 Interdisciplinary network in different time stages.

Table 11 Interdisciplinary pairs with the highest weights in different time stages.

Rank	1994–2008		2009–2014		2015–2021	
	Interdisciplinary pair	Weight	Interdisciplinary pair	Weight	Interdisciplinary pair	Weight
1	Software and application, area and development studies	78	Graph, area and development studies	523	Graph, area and development studies	1039
2	Software and application, economics	72	Software and application, area and development studies	452	Software and application, area and development studies	936
3	Software and application, history and theory	64	Software and application, economics	333	AI, linguistics and art studies	920
4	Graph, area and development studies	78	Algorithms and theory, area and development studies	235	Software and application, economics	764
5	AI, linguistics and art studies	60	Networks, area and development studies	200	AI, area and development studies	689
6	Software and application, linguistics and art studies	46	Software and application, linguistics and art studies	183	AI, economics	608
7	AI, area and development studies	44	Graph, economics	175	Algorithms and theory, area and development studies	602
8	Graph, economics	35	AI, linguistics and art studies	170	Software and application, linguistics and art studies	495
9	AI, area and development studies	31	Networks, economics	155	Graph, economics	405
10	Networks, economics	31	Software and application, social policy	145	Algorithms and theory, Economics	380
11	Algorithms and theory, economics	30	Software and application, history and theory	142	AI, social policy	355
12	Networks, area and development studies	30	Graph, history and theory	136	AI, health studies	348
13	Graph, history and theory	29	Algorithms and theory, economics	116	Software and application, social policy	330
14	Software and application, social policy	27	Graph, linguistics and art studies	113	Networks, area and development studies	325
15	Algorithms and theory, area and development studies	25	AI, area and development studies	109	Algorithms and theory, linguistics and art studies	308

literature is not yet included in DBLP at the time of our work, and these papers are not covered in our study. Still, our method can be conveniently applied in other online databases such as Web of Science.

5.2 Choice of keyword set

We apply a two-phase keyword filter manner to construct our social computing research literature dataset from DBLP. Though our method tries best to maintain balance between extensiveness and accuracy, we cannot avoid the subjectivity and bias in keyword choice. There are works[37, 38] focusing on generating high quality keywords from data source like publications or emails, but they are more focused on targeted keyword generation with known data source. However, social computing research is broad in range, which is hard and time-consuming to get a representative publication set for keywords generation. As a result, in this work we choose to mainly base on expert-generated keywords for efficiency and accuracy. In further work, we plan to generate a more comprehensive keyword set by expanding automatic generated keywords starting from a small representative publication set as seeds based on semantic similarity and proximity[39]. The resulting keyword set can in turn be used to retrieve more publication to expand the seeds for a new iteration. The comprehensive keyword set can be expanded through several iterations and be tested by comparing with expert-generated keywords[40].

5.3 Drawbacks of using an automated framework

Our bibliometric analysis is based on DBLP, which is a large-scale online literature database. The use of an automatic framework is necessary for extracting relevant papers from DBLP, and this also introduces some inaccuracy. For example, a small number of papers which contain certain field keywords in their abstract might turn out to be irrelevant to this field. Although we carefully choose keywords and algorithms to enhance the data reliability, there are still chances that irrelevant papers are selected or inappropriate labels (fields and research topics) are given to the paper.

6 Conclusion

In this paper we conduct a thorough research bibliometric analysis on the social computing discipline with literature data from DBLP platform. Using a two-phase framework, we extract a new representative dataset of social computing-related publications. In short, we conclude the following key findings on the development of social computing research.

First, we can see growing diversity in social computing research. More and more countries join social computing research with increasing world wide academic collaboration. Meanwhile, various publication venues have increased the diversity. The sharp increase in research volume has driven social computing research into a steady stage in which research is relatively saturated and more competitive. One issue that makes new research more challenging is the emerging ethical concerns in liberty and privacy.

Second, our findings show the role change and the trend of inseparability of the disciplines within social computing subject. Early social computing studies used traditional computational techniques as tools for classic social science issues like community management or network analysis. These studies often focused on analysis on social science-related datasets. An example is the research that focuses on calculating graph metrics for some OSN sites' social graphs. In recent years, evolution of social computing discipline has shaped computer science as core component of the subject. Computer science technologies, for example, machine learning[41–45], have been developing rapidly and smoothly integrating into social services and systems, shaping the new generation of social interaction, and bringing up new social issues. New forms of social interaction like recommendation systems, AI chatbots, and metaverse have become emerging topics in both academia and industry. What is more, as an interdisciplinary area, the evolution of social computing research comes with inseparability in traditional disciplines. Digital information and algorithms have become crucial in modern society, breeding novel social interaction, and governance metrics inseparable from various computational technologies. The blending trend acts as a binder making it hard to give strict classification for many specific social computing studies.

