IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 4, 2018, accepted December 12, 2018, date of publication December 18, 2018, date of current version January 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2887341

INVITED PAPER

Human Interactive Behavior: A Bibliographic Review

XIANGJIE KONG, (Senior Member, IEEE), KAI MA, SHEN HOU, DI SHANG, AND FENG XIA, (Senior Member, IEEE)
Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, School of Software, Dalian University of Technology, Dalian 116620, China

Corresponding author: Feng Xia (f.xia@ieee.org)

This work was supported in part by the National Natural Science Foundation of China under Grant 61572106, in part by the Dalian Science and Technology Innovation Fund under Grant 2018J12GX048, and in part by the Fundamental Research Funds for the Central Universities under Grant DUT18JC09.

ABSTRACT As a common human behavior, interaction is everywhere in human life. With the rise of big data and human–computer interaction in the 21st century, more and more researchers from different industries and disciplines pay great attention to the human interactive behavior research. From the perspective of computer science, scholars try to use computer technology to make research more meaningful. To the best of our knowledge, this paper is the first study to investigate the potential rules of human interactive behavior in the view of computer science, based on 16 top-tier journals of human interactive behavior from Microsoft Academic Graph dataset. We put forward a topic extraction and clustering model based on word2vec to infer key topics, which can be widely used in different fields of research. We find that the growth of human interactive behavior is in an uptrend on the whole. Besides, the cooperative relationship between authors and countries/regions is closer over time. We also make the mensurable evolution analysis of topics by a statistical method. Some topics are hot all the time, while some are unpopular as time goes by. Finally, we do rankings in the field of human behavior research from a new perspective. All these findings help researchers observe potential patterns and the topic evolution in half a century, which may shed dazzling light on the exploration of human interactive behavior.

INDEX TERMS Human interactive behavior, computer science, topic modeling, bibliographic review and topic trend.

I. INTRODUCTION

Human behavior is an unfailing research direction for a long time. Not a single subject, but an interdisciplinary product, human behavior is studied by the researchers from fields, including but not limited to biology, physic, psychology, sociology and computer science. However, there is still no quite accurate and authoritative definition of human behavior at present. A conclusive definition of human behavior is that it is the response of individuals or groups of humans to internal and external stimuli [1].

As a key area of human behavior research in both science and engineering, interactive behavior appears in every corner of human life. In the process of human interactive behavior research, researchers mainly focus on three aspects. One is aiming at interactive behavior among human, which means the interaction among persons in human groups, such as collaboration and competition. The next is concentrating on the impact of interaction on society, in terms of environment, economy, and politics etc. And the last one is devoted to interaction between human and machine, such as human-computer interaction and artificial intelligence.

In the 21st century, with the advances of technology, computer has been one of the most frequently-used tools both in work and life, making human interactive behavior research in view of computer science receive much attention. It is significantly necessary and meritorious to study human interactive behavior systematically, for it can reflect changes in universal behavior of human. What is more, in the big-data era, researchers of human communication behavior mainly focus on analyzing the information from the massive data, such as travel trajectory data, mobile data and spatio-temporal data. That gives the study of human interactive behavior a new meaning.

The bibliographic analysis is a usual research method to structure the academic framework and explore the topic evolution from an extensive number of publications. And we can get the nature of one field or discipline and change of the research emphasis by the quantitative analysis.

Some online academic systems, such as Google Scholar and Microsoft Academic Services, can provide platforms for users to acquire the useful information of scholars and papers. Besides, these systems also offer interfaces for researchers to download the academic datasets, which include but not limited to the names of papers and authors, the published year, and the references. For example, in the field of physics, Perc [2] identifies all unique words and phrases, and obtain quantifiable insights into the trends of physics discovery from the end of the 19th century to today. In addition, Kuhn *et al.* [3] do the analysis of memes in the scientific literature, and reveal that memes are governed by a surprisingly simple relationship between frequency of occurrence and the degree to which they propagate along the citation graph.

With the help of these online systems and datasets, researchers do a lot of related work about the quantitative document analysis in all kinds of fields. Liu *et al.* [4] quantify the temporal trends at the topic level and discover the inner connection among these topics in the field of artificial intelligence in the 21st century, by using publication metadata extracted from 9 top-tier journals and 12 top-tier conferences of this discipline. Topic evolution is not only a key factor but also a difficult point in the bibliographic analysis. In the field of transportation research, Sun and Yin [5] apply a latent Dirichlet allocation model on article abstracts to infer the key topics from the articles published in 22 leading transportation journals from 1990 to 2015. What is more, scientific collaboration among scholars is also a popular topic at present. Collaboration is not a single behavior. On the contrary, there are many important patterns in collaboration. Viana *et al.* [6] find the intermittency of collaborations in the research, which shows that researchers tend to exhibit different patterns of collaborations as far as the intermittency is concerned. Besides, Hennemann *et al.* [7] do the research which reveal a strongly decreasing relation between spatial distance and the probability of co-authoring a scientific publication. Xia *et al.* [8] examine the background and state of the art of big scholarly data, and analyze the relationships among scholars based on the big scholarly data. And Sun and Rahwan [9] construct the coauthorship network in transportation research from 22 transportation journals. Besides, Chen *et al.* [10] examine how research topics evolve by analyzing the topic trends, evolving dynamics, and semantic word shifts in the information retrieval domain. Yan [11] uses topic continuity and popularity to examine topic dynamic characteristics. Fister, Jr. *et al.* [12] construct citation network of scientific publishing, and discover the citation cartels using the modern semantic web tools for manipulating the knowledge on the Internet.

Although more and more efforts based on the theory and technology of bibliographic analysis are put forward, the analysis research of human interactive behavior is still undrafted. So far, there is little attention being paid to a statistical analysis to portray the field of human interactive behavior at the beginning of the 21st Century, especially in the view of computer science. Thus it is the time to understand the inner structure and topic evolution over time and region through the quantitative analysis of this area.

To meet the need of human interactive behavior subject research, we use the technology of bibliographic analysis to research the evolution of human interactive behavior in four levels. Firstly, we construct the topic extraction and clustering model named TEC, which is based on word2vec. Secondly, we survey the evolution in this field from the change in number of papers and authors. Thirdly, we examine the research related to the popular topic of this field from different aspects. Fourthly, we focus on the relationship among the papers, and build the network of topics co-presence. The scholarly dataset of our study consists of 21,372 publications and 1,526,096 citations from 16 top-tier journals. The main findings include:

- We construct the topic extraction and clustering (TEC) model based on the word2vec, which is widely used in the deep learning. We input the origin datasets, and combine similar keywords into a cluster, thus get 753 different themes clusters. The topics in one cluster can be regarded as same topic. Moreover, this model also can be used to do analysis beyond human behavior research.
- We find that the growth of human interactive behavior shows an uptrend on the whole, in terms of the number of publications, authors, references and citations. Besides, the cooperative relationships between authors and countries/regions are closer compared with relationships many years ago. It can be reflected by the increasing number of citations, co-authors from all over the world and references.
- We make the mensurable evolution analysis of topics by statistical method, and obtain top 25 popular topics of human interactive behavior, including "Psychology", "use study", "social identity" and so on. Besides, we get the dynamic change of hot topics by years and their popularity degree every 6 years. Some topics are hot all the time, while some are unpopular as time goes by.
- Some rankings are done to screen out the excellent authors, publications and countries/regions in the field of human interactive behavior research. We rank the authors by the number of citations per paper, while publications by the number of total citations, and countries/regions by the citations per paper.

In summary, we research the development of human interactive behavior research in 21st century. By the TEC model, we cluster the similar topics and do the quantitative analysis. We find that the research of human interactive behavior is becoming more and more multiply and popular with time going by. Researchers from different areas join the teams of human interactive behavior researching and make a great contribution to the society. Our results not only show the development of human interactive behavior, but also explore the growth of the number of journals, publications, authors and cooperations. Besides, some important changes are also displayed and discussed. For society, our research can help governments to make decisions of scientific development, and allocate the resources rationally. For researchers, our results will give them direction and inspiration to do more meaningful researches of human behavior. Last but not least, the model we construct is also suitable for other areas of research. In this research, we provide a valuable tool for researchers to make more informed decisions. We also hope this work can stimulate more discussion on the state of publishing in human interactive behavior research.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 provides methodologies and models we use to do analysis. We get the results and make bibliographic analysis based on our results in Section 4. Finally, Section 5 concludes the paper with some directions in future.

II. RELATED WORK

In this section, we introduce the relevant works about human interactive behavior and word2vec models used to do topic clustering.

A. HUMAN BEHAVIOUR

Human behavior is always a transdisciplinary and diversified theme, and the research of behavior change and prediction is quite widespread. As for complex and transdisciplinary science, researchers do a lot to make the research more efficient and useful. Helbing *et al.* [13] highlight that the complex and often counter-intuitive behavior of social systems and their macro-level collective dynamics can be better understood by means of complexity science. Besides, Perc [14] find that cumulative advantage and success-breads-success also both describe the fact that advantage tends to beget further advantage which is at the heart of self-organization across social and natural sciences.

There are many tools and techniques used to explore the human behavior. Some scholars do this research based on the human motion [15], they use Bayesian network to estimate motion parameters and use Markov chain models to process time data. With the popularity of mobile devices, such as smart phones and portable accessories, mobile data becomes significantly important information to record and analyze human behavior [16]. Pejovic and Musolesi [17] present that therapists can understand the behavior of patients and intervene their behavior change by the mobile phones, which are equipped with a range of sensors and carried by users at any time.

In the field of computer science, scholars put forward and design many algorithms to analyze human behavior. In the early days, model-fitting approach is one of the most outstanding approaches to analyze human behavior [18]. With the rise of machine learning, researchers often combine machine learning with human behavior. Shteingart and Loewenstein [19] use model-free reinforcement learning (RL) to explain the human behavior. As for the abnormal human behavior, the mining algorithm for intelligent home dataset [20] is proposed to explore new and changing behaviors, especially the abnormality. Other than the area of computer science, such as agriculture, environment, and economics, researchers use their own theory and practice to do the project research and policy design [21]. For example, there is a study [22], exploring the link between biological weather variables and the behavior of tourists.

