Multimedia Tools and Applications (2019) 78:33747–33804
https://doi.org/10.1007/s11042-019-08046-6

User behavior mining on social media: a systematic literature review

Rahebeh Mojtahedi Safari [1] • Amir Masoud Rahmani [2] • Sasan H. Alizadeh [3]

Received: 22 October 2018 / Revised: 9 May 2019 / Accepted: 26 July 2019 /
Published online: 17 August 2019

Abstract
User behavior mining on Social Media (UBMSM) is the process of representing, analyzing, and extracting operational and behavioral patterns from user behavioral data in social media. It discusses theories and methodologies from different disciplines such as combining theorems and techniques from computer science, data mining, machine learning, social network analysis, and other related disciplines. User behavior mining provides a deep understanding of user behavioral data such that we observe not only individual behavioral patterns, but also interaction and communication among users by considering collective behavior of users. The aim of this study is to provide a systematic literature review on the significant aspects and approaches in addressing user behavior mining on social media. A systematic literature review was performed to find the related literature, and 174 articles were selected as primary studies. We classified the surveyed studies into four categories based on their focused area: users, user-generated content, the structure of network that content spreads on it and information diffusion. The majority of the primary articles focus on user aspect (66%); 6% of them focus on content aspect; 6% of them focus on network structure aspect, 22% of them focus on information diffusion aspect.

Keywords Systematic literature review · User behavior mining · Behavioral data · Individual behavior · Collective behavior · Social media · User-generated content · Information diffusion

1 Introduction

In recent years, with the emersion of social media services, the traces of user activities are easily maintained, generating behavioral data on an extremely large scale [102, 178, 188].

✉ Amir Masoud Rahmani
rahmani@srbiau.ac.ir

Rahebeh Mojtahedi Safari
R.Mojtahedi@qiau.ac.ir

Sasan H. Alizadeh
s.alizadeh@itrc.ac.ir

Extended author information available on the last page of the article

From one side, user behaviors in social media breed big data and on the other side, social media has become a unique source of big data, hundreds of millions of users spending countless hours on social media to share, post, connect and interact at a considerable rate. Social media is defined as a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0 and that allow the creation and exchanges of user generated content [90]. The emergence of this big behavioral data involves analyzing, modeling and predicting of user behavior on social media with new challenges. User behavioral data are mass, unstructured, and noisy and have dynamic nature. Moreover, we can observe not only individual behavioral patterns, but also interaction and communication among users by considering collective behavior of users. User behavior mining is a new interdisciplinary viewpoint that has emerged from demanding for applying new techniques. UBMSM is the process of representing, analyzing, and extracting operational and behavioral patterns from user behavioral data. It discusses theories and methodologies from different disciplines such as combining theorems and techniques from computer science, social science, data mining, machine learning, social network analysis, statistics, optimization, cognitive science, and other related disciplines. UBMSM can be helpful for user behavior prediction in social networks, Fraud detection, recommender systems, etc.

The number of ongoing studies in UBMSM is rapidly increasing due to the increasing tendency of researchers within different areas of expertise to address the problem. A glimpse of reliable published works illustrates that researchers who are interested in this promising area face a tremendous number of novel ideas, algorithms and approaches, and further expanding the scope of problems. The main goal of this study is to survey researches done in addressing UBMSM from different viewpoints by considering different dimensions (spatial, temporal dimensions). This study classifies the published studies on UBMSM into the four categorizations which are user who share or receive the information, user-generated content that is diffused among users, the structure of online social network and the information diffusion process. In fact, this study identifies the four effective aspects which affect UBMSM based on the reliable published studies between 2007 and 2019 in journals and conferences. Therefore, due to the lack of related surveys in this area, presenting a systematic review on UBMSM is necessary and will facilitate future researches. The major achievement of this review is the description of the different objectives of performing user behavior mining studies and representation of the reasonable and beneficial classifications and taxonomies of applied approaches and mechanisms. Since, the research scope of considered subject is very wide, we ignore some related topics which do not fall within the research scope of this review but will be briefly mentioned.