Last but not least, apart from its significance in improving technology and efficiency, social computing discipline is an observation and reflection on the way of our life in modern society, while also introducing new possibilities for our future. Instead of providing empirical findings or technical solutions only, social computing research's unique characteristics allow us to perceive our everyday issues and reshape our lives from a higher level, telling us where we are standing and where we are going.

These findings will provide insights for researchers and practitioners in relevant area on the development and evolution of social computing. Moreover, we believe that the framework we develop in this paper could be further applied to other disciplines. For example, researchers who aim to conduct a data-driven analysis on other research areas can adopt our framework with own keyword set to gain insights from large literature source. We will further explore the development of other research areas. Also, we could dig deeper into some subareas of social computing and explore their development, for example, security problems in online social networks[41, 43] or social-aware prediction[46–48].

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 71731004, 62072115, 62102094, 62173095, and 61602122), Shanghai Science and Technology Innovation Action Plan Project (No. 22510713600), and Natural Science Foundation of Shanghai (No. 21ZR1404700).

References

[1] D. Schuler, Social computing, *Communications of the ACM*, vol. 37, no. 1, pp. 28–29, 1994.

[2] D. C. Dryer, C. Eisbach, and W. S. Ark, At what cost pervasive? A social computing view of mobile computing systems, *IBM Systems Journal*, vol. 38, no. 4, pp. 652–676, 1999.

[3] D. Zeng, F. -Y. Wang, and K. M. Carley, Guest editors' introduction: Social computing, *IEEE Intelligent Systems*, vol. 22, no. 5, pp. 20–22, 2007.

[4] G. Beigi and H. Liu, Identifying novel privacy issues of online users on social media platforms, *ACM SIGWEB Newsletter*, no. 4, pp. 1–7, 2019.

[5] X. Zhang, X. Chen, H. Yan, and Y. Xiang, Privacy-preserving and verifiable online crowdsourcing with worker updates, *Information Sciences*, vol. 548, pp. 212–232, 2021.

[6] T. Keipi, M. Näsi, A. Oksanen, and P. Räsänen, *Online Hate and Harmful Content: Cross-National Perspectives*. London, UK: Routledge, 2016.

[7] A. Santos and M. Figueras, Instagram and gender inequalities: The discourse of young women regarding social networks, in *Proc. Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*, Salamanca Spain, 2020, pp. 577–581.

[8] F. -Y. Wang, K. M. Carley, D. Zeng, and W. Mao, Social computing: From social informatics to social intelligence, *IEEE Intelligent Systems*, vol. 22, no. 2, pp. 79–83, 2007.

[9] R. K. F. Ip and C. Wagner, Weblogging: A study of social computing and its impact on organizations, *Decision Support Systems*, vol. 45, no. 2, pp. 242–250, 2008.

[10] M. Parameswaran and A. B. Whinston, Social computing: An overview, *Communications of the Association for Information Systems*, vol. 19, no. 1, pp. 762–780, 2007.

[11] H. Ali-Hassan and D. Nevo, Identifying social computing dimensions: A multidimensional scaling study, presented at the International Conference on Information Systems (ICIS 2009), Phoenix, AZ, USA, 2009.

[12] T. Wang, Q. Zhang, Z. Liu, W. Liu, and D. Wen, On social computing research collaboration patterns: A social network perspective, *Frontiers of Computer Science*, vol. 6, no. 1, pp. 122–130, 2012.

[13] M. R. Lee and T. T. Chen, Understanding social computing research, *IT Professional*, vol. 15, no. 6, pp. 56–62, 2012.

[14] Y. Li and K. Joshi, The state of social computing research: A literature review and synthesis using the latent semantic analysis approach, presented at Eighteenth Americas Conference on Information Systems, Seattle, WA, USA, 2012.

[15] T. K. Landauer, LSA as a theory of meaning, in *Handbook of Latent Semantic Analysis*, T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch, eds. New York, NY, USA: Psychology Press, 2007, pp. 3–34.