B. HUMAN INTERACTIVE BEHAVIOUR

Understanding and recognition of human behavior is also a significant direction. Some researches are based on some parts of the human own body activities. An efficient real-time gesture recognition mechanism based on hand motion trajectory and hidden Markov model classifier is developed to help analyze the human behavior and develop the human-computer interactive control system [23]. Besides, Wu *et al.* [24] argue that the human visual system can receive both RGB and depth information and make accurate judgments about human behavior. What is more, with the maturity of sensing technology and pervasive computing technology, extensive research is being carried out to understand human behavior using different sensing technologies [25]–[27].

In the age of information, privacy is an import issue of human behavior. Acquisti *et al.* [28] summarize and draw connections between diverse streams of empirical research on privacy behavior.

Human-computer interaction is also a topic worth focusing on, because there are a large number of computer users which make the human-machine operation be a crucial human behavior. Azvine and Wobcke [29] consider the abstract features of the human-computer interaction system, which are needed to generate intelligent human behavior. Many studies attempt to model and identify human behavior through human-computer interaction motion analysis. Jaouedi *et al.* [30] pay attention to human behavior analysis in video scenarios. In their work, they explain human behavior identification through K Nearest Neighbors approach. Human can not live without transportation, so vehicle navigation system is an important human-machine interface. By evaluating vehicle navigation systems [31], researchers and decision-makers can optimize the human travel behavior. Besides, more and more attention is given to scenarios in which human-machine cooperation is beneficial but non-trivial. Crandall *et al.* [32] indicate that general human-machine cooperation is achievable using a non-trivial, but ultimately simple, set of algorithmic mechanisms.

Apart from the individual behavior, social behaviour gets a lot of attention too. Human interaction in social networks is dominated by emergent statistical laws such as non-trivial correlations and temporal clustering [33]. Through the research of social human behaviour, We can analyze the current situation of the whole society and the trendency of society development. Carroll *et al.* [34] combine the community computing and human behaviour, which has much impact on society and daily collective life. Online communities are becoming a part of the lives of Internet users [35]. Bishop [36] builds a framework for human-computer interaction to explain human desires and behaviour. Social interaction is an important part of social behaviour. Sofia *et al.* [37] discuss the role and applicability of intelligent data captured in non-invasive ways in reasoning and contextualization of human behaviour and interactions.

C. WORD2VEC

One of the main computational and scientific challenges in the modern age is to extract useful information from unstructured texts. Topic models are one popular machine-learning approach that infers the latent topical structure of a collection of documents [38]. Li *et al.* [39] present a novel framework to compute the correlated sub-networks from a large information network, and to extract the influential topics from each correlated network. Word2vec is a group of related models used to generate word vectors, which is put forward firstly by Mikolov [40]. These models of word2vec are shallow and double-layered neural networks used to train the reconstruction of linguistic word text. The network is represented by words and needs to guess the input words in adjacent positions. After the training, the word2vec model can be used to map each word to a vector, which can be used to represent the relationship between word and word. This vector is the hidden layer of neural network.

With the development of computer applications, natural language processing has received high attention. The application requirements of machine translation, speech recognition and information retrieval have raised higher and higher requirements on the computer's natural language processing ability. In order for computers to be able to handle natural languages, natural languages need to be modeled first. Natural language modeling methods have undergone a transformation from rule-based methods to statistical methods.

The natural language model derived from statistical modeling is called statistical language model. There are many statistical language modeling techniques, including n-gram, neural network, log-linear model and so on. In the process of modeling natural language, there are some challenges, such as dimension disaster, word similarity, model generalization ability and model performance. Thus, finding solutions to the above problems is an intrinsic motivation for the development of statistical language models. Google presents the word2vec in 2013, which is a software tool for training word vector. Based on the given corpus, word2vec can express a word into vector form quickly and effectively through the optimized training model, providing a new tool for the application research in the field of natural language processing.

III. METHODOLOGY

Fig. 1 shows the framework of our TEC model. The whole schema can be divided into three parts: data pre-processing module based on data in MAG, similar topic clustering module based on word2vec, and topic trends analysis module based on statistical method.

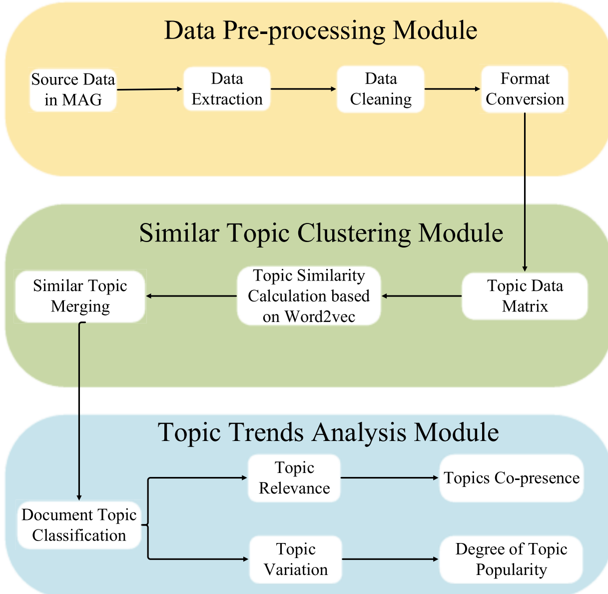

FIGURE 1. Framework of TEC.

In this section, we first introduce the situation of data we use, including the origin of dataset and data processing. Then we emphasize some parameters and computing methods used in our analysis. Finally, we describe the method of analyzing inner structure of human interactive behaviour by the topic evolution.

A. DATA PRE-PROCESSING MODULE BASED ON DATA IN MAG

Before the study begins, we should identify which papers and journals belong to the field of human interactive behaviour. To select the appropriate papers, we first divide the human interactive behaviour into three categories based on computer science perspective: human-human behaviour, human-society behavior, and human-computer behaviour. Human-human behaviour means the behaviour among humans, which contains the evolution of human behaviour, human decision processes, interactive technology and so on. As for human-society behaviour, the research raise the interactive behavior to the social level, mainly aiming at the social ethics, policy-making, and information interaction among cities or regions. Human-computer behaviour research is especially valuable to researchers in the field of computers, this type of interactive behaviour includes the human-computer interaction, artificial intelligence, and machine learning.

Based on the classification method above, we decide to choose the top journals which respectively aims at above behaviour research based on the journal aims and scopes. It is worth mentioning that, we do not choose the papers from conferences, because different conferences have different holding cycles, which may result in the fluctuation of publications and prevent us from reaching a correct conclusion. TABLE 1 shows the research directions of the journals we use.

TABLE 1. The classification of chosen journals based on research direction.

	Human-human behaviour	Human-society behaviour	Human-computer behaviour
1	Behaviour &Information Technology	IEEE Transactions on Computational Social Systems	ACM Transactions on Computer-Human Interaction
2	Evolution and Human Behavior	Journal of Information, Communication and Ethics in Society	Computers in Human Behavior
3	Multivariate Behavioral Research	IEEE Technology and Society Magazine	IEEE Transactions on Human-Machine Systems
4	Organizational Behavior and Human Decision Processes	Journal for The Theory of Social Behaviour	International Journal of Human Computer Studies
5	International Journal of Information and Communication Technology	International Journal of Social and Humanistic Computing	Human-computer interaction
6			Journal of the Association for Information Science and Technology

We obtain the data from the Microsoft Academic Graph (MAG), which is a heterogeneous graph containing scientific publication records, citation relationships among those publications, as well as authors, institutions, journals, conferences, and fields of study. To research the topic evolution of human behaviour and build the collaboration network of human behaviour, we select the papers from 16 journals, which are the top or famous journals closely related to human, behaviour, society or human-computer interaction. For the journals in computer field, we choose these journals which all recommended by the Chinese Computer Society (CCF) and the Australian Institute of Computing Research and Education (CORE). For the journals not in computer field, the journals we choose have a relatively high SCI (Science Citation Index) impact factor in there own academic sectors.

TABLE 2 lists the journals and their basic statistics including the total number of papers, the total citations of these papers, the total number of unique authors, the average number of authors per paper, the average number of published papers per author, and the average number of citations per paper. TABLE 3 shows the significant object schema in MAG dataset, every paper is stored as an object of JSON.

B. SIMILAR TOPIC CLUSTERING MODULE BASED ON WORD2VEC

For a paper, there are many different topics. In the MAG dataset used in this study, "keywords" and "fos" attributes of each paper are extracted as the topics. Because many topics have the same word meaning but different form, it is necessary to do the classification and clustering of similar topics.

There are many subject extraction and similarity calculation methods in the research. Chen *et al.* [41] present a novel approach for extracting hot topics from disparate sets of textual documents published in a given time period. Besides, Silva *et al.* [42] develop a new method which extracts keywords from abstracts and the citation network of papers in a topic, to generate automated taxonomies and visualizations of a scientific field.

As for the topic similarity calculation, there are many popular methods. Gomaa and Fahmy [43] do a survey of text similarity approaches. In this survey, three text similarity approaches are discussed: String-based, Corpus-based and Knowledge-based similarities. Especially, we use the Corpus-based text similarity approach based on word2vec in our research. Word2vec is the most popular word similarity analysis model. The advantage of this model is that it is dynamic, and the results of word similarity analysis are all related to the corpus provided, and the results obtained from different corpus are also different.

Studied by the researchers from fields, Human behaviour is not a single subject, but an interdisciplinary area. How to obtain and classify the topic of papers is a crucial adjective problem. In MAG datasets, every paper has its own keywords and field of study, which are stored as TABLE 3 shows. Combining these two characteristic value can reflect the topic of papers. We extract the information of keywords and field of study from every paper, and filtrate the repetitive topics which appear both in keywords and field of study. After that, we obtain the topics of every paper. Then we select the topics which occur more than 10 times to make the analysis more efficient, and get 3,973 topics finally.

Algorithm 1 Framework of Similar Topic Clustering Module.

```
Require: T_matrix: the list of topics from the papers;
Ensure: S_ij: the similarity index between two topics;
         C_matrix: the matrix of similar topic communities;
Extracting the topics which occurs number > 10 from
  T_matrix;
Training word2vec model, with help of data in google
  corpus;
for all i, j in T_matrix do
    calculate S_ij;
    if S_ij > 0.6 then
        Regard topic i and j as one community member;
        Put topic i and j into C_matrix;
    end if
end for
Choose the representative topics of one community;
```

Before we do the analysis of topic trends, we should divide these topics into different communities. The similar topic clustering module we use is based on word2vec. First, we train the word2vec model by using the google corpus.