The main contribution of this study is to provide the insights of UBMSM by four effective aspects, believes that the proposed work is able to contribute a better and more comprehensive understanding of UBMSM for user behavior prediction, fraud detection on social media.

The structure of this paper is organized as follows. After the introduction, the main aims of the paper and research questions will be defined and described in section 2. Some concepts of user behaviors on social media and a categorization of user behaviors on social media will be defined in section 3. In section 4, after definition of UBMSM problem, four main aspects to conduct user behavior mining on social media, the effective characteristics of each aspect describing their role in UBMSM, and a classification of the applied approaches and utilized datasets are investigated and effective information and statistics are extracted for future research. The final section of the paper contains the conclusion and future works.

2 Survey goals and execution

Survey goals and research questions are described in section 2.1, and the statistics on published and presented papers in different journals and conferences are presented in section 2.4. The authors of the present paper have profited from "Guidelines for performing Systematic Literature Reviews in Software Engineering" [28] and have also used [53] to conduct and perform this research.

2.1 Survey goals and research questions

The present research aims at collecting and investigating all of the credible and effective studies that have examined *user behavior mining on social media* (UBMSM). More specifically, the extraction of salient features and methods of papers will be considered, and their characteristics will be described. To achieve the above-mentioned goals and identify the methods that have been selected by researchers for their studies and result assessment methods, datasets and evaluation metrics are used. Most researchers have considered different features which affect some special behaviors of user which provide guidelines for analyzing, modeling, predicting and recommending user behavior on social media in the future. The following research questions (RQs) are raised.

> RQ1. What are the publication statistics and venue of the existing studies on UBMSM in the literature?

Answering this question enables us to identify the distributions over popular publishers in this field.

> RQ2. What are the existing studies of UBMSM, annual distribution, and their focused area?

Answering this question enables us to classify the studies based on their focused area in the four main aspects and it help us to understand the amount of attention of existing studies on each focused area.

> RQ3. Can be proposed a general classification of user behavior?

Answering this question guides the researchers to distinguish between varieties of behaviors which occur by users in interaction with social media.

> RQ4. What are the effective aspects which affect UBMSM?

Answering this question enables the researchers to deal with UBMSM from different points of view. In other words, with the aggregation of these aspects together the temporal and statistical patterns of the time-stamped user behavioral data are unfolded precisely and eventually, UBMSM process more accurately is done.

> RQ5. What are the characteristics of each aspect?

Answering this question enables us to be familiar with significant features of each aspect that can be effective in UBMSM process and consequently in its applications such as predicting of user behavior on social media over time.

RQ6. What classifications and taxonomies can be presented for the characteristics of each aspect?

Considering the all key dimensions of each characteristic and presenting them in the format of classification and taxonomy reveals a comprehensive understanding of each characteristic that helps researchers to facilitate UBMSM process is done more precisely.

These questions are a few of many whose answers require us to review the studies and approaches applied in UBMSM process to identify effective aspects and present a variety of classifications and taxonomies which are applicable in predicting a variety of behaviors on social media in future.

2.2 Search options

Search strings are applied for academic database sources. Search strings are defined by identifying synonyms and alternative spellings for each of the question components and aggregate them by using the Boolean logics (Boolean OR and Boolean AND). For our research, a number of search strings has been constructed using relevant terms based on our research topic. Therefore, five keywords have been selected which are "user behavior mining", "social media", "user influence", "information diffusion" and "user interest modeling". In order to expand the scope of search, search strings was applied based on a review of title, abstract and body of papers. All selected primary papers were reviewed by using a set of inclusion criteria to identify they were beneficial to answer the research questions.