[16] Y. Bu, W. Lu, Y. Wu, H. Chen, and Y. Huang, How wide is the citation impact of scientific publications? A cross-discipline and large-scale analysis, *Information Processing & Management*, vol. 58, no. 1, p. 102429, 2021.

[17] M. R. Frank, D. Wang, M. Cebrian, and I. Rahwan, The evolution of citation graphs in artificial intelligence research, *Nature Machine Intelligence*, vol. 1, no. 2, pp. 79–85, 2019.

[18] Z. Lin, Y. Zhang, Q. Gong, Y. Chen, A. Oksanen, and A. Y. Ding, Structural hole theory in social network analysis: A review, *IEEE Transactions on Computational Social Systems*, vol. 9, no. 3, pp. 724–739, 2022.

[19] S. P. Borgatti, Structural holes: Unpacking Burt's redundancy measures, *Connections*, vol. 20, no. 1, pp. 35–38, 1997.

[20] X. Chen, J. Chen, D. Wu, Y. Xie, and J. Li, Mapping the research trends by co-word analysis based on keywords from funded project, *Procedia Computer Science*, vol. 91, pp. 547–555, 2016.

[21] F. Madani and C. Weber, The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis, *World Patent Information*, vol. 46, pp. 32–48, 2016.

[22] M. Ley, DBLP: Some lessons learned, *Proc. VLDB Endow.*, vol. 2, no. 2, pp. 1493–1500, 2009.

[23] L. Ceriani and P. Verme, The origins of the Gini index: Extracts from Variabilità e Mutabilità (1912) by Corrado Gini, *The Journal of Economic Inequality*, vol. 10, no. 3, pp. 421–443, 2012.

[24] K. A. Nolan and J. E. Callahan, Beachcomber biology: The Shannon-Weiner species diversity index, in *Proc. 27th Workshop/Conference of the Association for Biology Laboratory Education* (*ABLE*), West Lafayette, IN, USA, 2006, pp. 334–338.

[25] Q. Gong, Y. Chen, J. Hu, Q. Cao, P. Hui, and X. Wang, Understanding cross-site linking in online social networks, *ACM Transactions on the Web*, vol. 12, no. 4, pp. 1–29, 2018.

[26] Y. Chen, J. Hu, Y. Xiao, X. Li, and P. Hui, Understanding the user behavior of foursquare: A data-driven study on a global scale, *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 1019–1032, 2020.

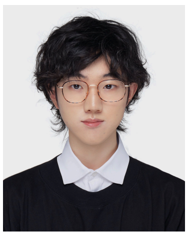
[27] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, YAKE! Keyword extraction from single documents using multiple local features, *Information Sciences*, vol. 509, pp. 257–289, 2020.

[28] Y. Huang, L. Yu, X. Wang, and B. Cui, A multi-source integration framework for user occupation inference in social media systems, *World Wide Web*, vol. 18, no. 5, pp. 1247–1267, 2015.

[29] V. Chandola, S. R. Sukumar, and J. C. Schryver, Knowledge discovery from massive healthcare claims data, in *Proc. 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Chicago, IL, USA 2013, pp. 1312–1320.