TABLE 2. Statistics of the choosen journals.

ID	journal Name	No. of Publication	No. of Authors	No. of Citations	Authors per Paper	Papers per Author	Citations per Paper
1	Behaviour & Information Technology	1693	3260	96684	2.47	1.27	57.11
2	Computers in Human Behavior	5379	11145	324387	2.85	1.39	60.31
3	IEEE Transactions on Human-Machine Systems	379	1248	13988	3.73	1.13	36.91
4	IEEE Transactions on Computational Social Systems	51	174	1450	3.93	1.08	28.43
5	Journal of Information, Communication and Ethics in Society	311	424	10291	1.67	1.35	33.09
6	Evolution and Human Behavior	1086	1957	87067	3.07	1.66	80.17
7	Human-computer interaction	621	1149	72430	2.50	1.37	116.63
8	IEEE Technology and Society Magazine	1168	1327	35413	1.66	1.36	30.32
9	Journal for The Theory of Social Behaviour	970	860	56803	1.28	1.43	58.56
10	Multivariate Behavioral Research	1710	2072	145539	2.06	1.69	85.11
11	Organizational Behavior and Human Decision Processes	1872	2778	293998	2.46	1.66	157.05
12	Journal of the Association for Information Science and Technology	3925	4706	216993	2.38	1.84	55.28
13	International Journal of Social and Humanistic Computing	37	105	1655	3.24	1.09	44.73
14	International Journal of Information and Communication Technology	304	719	6001	2.53	1.08	19.74
15	ACM Transactions on Computer-Human Interaction	510	1672	55655	3.28	1.21	33.29
16	International Journal of Human Computer Studies	1356	4093	107742	3.02	1.81	79.46

TABLE 3. Significant object schema in mag dataset.

No.	Attribute	Type	Description	Example
1	id	string	MAG ID	1e72503c-6df8-4072-84ee-caf1730bc2a1
2	title	string	paper title	Evolution of information systems in organizations
3	authors.name	string	author name	Charlie C. Chen
4	author.org	string	author institution	Department of Psychology, University of Bologna, Cesena, FC, Italy
5	venue	string	paper venue	Behaviour & Information Technology
6	abstract	string	abstract	Previous studies have shown that hypertext users generate a mental representation...
7	year	int	published year	2016
8	keywords	list of strings	keywords	["ease of use", "usefulness self efficacy", "elderly people", "tam"]
9	fos	list of strings	fields of study	["Psychology", "Three-dimensional space", "Computer vision", "Artificial intelligence", "Communication"]
10	n_citation	int	number of citation	50
11	reference	list of strings	citing paper's ID	["06e92b21-5a5b-4d2e-9138-bc2975807959", ..."2e5ebde6-2494-47bb-8733-94e2e4f74464"]

Then, we input the topics data we get to the model, and calculate the similarity index S_{ij} of each two topics.S_{ij} is a decimal range from 0 to 1. Bigger it is, more similar two topics are. If $S_{ij} > 0.6$, we regard these two topics as the members of same community.

After calculate S_{ij}, we divide the 552 communities, where the S_{ij} of each two members is bigger than 0.6. We use the topic which occurs the most times in one community as the representative of the community. Algorithm. 1 is the pseudocode of Similar Topic Clustering Module, and we can get the set of topic communities.

C. TOPIC TRENDS ANALYSIS MODULE BASED ON STATISTICAL METHODS

In this subsection, we introduce our topic trends analysis module of human interactive behaviour. In this module, We use various analysis methods to characterize internal structure of research network and topic trends from different angles.

1) PARAMETERS

Firstly, we describe some parameters used to quantize the index of topic evolution and academic popularity.

- The number of authors per paper:
  We denote N_{AA} as the average number of authors in one paper for all articles:
  $$N_{AA} = \frac{\sum_{p \in P} |a_p|}{|P|} \tag{1}$$
  where $|P|$ is the number of papers in the journals, and $|a_p|$ means the number of authors in paper p.

- The number of papers per author:
  We denote N_{AP} as the average number of papers published by one author for all scholars:
$$N_{AP} = \frac{\sum_{a \in A} |a_a|}{|A|} \tag{2}$$
  where $|A|$ is the number of authors in the journals, and $|a_a|$ means the number of papers of author a.
- The number of papers per author:
  We denote N_{AC} as the average number of citations in one paper for all articles:
$$N_{AC} = \frac{\sum_{p \in P} |c_p|}{|P|} \tag{3}$$
  where $|P|$ is the number of papers in the journals, and $|c_p|$ means the number of citations in paper p.

2) PAPER TOPIC RELEVANCE

The topics of papers are not isolated. On the contrary, many topics are related to each other, which we can structure the topic network based on. In order to research the paper topic relevance, we calculate the probability of topic A's occurrence on the condition that topic B's occurrence. We denote P_A as the probability of topic A's occurrence:
$$P_A = \frac{|N_A|}{|N|} \tag{4}$$
where $|N|$ means the total number of papers in journals, and $|N_A|$ means the number of papers which have the topic A.

Analogously, we can compute P_B:
$$P_B = \frac{|N_B|}{|N|} \tag{5}$$
Then, we define P_{AB} as the probability that topic A and B occur in the same time:
$$P_{AB} = \frac{|N_{AB}|}{|N|} \tag{6}$$
where $|N|$ means the total number of papers in journals, and $|N_{AB}|$ means the number of papers which have both the topic A and B.

Finally, we calculate $P(A|B)$, which means probability that A appears under the condition that B appears based on Equation. 4 to 6:
$$P(A|B) = \frac{|P_{AB}|}{|P_B|} \tag{7}$$

3) TOPIC VARIATION

The popular topics of papers are constantly changing. We try to analyze the evolution of topic based on time, country/region, and journals respectively [5].

- To research the topic distribution over time, we denote $\theta_k^{[t]}$ as the proportion of topic k in all journals at year t:
$$\theta_k^{[t]} = \frac{|N_{kt}|}{|N_t|} \tag{8}$$
  where $|N_t|$ means the total number of all topic appearance in year t, and $|N_{kt}|$ presents the number of topic k appearance in year t.
- Similarly, we define $\theta_k^{j[t]}$ as the proportion of topic k in journal j at year t. So that we can quantify the topic evolution of every journals:
$$\theta_k^{j[t]} = \frac{|N_{jkt}|}{|N_{jt}|} \tag{9}$$
  where $|N_{jt}|$ is on behalf of the total number of topic appearance of journal j in year t, and $|N_{jkt}|$ presents the number of topic k appearance of journal j in year t.
- Because every author has the affiliation, we can probe the topic evariation of different countries or regions based on it. We use $\theta_k^{c[t]}$ as the proportion of topic k in country/region c at year t:
$$\theta_k^{c[t]} = \frac{|N_{ckt}|}{|c_{jt}|} \tag{10}$$
  where $|N_{ct}|$ means the total number of topic appearance from country/region c in year t, and $|N_{ckt}|$ presents the number of topic k appearance from country/region c in year t.

4) DEGREE OF TOPIC POPULARITY

With the development of the times, some topics of computer based human behaviour are hotter and hotter, while some are unpopular. It is valuable to define a parameter to describe the changing degree of topic popularity. We use R_k to reflect the degree of topic k popularity growth between three-year intervals. For example, we can get the change from 2012 to 2017:
$$R_k = \frac{\sum_{t=2015}^{2017} \theta_k^{[t]}}{\sum_{t=2012}^{2014} \theta_k^{[t]}} \tag{11}$$
According to the Equation. 11, we can obtain the degree of topic k popularity growth. If $R_k < 1$, that means topic k becomes less popular from 2012 to 2017, and vice versa.

5) TOPICS CO-PRESENCE

Topics co-presence describes the condition that two topics appear in one paper at the same time. It can help us research the inner connection between two topics. We focus on the phenomenon of topics co-presence, and construct the network of topics co-presence based on the method showed in [4]. From Equation 4 to 6, we obtain P_A, P_B and P_{AB}, which means the probability of topic A's and B's severally and simultaneously occurrence. Based on these, we use the coefficient of co-presence $C(A, B)$:
$$C(A, B) = \frac{P_{AB}^2}{min\{P_A, P_B\} * mean\{P_A, P_B\}} \tag{12}$$
When $C(A, B) > 0.1$, we consider that topic A and B have the topics co-presence, and add this link to the network of topics co-presence.

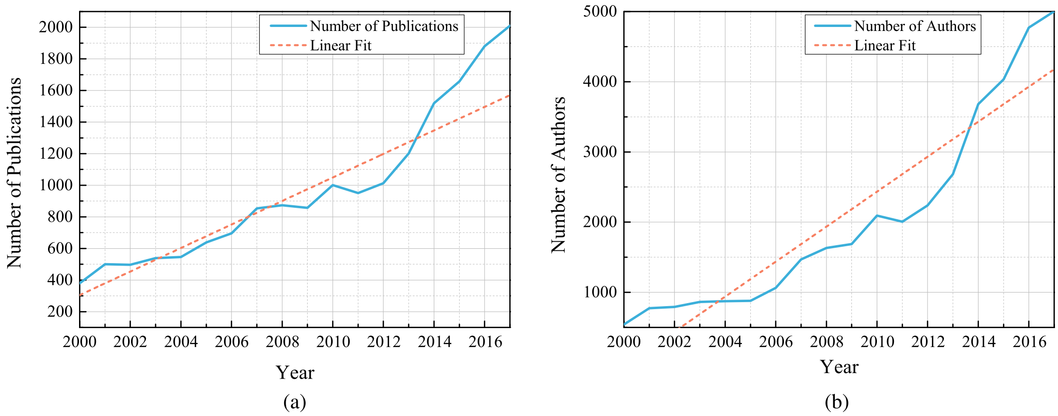

FIGURE 2. The evolution of human behavior papers in the 21 century. (a) The change of the number of publications from 2000 to 2017. (b) The change of the number of authors from 2000 to 2017.