2.3 Database sources

Some database sources have been used for searching research publications on (UBMSM); these database sources are shown in Table 1. In the table, we have shown the name and the URL of the database sources. It was decided to search for publications in the period from 2007 up to April 2019.

2.4 Publication statistics

In this study, we searched and selected 174 articles on UBMSM in particular that were published since 2007 until April 2019 from different high-level refereed journal articles, conference proceedings, workshop proceedings, thesis and book chapters which are considered in Sections 3 and 4.We classified the studies based on their focus area in order to answer our RQ. The result of this effort is a comprehensive collection of resources that can provide an acceptable level of concepts and information about user behavior mining problem on social

Table 1 Database sources

Source	URL
Elsevier	https://www.elsevier.com
Springer	http://link.springer.com/
IEEE	http://ieeexplore.ieee.org/
ACM	http://dl.acm.org/
Google scholar	https://scholar.google.com

media and the different aspects of addressing this problem that are introduced in the literature. As shown in Fig. 1, our classification is based on the main aspects which conduct user behavior mining on social media: user, content, network structure and information diffusion. Figure 1 shows the classification of studies.

Answer to question1:

For answering RQ1, we indicate Figs. 2 and 3. In Fig. 2, we observe the majority of reviewed studies in our paper are published by ACM. The publication statistics are as follow: 36% of studies are published by ACM, 20% by Springer, 14% by IEEE, 15% by Elsevier and 15% by other publishers. As can be seen in Fig. 3, 53% of studies are published in journals, 40% of papers are presented at conferences, 3% of articles are presented in workshops, 2% of studies include thesis and 3% of them are published as books.

Answer to question2:

For answering RQ2, we selected 174 papers via searching in the data sources of Table 1, and then we classified them based on their focus area into four categories which are shown in Table 2. Figure 4 shows distribution of studies per category. In this figure, we observe that the majority of the studies were focused on user aspect (66%), 22% focused on information diffusion aspect, 6% accounted network structure aspect and 6% focused on content aspect. Moreover, Fig. 5 indicates annual distribution of the reviewed studies in UBMSM since 2007 until April 2019.

3 User behavior on social media

3.1 Human behavior dynamics definition

Human dynamics (abbreviations HD, HBD, HB) is related to a branch of complex systems research in statistical physics such as systems of complex human interactions. The main goal of human dynamics is to take in to account distinctive behaviors of humans using statistical tools and methods in statistical physics. White, et.al, indicated that human behavior dynamics refers to the impacts of various causal forces in human behavior such as, human interactions, groups, social movements and historical transitions. He investigated dynamics of human behavior in different level of social entities may consider spatial and temporal, local and long distance interactions, changes in distributional properties, synchronous and asynchronous causality in dynamical evolution. The researches in understanding human behavior both at the individual and collective scales have been enhanced after the Barabási and Albert's paper as "*The origin of bursts and heavy tails in human dynamics*" [21]. Barabási and Albert defined a model based on queue that was capable to demonstrate bursts and heavy-tailed or power-law distribution of human's inter-event times that occur in human activity.

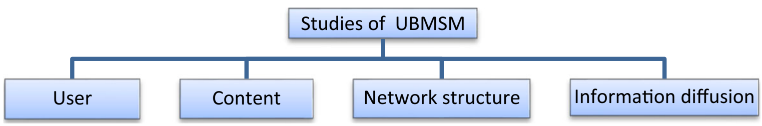
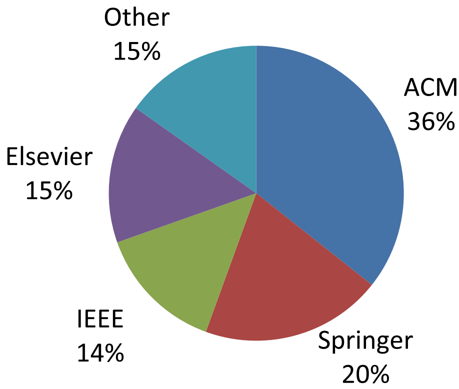