[30] N. Albadi, M. Kurdi, and S. Mishra, Investigating the

effect of combining GRU neural networks with handcrafted features for religious hatred detection on Arabic Twitter space, *Social Network Analysis and Mining*, vol. 9, no. 1, p. 41, 2019.
[31] A. Kumar, K. Srinivasan, W. -H. Cheng, and A. Y. Zomaya, Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data, *Information Processing & Management*, vol. 57, no. 1, p. 102141, 2020.
[32] H. Liu, Y. Ge, Q. Zheng, R. Lin, and H. Li, Detecting global and local topics via mining Twitter data, *Neurocomputing*, vol. 273, pp. 120–132, 2018.
[33] C. Y. Chiu, H. Y. Lane, J. L. Koh, and A. L. Chen, Multimodal depression detection on instagram considering time interval of posts, *Journal of Intelligent Information Systems*, vol. 56, no. 1, pp. 25–47, 2021.
[34] A. Al-Zoubi, H. Faris, J. Alqatawna, and M. A. Hassonah, Evolving support vector machines using whale optimization algorithm for spam profiles detection on online social networks in different lingual contexts, *Knowledge-Based Systems*, vol. 153, pp. 91–104, 2018.
[35] J. Cao, Social context modelling and recognition: Current work and future directions, in *Proc. 2017 IEEE International Conference on Pervasive Computing and Communications Workshops* (*PerCom Workshops*), Kona, HI, USA, 2017, pp. 243–243.
[36] J. Moody, The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999, *American Sociological Review*, vol. 69, no. 2, pp. 213–238, 2004.
[37] J. Wang, G. Su, C. Wan, X. Huang, and L. Sun, A keyword-based literature review data generating algorithm—Analyzing a field from scientific publications, *Symmetry*, vol. 12, no. 6, p. 903, 2020.
[38] M. Dredze, H. M. Wallach, D. Puller, and F. Pereira, Generating summary keywords for emails using topics, in *Proc. 13th International Conference on Intelligent User Interfaces*, Gran Canaria, Spain, 2008, pp. 199–206.
[39] A. Joshi and R. Motwani, Keyword generation for search engine advertising, in *Proc. Sixth IEEE International Conference on Data Mining - Workshops* (*ICDMW'06*), Hong Kong, China, 2006, pp. 490–496.
[40] C. D. Hurt, Automatically generated keywords: A comparison to author-generated keywords in the sciences, *Journal of Information and Organizational Sciences*, vol. 34, no. 1, pp. 81–88, 2010.
[41] Q. Gong, Y. Chen, X. He, Z. Zhuang, T. Wang, H. Huang, X. Wang, and X. Fu, DeepScan: Exploiting deep learning for malicious account detection in location-based social networks, *IEEE Communications Magazine*, vol. 56, no. 11, pp. 21–27, 2018.
[42] Q. Gong, Y. Chen, X. He, Y. Xiao, P. Hui, X. Wang, and X. Fu, Cross-site prediction on social influence for cold-start users in online social networks, *ACM Transactions on the Web*, vol. 15, no. 2, pp. 1–23, 2021.
[43] X. He, Q. Gong, Y. Chen, Y. Zhang, X. Wang, and X. Fu, DatingSec: Detecting malicious accounts in dating apps using a content-based attention network, *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 5, pp. 2193–2208, 2021.
[44] T. K. Balaji, C. S. R. Annavarapu, and A. Bablani, Machine learning algorithms for social media analysis: A survey, *Computer Science Review*, vol. 40, p. 100395, 2021.
[45] J. Grimmer, M. E. Roberts, and B. M. Stewart, Machine learning for social science: An agnostic approach, *Annual Review of Political Science*, vol. 24, pp. 395–419, 2021.
[46] C. Li, S. Zhang, and X. Li, Can multiple social ties help improve human location prediction? *Physica A*: *Statistical Mechanics and its Applications*, vol. 525, pp. 1276–1288, 2019.
[47] Y. Liu, X. Shi, L. Pierce, and X. Ren, Characterizing and forecasting user engagement with in-app action graph: A case study of snapchat, in *Proc. 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA, 2019, pp. 2023–2031.
[48] Y. Chen, X. Wu, A. Hu, G. He, and G. Ju, Social prediction: A new research paradigm based on machine learning, *The Journal of Chinese Sociology*, vol. 8, p. 15, 2021.



Jiaqi Wu received the bachelor degree from Fudan University, Shanghai, China in 2022. During his undergraduate study, he worked as a research assistant at the Mobile Systems and Networking (MSN) Group with Prof. Yang Chen. His research interests include human-computer interaction and social computing.

Bodian Ye is currently pursuing the master degree at the School of Computer Science, Fudan University, Shanghai, China. She received the bachelor degree from Hangzhou Normal University in 2022. Now she works as a research assistant at the Mobile Systems and Networking (MSN) Group with Prof. Yang Chen. Her research interests include social networks and data mining.

Qingyuan Gong received the PhD degree in computer science from Fudan University in 2020, and worked as a postdoctoral researcher in the School of Computer Science, Fudan University from 2020 to 2022. Now she is a pre-tenure associate professor within the Research Institute of Intelligent Complex Systems at Fudan University. Her research interests include network security, user behavior analysis, and computational social systems. She published referred papers in *IEEE Communications Magazine*, *ACM TWEB*, *IEEE TDSC*, *IEEE TMC*, *IEEE TKDE*, *Springer WWW Journal*, *ACM CIKM*, and *ICPP*. She was a visiting student at University of Göttingen in 2015 and 2019, and also University of Chicago in 2018.