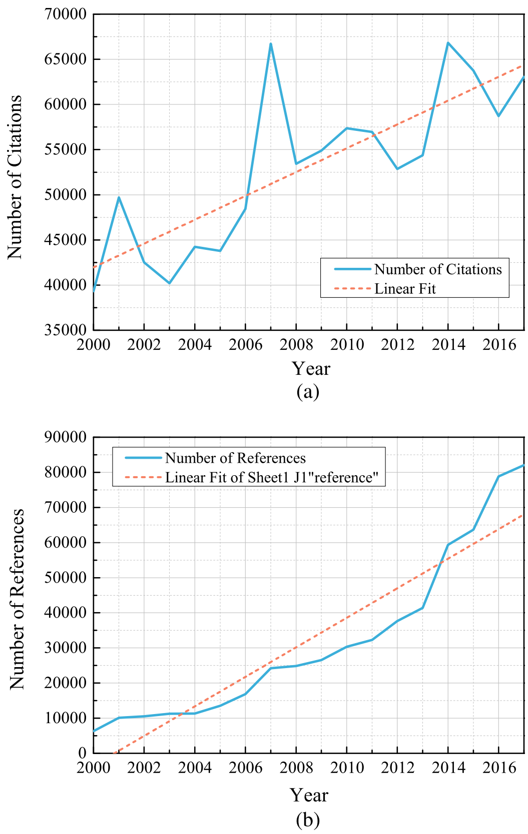

FIGURE 3. The evolution of citations and references. (a) The number of citations from 2000 to 2017. (b) The number of references from 2000 to 2017.

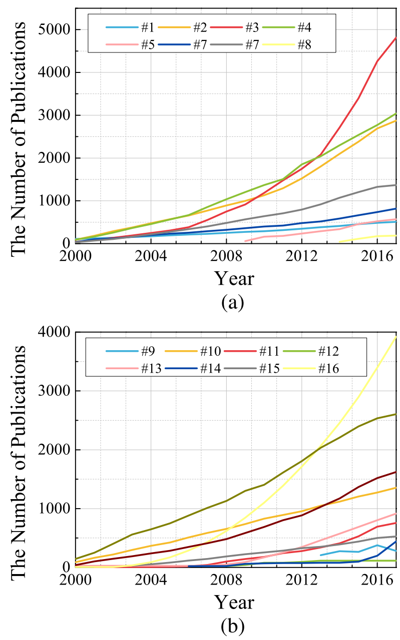

FIGURE 4. The cumulative number of papers of 16 journals in our research. (The name of every journal ID shows in TABLE 2).

IV. RESULTS AND ANALYSIS

In this section, the results of topic trends of human interactive behaviour research are introduced in the following experimental result analysis. The results include the growth of human interactive behaviour research, the evolution analysis of topic, and the ranking of authors/publications/countries.

A. THE GROWTH OF HUMAN INTERACTIVE BEHAVIOUR RESEARCH

To observe the development of human interactive behaviour, we count the number of papers published and authors per year(see in Fig. 2). Fig. 2 reflect the evolution of human interaction behaviour papers in the 21 century. FIg.2(a) shows the change of the number of publications from 2000 to 2017. Through the outcome, we find that human interactive behaviour have gained more and more attention since 2000 in Fig. 2(a), as the growth rate is unprecedentedly high during this period. Although the growth rate decreased slightly later, the speed of rising increased. With the development of computer technology and big data, human interactive behaviour resulted in explosion of publications, and

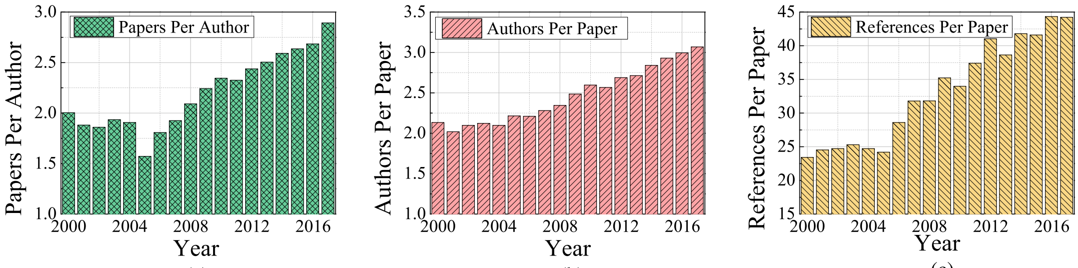

FIGURE 5. The evolution of the per number. (a) The average number of papers per author from 2000 to 2017. (b) The average number of authors per paper from 2000 to 2017. (c) The average number of references per paper from 2000 to 2017.

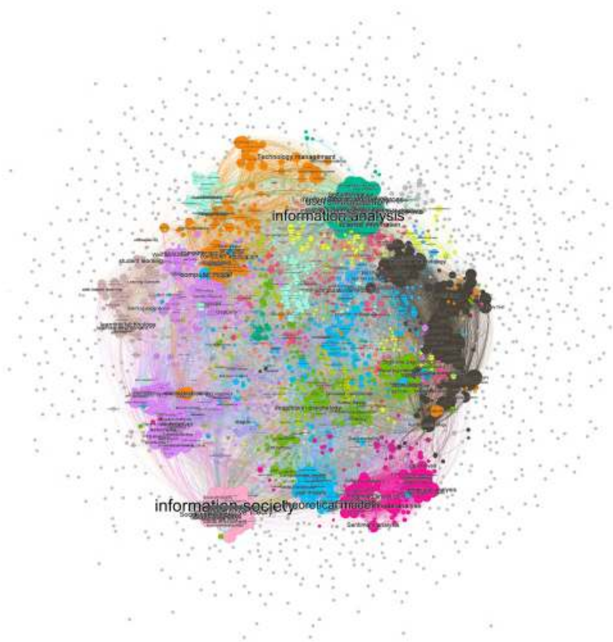

FIGURE 6. Topic communities network.

TABLE 4. Ranking of authors based on the number of publications in the 21st century.

Names	Citations	Publications	H-index
Mike Thelwall	8272	88	79
Loet Leydesdorff	6724	84	91
Gregory D. Abowd	5983	6	80
Linda Argote	5887	6	24
Kristopher J. Preacher	5765	13	35
Andrew F. Hayes	5095	4	24
David P. MacKinnon	4931	10	42
Bernard J. Jansen	4727	26	57
Paul A. Kirschner	4536	31	64
Derek D. Rucker	4412	1	35
Paul Ingram	4145	2	26
Amanda Spink	4114	26	42
Anind K. Dey	3977	7	72
Lutz Bornmann	3715	54	53
Noam Tractinsky	3640	14	26

reached the climax of publications in 2017. It is worth mentioning that the fold line appeared a more obvious increase from 2013 to 2014. The reason is that the journal ''IEEE Transactions on Computational Social Systems'' establishes in 2014(see in Fig. 4(a)). The result indicates researches of human interactive behaviour have increased since 2000. In addition, we count the publication numbers of all 16 chosen journals, and explore the change of the number of publications by years. Fig. 4 shows the cumulative number of papers of 16 journals, which can help us observe the journals' evolution in terms of publication number.

What is the relationship among the growth of papers and that of scientists? To solve this problem, we analyze the number of authors in the database(see in Fig. 2(b)) and find that the growth of authors have the same trend as that of publications. Fig. 2(b) shows the change of the number of authors from 2000 to 2017. What is more, from Fig. 2(b), we find that the number of authors is several times as big as the number of publications. According to the statics, we can get the conclusion that the increase of human interactive behaviour publications may be driven by the increasing number of authors.

We also do a explore the relation between authors and publications, which are showed in Fig. 5), and observe that the average number of papers per author have an overall rising trend from 2000 to 2017(see in Fig. 5(a)), and there is also a rise trend of the average number of authors per paper(see in Fig. 5(b)). Each author has about 3 publications the last two years, and there are about 3 authors collaborate to write papers together at average. It suggests that collaboration among scientists is more and more common in this field, and the average productivity is becoming better and better over the development of human interactive behaviour.

The number of citations can reflect the importance of the papers and journals. Fig. 3 shows the evolution of citations and references. We display the change of the number of citations over years and its polynomial fits in Fig. 3(a). It shows that the overall trend of citations in the field of computer based human behaviour has increased between 2000 and 2017. But the rising trend is not obvious, even a little messy, which shows the papers publication after 2000 have

FIGURE 7. Top 25 popular topics and their relevance. (a) Psychology. (b) Computer. (c) WWW. (d) Knowledge Management. (e) Data Mining. (f) Cognition. (g) Learning. (h) Social Science. (i) Simulation. (j) Communication. (k) Information Retriev-al. (l) Multimedia. (m) Theory. (n) Empirical Research. (o) Human. (p) Internet. (q) Law. (r) Statistics. (s) System Information. (t) Models. (u) Technology. (v) Correlation Analysis. (w) User. (x) Emotion. (y) Social Identity.

higher level compared with the 20th century. Fig. 3(b) shows the change of the number of references in the 21 century. The curve in Fig. 3(b) shows that the number of references increase significantly and rapidly. Observing the number of publications with the Fig. 2, we can find that the number of references per paper is increasing from 2000 to 2017 (see in Fig. 5(c)). That shows the scholars focus on the others work in their own research, and the papers are more and more rigorous and trustworthy.