Fig. 1 A classification scheme of studies based on the focus area

Fig. 2 Studies per publisher



3.2 Behavior definition

Generally, behavior refers to the action or reaction of an actor, entity, human, user or otherwise, to situations or stimuli in its environment [31]. It is pervasive and can be widely appeared anywhere at any time. In other words, behavior has been highly regarded for complex problem-solving within virtual and physical world. Behavior analysis is a crucial concept in perceiving the driving forces and cause-effects in many disciplines such as behavioral science, computer science, social networks, recommender systems, customer relationship management, multiagent organizations, community detection, social computing, etc. In social networks, for instance, it is extensively agreed that user behavior analysis is significant topic for deeply understanding of users. Behaviors show special characteristics and features in different applications. Actor and entity can be considered as the body of behavior, action and property are indicators of behavior because they describe the body. Behaviors are categorized as qualitative and quantitative behaviors based on involved behavior type. Qualitative behavior is characterized by actions of actors and quantitative behavior is quantified by properties of entities. Quantitative behaviors can be divided into numerical and discrete behaviors which are described by numerical and categorical properties respectively.

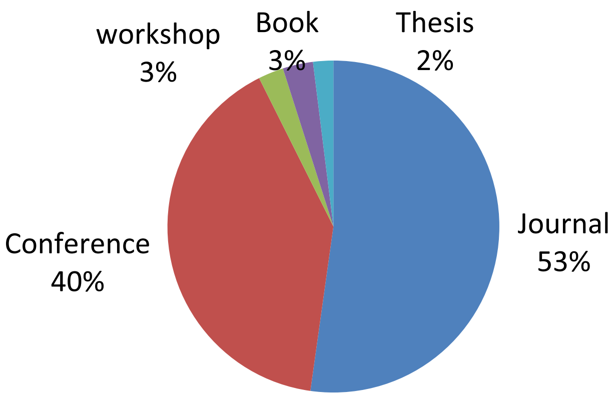


Fig. 3 Studies per venue

Table 2 A classification of studies based on the focus area

Category	UBMSM	Number of studies
User	Tie strength, limited attention, social influence measures and homophily process	113
Content	Topic detection, topic and event clustering approaches	13
Network structure	Network evolution models	11
Information diffusion	Information diffusion process and information diffusion modeling	37

3.3 User behavior categorization on social media

Users expose different behaviors based on the number of users whose participate in behavior on social media.

Answer to question3:

For answering RQ3, we can categorize user behavior on social media into two categories namely individual and collective behavior that has been shown in Fig. 6.

3.3.1 Individual behavior

Individual behavior is the behavior that one user exhibits surfing on social media. In individual behavior, our focus is on one user who does an activity in regarding to another user, an entity and a community. Therefore, we can classify individual behavior into three categories that has been shown in Fig. 7.

- **User-user behavior.** This behavior is the behavior that a user exhibits with regard to another user and can be considered as relationship behavior and communication behavior.

Relationship behaviors deal with the link structures between users. For example befriending, following, unfollowing, subscribing and blocking someone are a relationship behavior. These behaviors generate the social graph of users. Communication behavior is the behavior that