Atte Oksanen is a professor of social psychology at Tampere University (Finland) and the leader of Emerging Technologies Lab. His research focuses on emerging technologies, new modes of social interaction, and deviant behavior online. He has led major cross-national projects over his career and his recent research projects investigate the increasing role of intelligent technologies in everyday life. He has published almost 200 peer reviewed articles and has over 300 publications to his name. His publications are published in journals such as *Criminology*, *Addiction*, *Pediatrics*, and *International Journal of Information Management*, *New Media & Society*, *Social Science and Medicine*, and *Computers in Human Behavior*. He also has monographs published by Routledge and Springer.

Cong Li received the PhD degree in intelligent systems from Delft University of Technology (TUDelft), Delft, the Netherlands in 2014. She is currently an associate professor in the School of Information Science and Technology, Fudan University, where she involves in complex network theory and applications. Her research focuses on analysis and modeling of complex networks, including network properties, dynamic processes, network of networks, etc. Her work on these subjects include 2 (co-)authored research monographs, 1 book chapter, and more than 40 international journal papers and conference papers.

Jingjing Qu is an associate professor at Shanghai Artificial Intelligence Laboratory. She received the PhD degree in entrepreneurship from University of Essex degree, UK in 2018. Her research interests include artificial intelligence ethics and governance, artificial intelligence innovation, social media analysis, and social network.

Felicia F. Tian is an associate professor of sociology at the School of Social Development and Public Policy, Fudan University. She received the BS degree in information management and information system from Renmin University of China in 2007 and the PhD degree in sociology from Duke University, USA in 2014. Her research interests include social networks and social capital, community studies, and social stratification and inequality. She serves as the board member of Social Stratification and Mobility Section in Chinese Sociological Association.

Xiang Li received the BS and PhD degrees in control theory and control engineering from Nankai University, China in 1997 and 2002, respectively. He was with City University of Hong Kong, Int. University Bremen, Shanghai Jiao Tong University, and Fudan University as a postdoctoral researcher, Humboldt research fellow, associate professor, professor/distinguished professor in 2002–2004, 2005–2006, 2004–2007, and 2008–2021, respectively. Currently, he is the founding director of the Institute of Complex Networks and Intelligent Systems, Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai, China. He served/serves as an associate editor for the *IEEE Transactions on Circuits and Systems-I*: *Regular Papers* (2010–2015), *Research*, the *Journal of Complex Networks*, and the *IEEE Circuits and Systems Society Newsletter*, and the associate editor (2018–2021) and area editor (since 2022) of the *IEEE Transactions on Network Science and Engineering*. His main research interests cover network science and systems control in both theory and applications. He has (co-)authored 6 research monographs, 7 book chapters, more than 150 peer-refereed journal publications, and 100+ indexed conference papers. He received the IEEE Guillemin-Cauer Best Transactions Paper Award from the IEEE Circuits and Systems Society in 2005, Shanghai Natural Science Award (1st class) in 2008, Shanghai Science and Technology Young Talents Award in 2010, National Science Foundation for Distinguished Young Scholar of China in 2014, National Natural Science Award of China (2nd class) in 2015, Ten Thousand Talent Program of China in 2017, TCCT CHEN Han-Fu Award of Chinese Automation Association in 2019, the Excellent Editor Award of the *IEEE Trans. Network Science and Engineering* in 2021, among other awards and honors.

Yang Chen received the BS and PhD degrees from Tsinghua University in 2004 and 2009, respectively. He is an associate professor within the School of Computer Science at Fudan University and leads the Mobile Systems and Networking (MSN) group at Fudan University. From April 2011 to September 2014, he was a postdoctoral associate at the Department of Computer Science, Duke University, USA, where he served as a senior personnel in the NSF MobilityFirst project. From September 2009 to April 2011, he has been a research associate and the deputy head of Computer Networks Group, Institute of Computer Science, University of Göttingen, Germany. He visited Stanford University (in 2007) and Microsoft Research Asia (2006–2008) as a visiting student. He was a Nokia visiting professor at Aalto University in 2019. His research interests include online social networks, Internet architecture, and mobile computing. He serves as an associate editor-in-chief of the *Journal of Social Computing*, an associate editor of *Computer Communications*, and an editorial board member of the *Transactions on Emerging Telecommunications Technologies* (*ETT*). He served as an OC/TPC member for many international conferences, including SOSP, SIGCOMM, WWW, IJCAI, AAAI, ECAI, DASFAA, IWQoS, ICCCN, GLOBECOM, and ICC. He is a senior member of the IEEE, ACM, and CCF.