B. THE EVOLUTION ANALYSIS OF TOPIC

The research of publication topic evolution is valuable and important. we can obtain some hot topics at present by the study and do targeted research and learning. By the similar topic clustering module we mentioned above, we divide the topics into 552 communities, in which topics have similar meanings. Fig. 6 shows the topic communities network we construct based on the TEC model. As show in Fig. 6, topic nodes in one community have the same color, and the distance

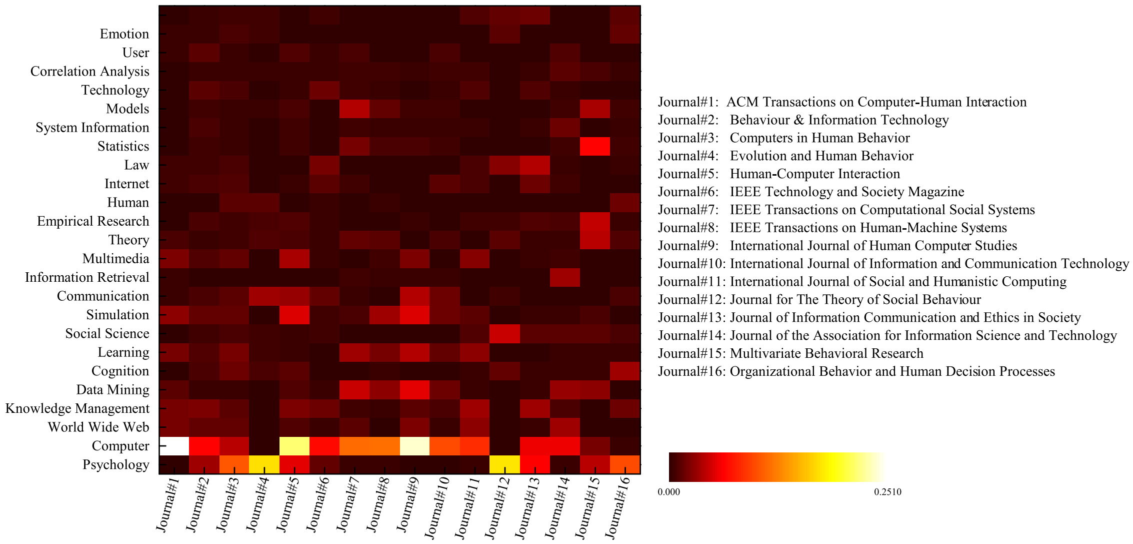

FIGURE 8. Top 25 topics distribution of journals.

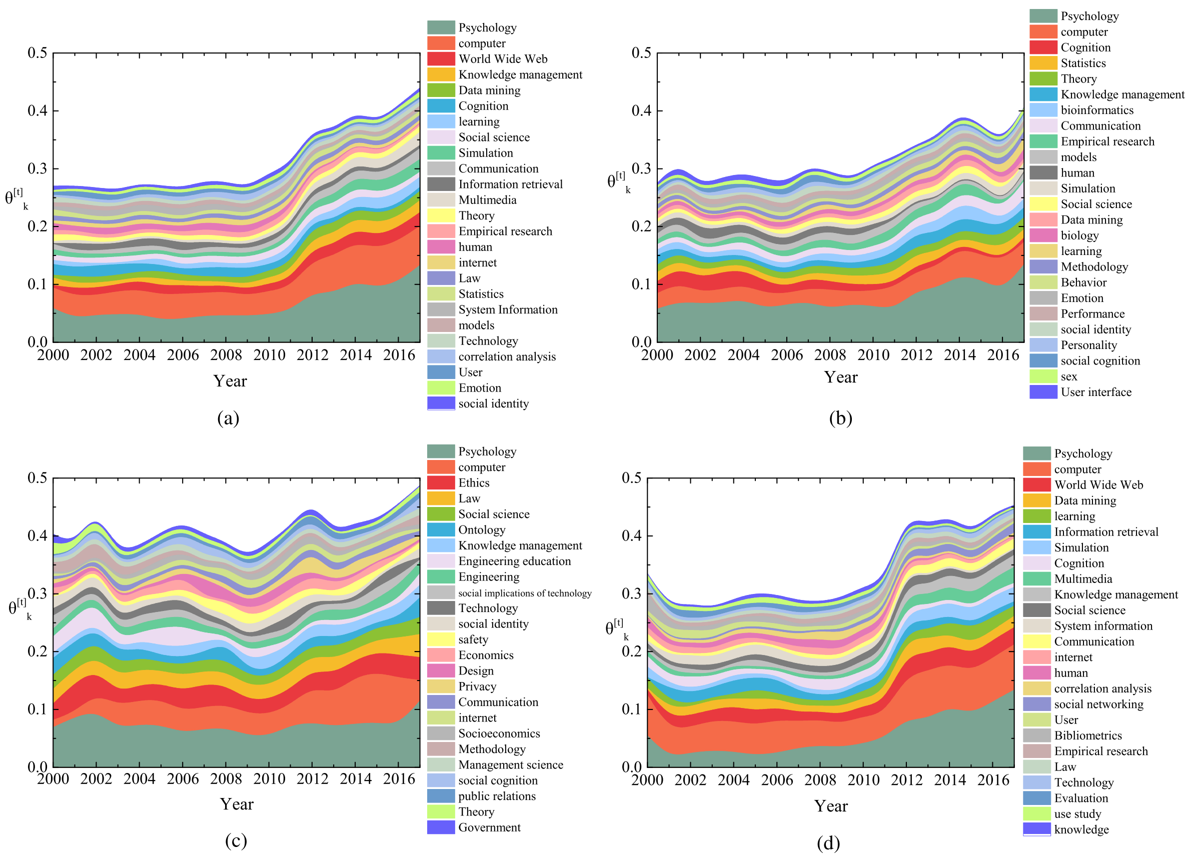

FIGURE 9. The evolution of top 25 popular topics from 2000 to 2017. (a) The proportion topics in all chosen journals. (b) The proportion topics in human-human interactive behaviour research. (c) The proportion topics in human-society interactive behaviour research. (d) The proportion topics in human-computer interactive behaviour research.

between two topics in the same community reflects the degree of similarity. Then we analyze the topic trend from human interactive behaviour research based on the relation of topic communities.

TABLE 5 shows ranking of top 25 popular topic communities based on frequency of occurrence every 2 years. By analyzing the keywords in the table, we can know how popular topics in this field have changed from 2000 to 2017.

TABLE 5. Ranking of top 25 popular topic communities based on frequency of occurrence every 2 years.

Year	Rate of Popular Topic	Top 25 Popular Topic
2000-2001	0.295596816976127	Psychology, Computer, Cognition, Knowledge Management, Theory, World Wide Web, Data Mining, Human, Statistics, Information Retrieval, Social Science, Simulation, System Information, Learning, Design, Methodology, Law, Technology, User, Models, Social Environment, Empirical Research, Communication, Internet, User Interface
2002-2003	0.294094221655848	Psychology, Computer, Cognition, World Wide Web, Human, System Information, Information Retrieval, Data Mining, Social Science, User, Evaluation, Knowledge Management, Statistics, Models, Communication, Theory, Internet, Problem Solving, Performance, Simulation, Methodology, Technology, Law, Use Study, Election
2004-2005	0.357873449228017	Psychology, Computer, World Wide Web, Cognition, Information Retrieval, System Information, Human, Data Mining, Social Science, User, Internet, Communication, Law, User Interface, Simulation, Evaluation, Personality, Language, Use Study, Empirical Research, Statistics, Models, Technology, Knowledge Management, Theory
2006-2007	0.350089281025156	Psychology, Computer, Cognition, World Wide Web, Information Retrieval, Human, System Information, Social Science, Data Mining, Learning, Empirical Research, Evaluation, Knowledge Management, Theory, Models, Correlation Analysis, Use Study, Communication, Design, User, Internet, Social Cognition, Law, Simulation, Social Identity
2008-2009	0.360885157121078	Psychology, Computer, Cognition, Human, Social Science, World Wide Web, Correlation Analysis, Resultat, Data Mining, Internet, Empirical Research, System Information, Communication, Learning, Knowledge Management, Law, Simulation, Information System, Models, Theory, Information Retrieval, Statistics, Emotion, Evaluation, Social Cognition
2010-2011	0.374513183843401	Psychology, Computer, World Wide Web, Data Mining, Learning, Social Science, Knowledge Management, Cognition, Empirical Research, Internet, Human, Communication, Information Retrieval, Statistics, Simulation, Correlation Analysis, Social Identity, Theory, System Information, Models, Emotion, Multimedia, Technology, Law, Knowledge
2012-2013	0.38319298245614	Psychology, Computer, World Wide Web, Knowledge Management, Data Mining, Communication, Social Science, Simulation, Multimedia, Learning, Empirical Research, Statistics, Cognition, Law, Information Retrieval, Theory, Social Networking, Internet, Bioinformatics, Behavior, Technology, Knowledge, Personality, User Interface, Social Identity
2014-2015	0.423286061699046	Psychology, Computer, World Wide Web, Knowledge Management, Simulation, Learning, Multimedia, Data Mining, Communication, Social Science, Safety, Suicide Prevention, Social Networking, Empirical Research, Social Media, Theory, Law, Technology, Bioinformatics, Information Retrieval, Statistics, Social Network, Cognition, Behavior, User Interface
2016-2017	0.447067328848934	Psychology, Computer, World Wide Web, Simulation, Multimedia, Learning, Knowledge Management, Communication, Data Mining, Social Networking, Social Media, Theory, Social Science, Law, Technology, Statistics, Emotion, Behavior, Information Retrieval, Anxiety, Cognition, Social Network, Models, Personality, Internet

TABLE 6. Ranking of topic communities based on the number of publications in the 21 century.

Topic	No.	Topic	No.	Topic	No.	Topic	No.	Topic	No.
Psychology	18621	Cognition	3508	Information Retrieval	2555	Internet	2167	Technology	1877
Computer	14043	Learning	3228	Multimedia	2487	Law	2162	Correlation Analysis	1820
World Wide Web	4554	Social Science	3187	Theory	2304	Statistics	2147	User	1719
Knowledge Management	3650	Simulation	3127	Empirical Research	2304	System Information	2089	Emotion	1682
Data Mining	3601	Communication	2953	Human	2220	Models	1878	Social Identity	1668

We can learn that the rates of top 25 topics each 2 years are more than One-Third, which means these top 25 topics can reflect the evolution of whole human interactive behaviour research effectively. According to the TABLE 5 we see that Popular topics only in the first half of this period are: "Learning", "Social Science", "Performance" and so on. The popular topics that come later are: "Internet", "Emotion"", "Knowledge Management" and so on. The topics that have been popular throughout the time are: "Psychology", "Computer", "Cognition", "Law" and so on. Besides, we rank top 25 topic communities by their frequency of occurrence in TABLE 6 to further investigate the popularity of topics. We can know that "Psychology", "Computer" and "WWW" have the top 3 times of occurrence from TABLE 6.