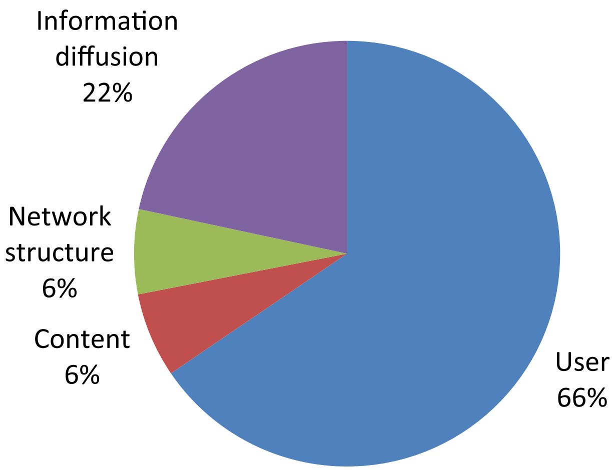


Fig. 4 Distribution of studies per category

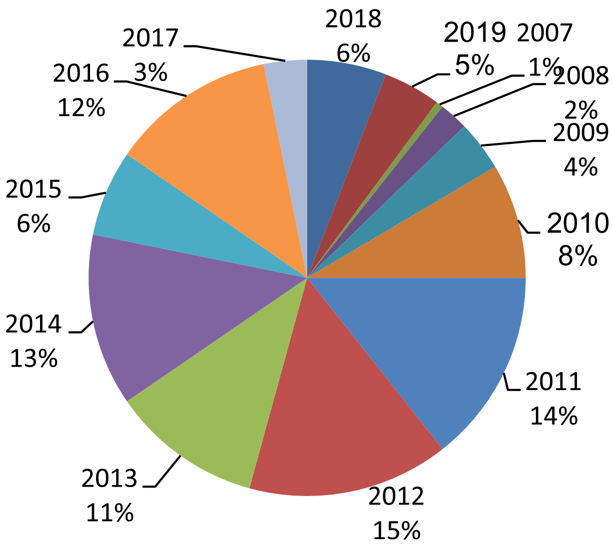


Fig. 5 Annual distribution of studies

exposes the communications between users on the social graph. For instance, when sending a message to another user, chatting, playing games, mentioning users, replying other user's tweets, etc., we are illustrating a communication behavior.

- **User-entity behavior.** This behavior appears when a user does one action on an entity. We can consider two forms for these behaviors. First one, can be seen as propagation behaviors, for instance when the user shares some content items or re-shares a content item, re-tweets tweets posted by friends, etc. second one, refers to linguistic behavior of users. For example, users may use some special terms in their profiles.
- **User-community behavior.** This behavior refers to relation between the user and a community. For example, when a user joins, leaves a community, or sends some content items to a community, we observe user-community behavior. User-community behavior is combination of two types of behaviors which above-mentioned.

3.3.2 Collective behavior

The expression collective behavior was first used by Franklin Henry Giddings (1908) and employed later by Robert E. Park (1921), Herbert Blumer (1939), Ralph Turner and Lewis

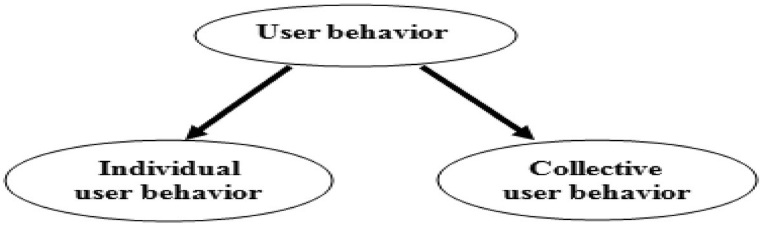


Fig. 6 User behavior categorization

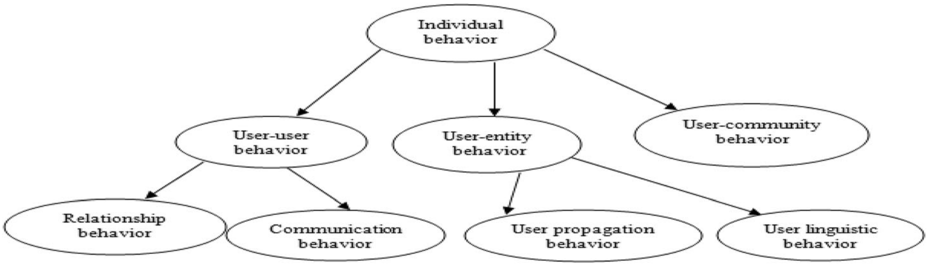


Fig. 7. Individual behavior categorization.