Because of the development of big data and computer technology, some topics become hot, while some become obscure. To distinguish the rising and falling tendency of

TABLE 7. Ranking of publications based on the number of citations in the 21st century.

No.	Title	Year	Citation	Journal
1	Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions	2007	4412	Multivariate Behavioral Research
2	Knowledge Transfer: A Basis for Competitive Advantage in Firms	2000	4095	Organizational Behavior and Human Decision Processes
3	A Conceptual Framework and A Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications	2001	3606	Human-Computer Interaction
4	Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods	2004	3528	Multivariate Behavioral Research
5	The Role of Justice in Organizations: A Meta-Analysis	2001	3270	Organizational Behavior and Human Decision Processes
6	The LinkâĂŘprediction Problem for Social Networks	2007	3264	Journal of the Association for Information Science and Technology
7	Ethical Leadership: A Social Learning Perspective for Construct Development and Testing	2005	2283	Organizational Behavior and Human Decision Processes
8	An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies	2011	2241	Multivariate Behavioral Research
9	Distributed Cognition: Toward A New Foundation for Human-computer Interaction Research	2000	2035	ACM Transactions on Computer-Human Interaction
10	The Process of Knowledge Transfer: A Diachronic Analysis of Stickiness	2000	2034	Organizational Behavior and Human Decision Processes
11	Distance Matters	2000	2026	Human-Computer Interaction
12	Charting Past, Present, and Future Research in Ubiquitous Computing	2000	1956	ACM Transactions on Computer-Human Interaction
13	Twitter Power: Tweets as Electronic Word of Mouth	2009	1899	Journal of the Association for Information Science and Technology
14	User Experience - A Research Agenda	2006	1753	Behaviour & Information Technology
15	Predicting E-Services Adoption: A Perceived Risk Facets Perspective	2003	1694	International Journal of Human-computer Studies
16	CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature	2006	1652	Journal of the Association for Information Science and Technology
17	The Evolution of Prestige: Freely Conferred Deference as A Mechanism for Enhancing the Benefits of Cultural Transmission	2001	1497	Evolution and Human Behavior
18	Third-party Punishment and Social Norms	2004	1474	Evolution and Human Behavior
19	Toward an Understanding of the Behavioral Intention to Use Mobile Banking	2005	1458	Computers in Human Behavior
20	Animation: Can It Facilitate?	2002	1391	International Journal of Human-computer Studies
21	The Evolution of Protégé: An Environment for Knowledge-based Systems Development	2003	1360	International Journal of Human-computer Studies
22	Conceptualizing Religion and Spirituality: Points of Commonality, Points of Departure	2000	1354	Journal for The Theory of Social Behaviour
23	A Cognitive-behavioral Model of Pathological Internet Use	2001	1312	Computers in Human Behavior
24	Searching the Web: The Public and Their Queries	2001	1287	Journal of the Association for Information Science and Technology
25	Identifying the Pitfalls for Social Interaction in Computer-supported Collaborative Learning Environments: A Review of the Research	2003	1287	Computers in Human Behavior
26	Personality and Motivation Associated with Facebook Use	2009	1266	Computers in Human Behavior
27	Who Interacts on the Web?: The Intersection of Users' Personality and Social Media Use	2010	1192	Computers in Human Behavior
28	Identity Construction on Facebook: Digital Empowerment in Anchored Relationships	2008	1185	Computers in Human Behavior
29	Predicting the Use of Web-based Information Systems: Self-efficacy, Enjoyment, Learning Goal Orientation, And the Technology Acceptance Model	2003	1110	International Journal of Human-computer Studies
30	Following You Home from School: A Critical Review and Synthesis of Research on Cyberbullying Victimization	2010	1100	Computers in Human Behavior

topic evolution, we use the parameter $\theta_k^{[t]}$ of different years and topics. We mainly focus on the dynamic change of topic. As the above-mentioned, human interactive behaviour is a large research filed, so we divide this behaviour into human-human behaivour, human-society behaviour and human-computer behaviour and survey the evolution of these three

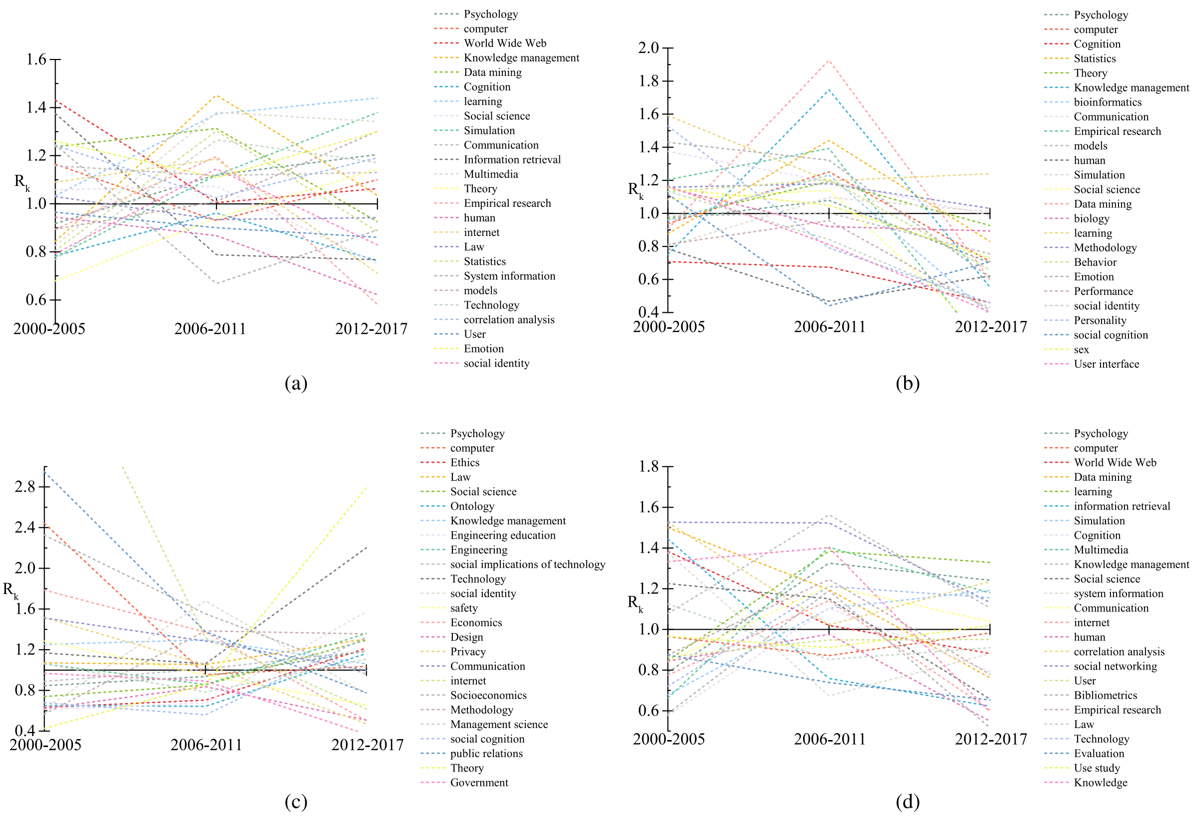

FIGURE 10. The change of R_k of top 25 topic every 6 years. (a) The change of all chosen journals. (b) The change of human-human interactive behaviour research. (c) The change of human-society interactive behaviour research. (d) The change of human-computer interactive behaviour research.

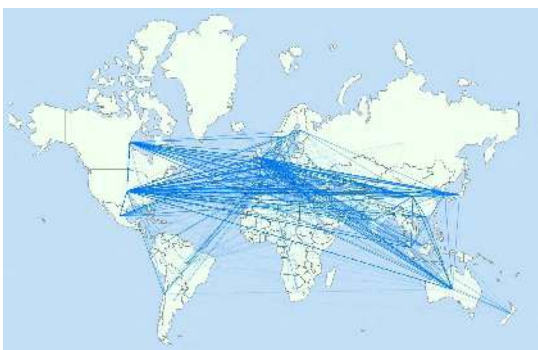

FIGURE 11. The growth rate of publication in the 21st century.

type of behaviors respectively. Fig. 9 reflects the value and variation of $\theta_k^{[t]}$ of 25 hot topics from 2000 to 2017. And These topic communities are sorted by quantity from bottom to top. From Fig. 9(a), we can see the evolution tendency of topics in all chosen journals are different. For example, the topics which become more and more popular are ''Computer'', ''WWW'', ''Data mining'' and so on. While the topics which are less popular are ''Cognition'', ''Empirical Research'' and the like. There are also some topics which are popular all the time, such as ''Simulation"'' and ''Information Science''.

Fig. 9(b), Fig. 9(c) and Fig. 9(d) shows the evolution of top topics of human-human behaviour, human-society behaviour and human-computer behaviour. We can see that there are obviously different among these behaviour. Topic ''Theory'' and ''Bioinformatics'' are popular in the H-H behaviour research, that means H-H interactive behaviour is mainly focus on the biomedical and theoretical research. Similarly, topic ''Ontology'' and ''safety'' are only hot in H-S interactive behaviour, which shows that researchers pay more attention to national development and social progress. Besides, Topic ''Learning'' and ''Correlation Analysis'' are only popular in H-C interactive behaviour, that means AI, man-machine interaction and machine learning are hot in H-C behaviour research filed.

To further explore the changing popularity of the topics over years, we use the method described in Section III-C4. We have introduced the degree of topic popularity to quantitative analyze the topic popularity. Fig. 10 shows the change of R_k of each hot topic community every 6 years. When $R_k > 1$, it means that topic k in current 3 years is more popular than the last 3 years. In Fig. 10(a), the topic ''Learning'' and ''Simulation'' have a big R_k at present in all chosen journals. Fig. 10(b), Fig. 10(c) and Fig. 10(d) shows the change of R_k of top 25 topics respectively in H-H, H-S and H-C behavior.

IEEE Access

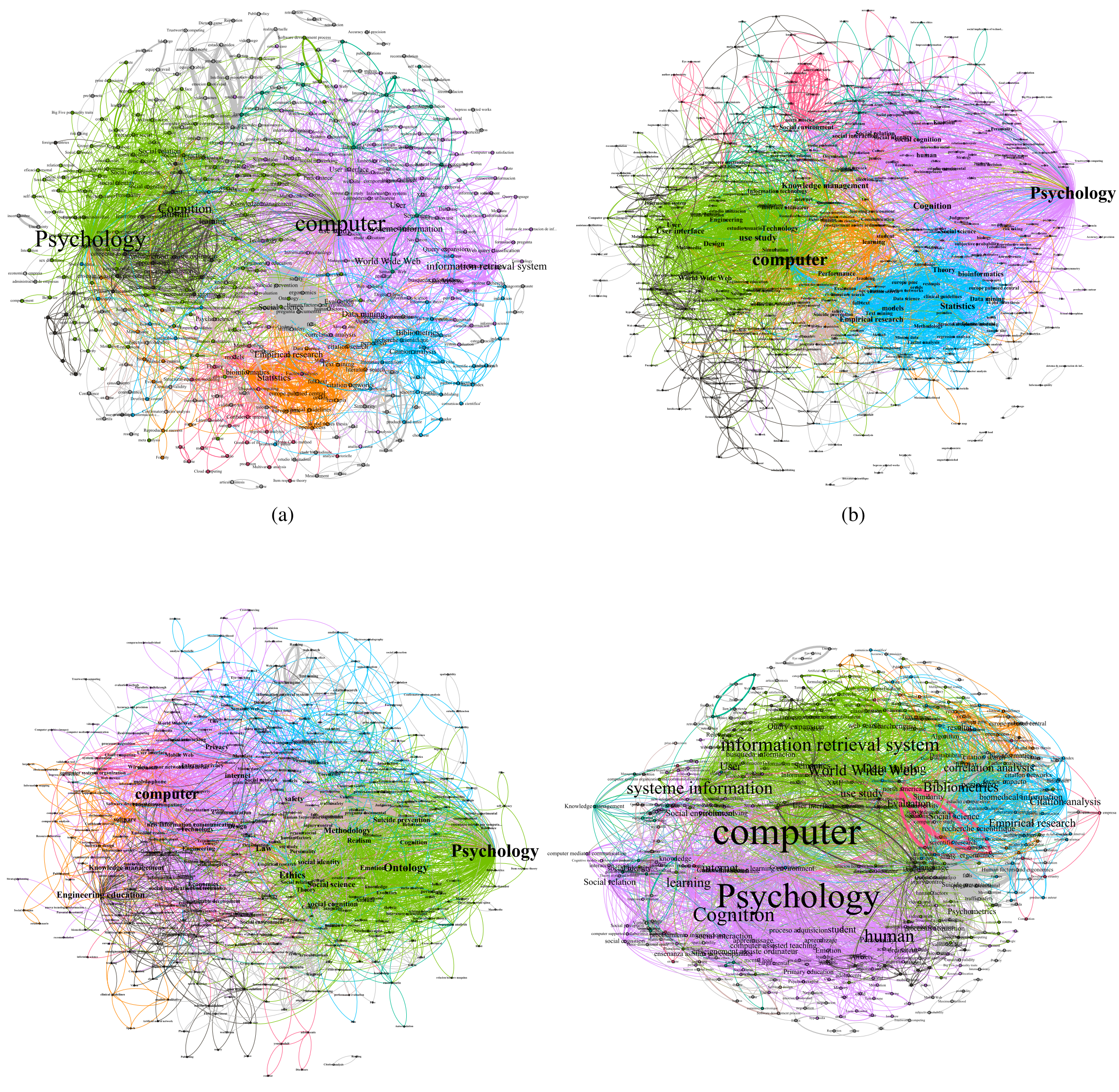

FIGURE 12. Topics co-presence network of human interactive behaviour. (a) The network of all chosen journals. (b) The network of human-human interactive behaviour research. (c) The network of human-society interactive behaviour research. (d) The network of human-computer interactive behaviour research.

Fig. 7 presents the proportion of popular topics related to human interactive behaviour in 21st century. We can prioritize these topics clearly and precisely by analyzing their frequency. Besides, these topic clouds consist of the popular topic and other topics related to it. The popularity of topics is based on the methods we have introduced in Section III-C3. The extend of relativity which is computed in the method in Section III-C5 can be represented by the size of the word. Bigger the size is, closer relationship of two topics is. For example, topic ''Emotion'' has a close connection with topic ''User'' and ''computer'', while ''Theory'' has a close connection with ''Models'' and ''Law''.

Fig. 8 shows $\theta_k^{[t]}$ as a matrix, with each column representing the top 25 topic communities distribution of a particular journal. We find that for most journals topics are widely distributed, because the $\theta_k^{[t]}$ of top 25 topics in all journals are bigger than 0. What is more, After comparing with every pair of chosen journals, we measured topic similarity between topics. If two journals have more similar color of topic, the research filed of two journals is closer.

We learn that there are co-presence phenomenon between two topics. Fig. 12 displays the topics co-presence network of human interactive behaviour, defined in Section III-C5. To ensure topics have obvious co-presence, we choose the topics A and B when $C(A, B) > 0.1$. After filtering, we get 440 vertices and 3398 edges in Fig. 12(a), 461 vertices and 4220 edges in Fig. 12(b), 404 vertices and 3366 edges in Fig. 12(c), 441 vertices and 3590 edges in Fig. 12(d).

The thickness of edge represents the value of coefficient of co-presence between two topics. And the size of label denotes the frequency of topic occurrence. For example, in Fig. 12(a), the topic "Psychology"" and "cognition" is popular in the network, and the topic "WWW" and "Web modeling" have a close relationship.

C. RANKING OF AUTHORS/PUBLICATIONS/COUNTRIES OR REGION

In order to quantify the importance of the authors, we count the times of citation for each authors from 2000 to 2017 and rank authors based on it(TABLE 4). What is more, every entry has the number of citations, publications and h-index for corresponding author attached to. Although some authors have the most publication, their h-index are not the most, just like Mike Thelwall, the first entry in TABLE 4. Besides, h-index can also reflect the academic influence of scholars. Bigger h-index is, more influential people is. This outcome maybe result from some famous papers which have large citations.

In a similar way, we quantify the influence of publications by the number of citation from 2000 to 2017, and rank papers based on it(TABLE 7). In our opinion, more times the paper has, more influential it is. For example, "Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions" is the first in TABLE 7. It has no doubt that the first is regards as one influential paper in this period. We also mark the journals where papers published in to indicate some important journals in this field.

Fig. 11 displays the growth rate of publication in the 21st century, which is drawn with the country as the node and the paper cooperation among authors from different countries as the edge. The figure consists of 43,261 nodes and 59,923 edges. The more edges one country/region has, the more connected it is to the rest of the world in the research field of human interactive behaviour. The figure clearly shows that the United States, Canada, China, Korea, Australia and Europe have more edges than other countries. This suggests that researchers in these countries are more connected to researchers in other countries than others. Researchers in these countries are more collaborative than others.

TABLE 8 shows the ranking of countries by the number of authors in the 21st century. Nearly half of the countries in this table are European and the rest are distributed in Asia, North America and Oceania. Through analyzing the data in the table, we can know that the United States that ranks first in the table has far more researchers than any other country, which is three to ten times as many as the others in the table. In addition, we can also see that the number of researchers in this field in China accounts for about half of the number of researchers in Asian countries in TABLE 8. Especially, on the one hand, China has a big number of authors and publications, that shows the improvement of academic strength. On the other hand, the average number of citations per paper of China is as high as the authors and publications, which means the paper quality needs to be further improved and there is still a long way for China to be the a powerful science and technology publisher from a large science and technology publisher.

TABLE 8. Ranking of countries based on the number of authors in the 21st century.

Country	Author	Publications	Citation	Citations per Paper
United States	6292	3909	617180	157.89
United Kingdom	1824	1412	196145	138.91
China	1277	684	70325	102.81
Canada	1048	779	118208	151.74
Spain	1001	529	77598	146.66
Netherlands	986	727	115019	158.21
Germany	906	641	80105	124.97
Australia	765	532	59005	110.91
India	670	511	48263	94.45
Korea	600	435	43717	100.50

V. CONCLUSION

Human interactive behaviour is an enduring scientific area, and many researchers from different fields pay attention to making achievements about it. However, few researches are done to explore the potential patterns and the inner structures of human behaviour. To the best of our knowledge, this paper is the first to investigate the potential rules of human interactive behaviour in the view of computer science. We do a bibliographic analysis of human interactive behaviour in the 21th century. We divide human interactive behaviour into human-human behaviour, human-society behaviour and human-computer behaviour, and choose 16 top-tier journals which belong to these three behaviour research fields. The dataset we used includes 21,372 papers and the data of important information such as authors, citations and references. We use many statistics parameters and define some indexes to make our analysis more accurate and quantitative than conventional analysis. We construct the TEC model composed of three layer modules. we input the origin datasets, and combine similar keywords into a community. Then we get 782 different themes clusters. The topics in one community can be regarded as the same kind of topics. Moreover, this model can be used to do analysis beyond human behaviour research.

From our research, we find that the growth of human interactive behaviour is in an uptrend on the whole, in terms of the number of publications, authors, references and citations. Besides, the cooperative relationships between authors and countries/regions are closer compared with relationships many years ago. It can be reflected by the increasing number of citations, co-authors from all over the world and references. We also make the mensurable evolution analysis of topics by statistical method, and not only obtain top 25 popular topics, but also get the dynamic change of hot topics by years. Some topics are hot all the time, while some are unpopular as time goes by. Finally, we do rankings to screen out the excellent authors, publications and countries/regions in the field of human behaviour research. All these findings help researchers observe the potential patterns and the topic evolution in the 21st century, which may shed dazzling light on the exploration of human interactive behaviour.

However, there are still some challenges in the future. First, the relationship between scholars not just includes the cooperation, we should consider the competition in the following work, so that we can make our academic network of human behaviour integrated. Second, the rank of authors and publications can be sorted in different ways, we will find some methods to make the ranking more reasonable than before. Finally, we hope to apply this research to practice, which is conducive to resource allocation and decision making.

REFERENCES

[1] I. Ajzen, "The theory of planned behavior," *Org. Behav. Hum. Decis. Processes*, vol. 50, no. 2, pp. 179–211, 1991.
[2] M. Perc, "Self-organization of progress across the century of physics," *Sci. Rep.*, vol. 3, Apr. 2013, Art. no. 1720.
[3] T. Kuhn, M. Perc, and D. Helbing, "Inheritance patterns in citation networks reveal scientific memes," *Phys. Rev. X*, vol. 4, no. 4, p. 041036, 2014.
[4] J. Liu *et al.*, "Artificial intelligence in the 21st century," *IEEE Access*, vol. 6, pp. 34403–34421, 2018.
[5] L. Sun and Y. Yin, "Discovering themes and trends in transportation research using topic modeling," *Transp. Res. C, Emerg. Technol.*, vol. 77, pp. 49–66, Apr. 2017.
[6] M. P. Viana, D. R. Amancio, and L. D. F. Costa, "On time-varying collaboration networks," *J. Inform.*, vol. 7, no. 2, pp. 371–378, 2013.
[7] S. Hennemann, D. Rybski, and I. Liefner, "The myth of global science collaboration—Collaboration patterns in epistemic communities," *J. Inform.*, vol. 6, no. 2, pp. 217–225, 2012.
[8] F. Xia, W. Wang, T. M. Bekele, and H. Liu, "Big scholarly data: A survey," *IEEE Trans. Big Data*, vol. 3, no. 1, pp. 18–35, Mar. 2017.
[9] L. Sun and I. Rahwan, "Coauthorship network in transportation research," *Transp. Res. A, Policy Pract.*, vol. 100, pp. 135–151, Jun. 2017.
[10] B. Chen, S. Tsutsui, Y. Ding, and F. Ma, "Understanding the topic evolution in a scientific domain: An exploratory study for the field of information retrieval," *J. Inform.*, vol. 11, no. 4, pp. 1175–1189, 2017.
[11] E. Yan, "Research dynamics: Measuring the continuity and popularity of research topics," *J. Inform.*, vol. 8, no. 1, pp. 98–110, 2014.
[12] I. Fister, Jr., I. Fister, and M. Perc, "Toward the discovery of citation cartels in citation networks," *Frontiers Phys.*, vol. 4, p. 46, Dec. 2016.
[13] D. Helbing *et al.*, "Saving human lives: What complexity science and information systems can contribute," *J. Stat. Phys.*, vol. 158, no. 3, pp. 735–781, 2015.
[14] M. Perc, "The Matthew effect in empirical data," *J. Roy. Soc. Interface*, vol. 11, no. 98, p. 20140378, 2014.
[15] K. Khoshhal, H. Aliakbarpour, K. Mekhnacha, J. Ros, J. Quintas, and J. Dias, "LMA-based human behaviour analysis using HMM," in *Technological Innovation for Sustainability* (IFIP Advances in Information and Communication Technology), vol. 349, no. 1. 2011, pp. 189–196.
[16] D. Choujaa and N. Dulay, "Predicting human behaviour from selected mobile phone data points," in *Proc. 12th ACM Int. Conf. Ubiquitous Comput.*, 2010, pp. 105–108.
[17] V. Pejovic and M. Musolesi, "Anticipatory mobile computing for behaviour change interventions," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2014, pp. 1025–1034.
[18] L. J. Eaves, K. A. Last, P. A. Young, and N. G. Martin, "Model-fitting approaches to the analysis of human behaviour," *Heredity*, vol. 41, no. 3, pp. 249–320, 1978.
[19] H. Shteingart and Y. Loewenstein, "Reinforcement learning and human behavior," *Current Opinion Neurobiol.*, vol. 25, no. 25C, pp. 93–98, 2014.
[20] S. Lühr, G. West, and S. Venkatesh, "Recognition of emergent human behaviour in a smart home: A data mining approach," *Pervasive Mobile Comput.*, vol. 3, no. 2, pp. 95–116, 2007.
[21] J. C. Cárdenas, "Human behavior and the use of experiments to understand the agricultural, resource, and environmental challenges of the XXI century," *Agricult. Econ.*, vol. 47, no. S1, pp. 61–71, 2016.
[22] C. R. de Freitas, "Weather and place-based human behavior: Recreational preferences and sensitivity," *Int. J. Biometeorol.*, vol. 59, no. 1, pp. 55–63, 2015.
[23] C.-Y. Kao and C.-S. Fahn, "A human-machine interaction technique: Hand gesture recognition based on hidden Markov models with trajectory of hand motion," *Procedia Eng.*, vol. 15, pp. 3739–3743, Jan. 2011.
[24] Y. Wu, Z. Jia, Y. Ming, J. Sun, and L. Cao, "Human behavior recognition based on 3D features and hidden Markov models," *Signal, Image Video Process.*, vol. 10, no. 3, pp. 495–502, 2016.
[25] L. Atallah and G.-Z. Yang, "The use of pervasive sensing for behaviour profiling—A survey," *Pervasive Mobile Comput.*, vol. 5, no. 5, pp. 447–464, 2009.
[26] H. W. Guesgen and S. Marsland, "Using contextual information for recognising human behaviour," *Int. J. Ambient Comput. Intell.*, vol. 7, no. 1, pp. 27–44, 2016.
[27] G. Batchuluun, J. H. Kim, H. G. Hong, J. K. Kang, and K. R. Park, "Fuzzy system based human behavior recognition by combining behavior prediction and recognition," *Expert Syst. Appl.*, vol. 81, pp. 108–133, Sep. 2017.
[28] A. Acquisti, L. Brandimarte, and G. Loewenstein, "Privacy and human behavior in the age of information," *Science*, vol. 347, no. 6221, pp. 509–514, 2015.
[29] B. Azvine and W. Wobcke, "Human-centred intelligent systems and soft computing," *BT Technol. J.*, vol. 16, no. 3, pp. 125–133, 1998.
[30] N. Jaouedi, N. Boujnah, O. Htiwich, and M. S. Bouhlel, "Human action recognition to human behavior analysis," in *Proc. 7th Int. Conf. Sci. Electron., Technol. Inf. Telecommun.*, Dec. 2017, pp. 263–266.
[31] T. Ross and G. Burnett, "Evaluating the human–machine interface to vehicle navigation systems as an example of ubiquitous computing," *Int. J. Hum.-Comput. Stud.*, vol. 55, no. 4, pp. 661–674, 2001.
[32] J. W. Crandall *et al.*, "Cooperating with machines," *Nature Commun.*, vol. 9, no. 1, 2018, Art. no. 233.
[33] D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, "Communication activity in a social network: Relation between long-term correlations and inter-event clustering," *Sci. Rep.*, vol. 2, no. 8, 2012, Art. no. 560.
[34] J. M. Carroll, "Community computing as human—Computer interaction," *Behav. Inf. Technol.*, vol. 20, no. 5, pp. 307–314, 2001.
[35] J. Zittrain, "Ubiquitous human computing," *Philos. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 366, no. 1881, pp. 3813–3821, 2008.
[36] J. Bishop, "Increasing participation in online communities: A framework for human–computer interaction," *Comput. Hum. Behav.*, vol. 23, no. 4, pp. 1881–1893, 2007.
[37] R. C. Sofia, L. I. Carvalho, F. de Melo Pereira, and S. Dattagupta, "The role of smart data in inference of human behavior and interaction," in *Smart Data: State-of-the-Art and Perspectives in Computing and Applications*, K.-C. Li, Q. Zhang, L. T. Yang, and B. Di Martino, Eds. Boca Raton, FL, USA: CRC Press, 2018.
[38] M. Gerlach, T. P. Peixoto, and E. G. Altmann, "A network approach to topic models," *Sci. Adv.*, vol. 4, no. 7, p. eaaq1360, 2018.
[39] J. Li, C. Liu, L. Chen, Z. He, A. Datta, and F. Xia, "iTopic: Influential topic discovery from information networks via keyword query," in *Proc. Int. Conf. World Wide Web Companion*, 2017, pp. 231–235.
[40] T. Mikolov, K. Chen, G. Corrado, and J. Dean. (2013). "Efficient estimation of word representations in vector space." [Online]. Available: https://arxiv.org/abs/1301.3781
[41] K.-Y. Chen, L. Luesukprasert, and S. C. T. Chou, "Hot topic extraction based on timeline analysis and multidimensional sentence modeling," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 8, pp. 1016–1025, Aug. 2007.
[42] F. N. Silva, D. R. Amancio, M. Bardosova, L. da F. Costa, and O. N. Oliveira, Jr., "Using network science and text analytics to produce surveys in a scientific topic," *J. Inform.*, vol. 10, no. 2, pp. 487–502, 2016.
[43] W. H. Gomaa and A. A. Fahmy, "A survey of text similarity approaches," *Int. J. Comput. Appl.*, vol. 68, no. 13, pp. 13–18, 2014.

XIANGJIE KONG (M'13–SM'17) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He is currently an Associate Professor with the School of Software, Dalian University of Technology, China. He has published over 70 scientific papers in international journals and conferences (with more than 50 indexed by ISI SCIE). His research interests include intelligent transportation systems, social computing, and computational social science. He is a Senior Member of the IEEE and CCF and a member of ACM. He has served as a (Guest) editor for several international journals and as a workshop chair or a PC member for a number of conferences.

KAI MA received the bachelor's degree in software engineering from the Dalian University of Technology, Dalian, China, in 2017, where he is currently pursuing the M.Sc. degree with The Alpha Lab, School of Software. His research interests include vehicular social networks, trajectory data generation, and human behavior.

SHEN HOU is currently pursuing the bachelor's degree in software engineering with the Dalian University of Technology, Dalian, China. He is also a Research Assistant with The Alpha Lab. His research interests include data mining and data analysis.

DI SHANG is currently pursuing the bachelor's degree in software engineering with the Dalian University of Technology, Dalian, China. She is also a Research Assistant with The Alpha Lab. Her research interests include data visualization and data analysis.

FENG XIA (M'07–SM'12) received the B.Sc. and Ph.D. degrees from Zhejiang University, Hangzhou, China. He was a Research Fellow with the Queensland University of Technology, Australia. He is currently a Full Professor with the School of Software, Dalian University of Technology, China. He has published two books and over 200 scientific papers in international journals and conferences. His research interests include data science, big data, knowledge management, network science, and mobile social networks. He is a Senior Member of the IEEE and ACM, and a member of AAAS.