Killian (1957), and Neil Smelser (1962) to refer to social processes and events which do not reflect existing social structure (laws, conventions, and institutions), but which emerge in a "spontaneous" way (By Wikipedia). Collective behavior is always driven by group dynamics, encouraging people to engage in acts they might consider unthinkable under typical social circumstances [111]. This behavior appears when a population of individuals makes a similar action and behaves in a same way, not necessarily ignoring their private information signals. This behavior may be planned but is often unplanned. User migration behavior across the social media sites is an example of collective behavior.

4 The user behavior mining on social media problem (UBMSM)

4.1 UBMSM definition

UBMSM is the process of representing, analyzing and extracting user behavior patterns from big raw user-generated data in social media. It is based on theories and methodologies from social network analysis, network science, sociology, ethnography, optimization and mathematics. Also it requires human data analysts and automated software programs to sift through large-scale user-generated data (e.g., on social media usage, online behaviors, sharing of content, interaction between individuals, etc.) to discover patterns. UBMSM depicts the virtual world of a variety of behaviors which each user shows on social media in a computable way, and designs models that can help us understand its interactions over time. In other word, user behavior mining, provides necessary tools which can use effective factors such as social network topology, user communities, individual and collective attributes, the content of user activities, events, influence, homophily and the mechanism of information diffusion process to mine this world of behavioral data for discovering the interesting user activity patterns. The discovered patterns can be modeled and used in many applications such as user behavior prediction, identifying users with non-human behavior, such as malicious users, etc. As a result, UBMSM plays a significant role in social fraud detection, social emotion analysis, insight of the user behavior pattern and prediction of user behavior in future, etc. However, we encounter with many challenges in UBMSM that hinder its potential.

4.2 UBMSM challenges

Mining user behavior data is the task of mining user-generated content by considering his/her topic/interest distribution, current state of user (user attributes), social relations in the entire network, and underlying information diffusion process. Thus, this behavioral data exposes user behavior mining to different challenges on social media. The most remarkable challenges are the following:

4.2.1 Data sparsity

We encounter with the big volume of data in social media. However, when we concentrate on a particular user, we often have little data and available data is sparse. Data sparsity leads to over-fitting problems.

4.2.2 Dynamics of the network structure

The social network is an instance of complex dynamics systems, which the structure of the network and attributes of nodes change over time. Therefore, it is necessary to mine the user behavior data for discovering the evolving patterns, in order to predict user behavior and determine the state of user over time.

4.2.3 The aggregation of the effective aspects in UBMSM

There are four important aspects to conduct UBMSM which are topic/interest distribution of user, current state of user, network structure, and information diffusion process. In other word, the dynamics of user behavior interweaves with these effective aspects on UBMSM process in order to collectively determine user behavior over time. Consequently, in UBMSM process, it is hard to aggregate these aspects together to unfold temporal and statistical patterns of the time-stamped user behavioral data precisely.

4.2.4 Complexity of removing the noisy data

User behavior data include a large portion of noisy data in nature. Thus it is necessary to remove noisy data in preprocessing phase consciously, because the wrongly removing noisy data may eliminate valuable knowledge.

4.3 Effective aspects in UBMSM process

Answer to question4:

For answering RQ4, we can say the studies of UBMSM incorporate four main aspects over time: user who share or receive the information, user-generated content that diffuses among users, the structure of online social network and the information diffusion process.

Therefore, it is important to explain these effective aspects and the role of them to unfold temporal and statistical patterns behind of the time-stamped user behavioral data. Figure 8 shows a general schema of effective aspects in UBMSM issue, and its application. In the following, first we introduce some characteristics for each aspect separately that clarify to answer RQ5. Then for answering RQ6, we investigate the role of the characteristics of each aspect in UBMSM process and we present some classifications and taxonomies for each of them.

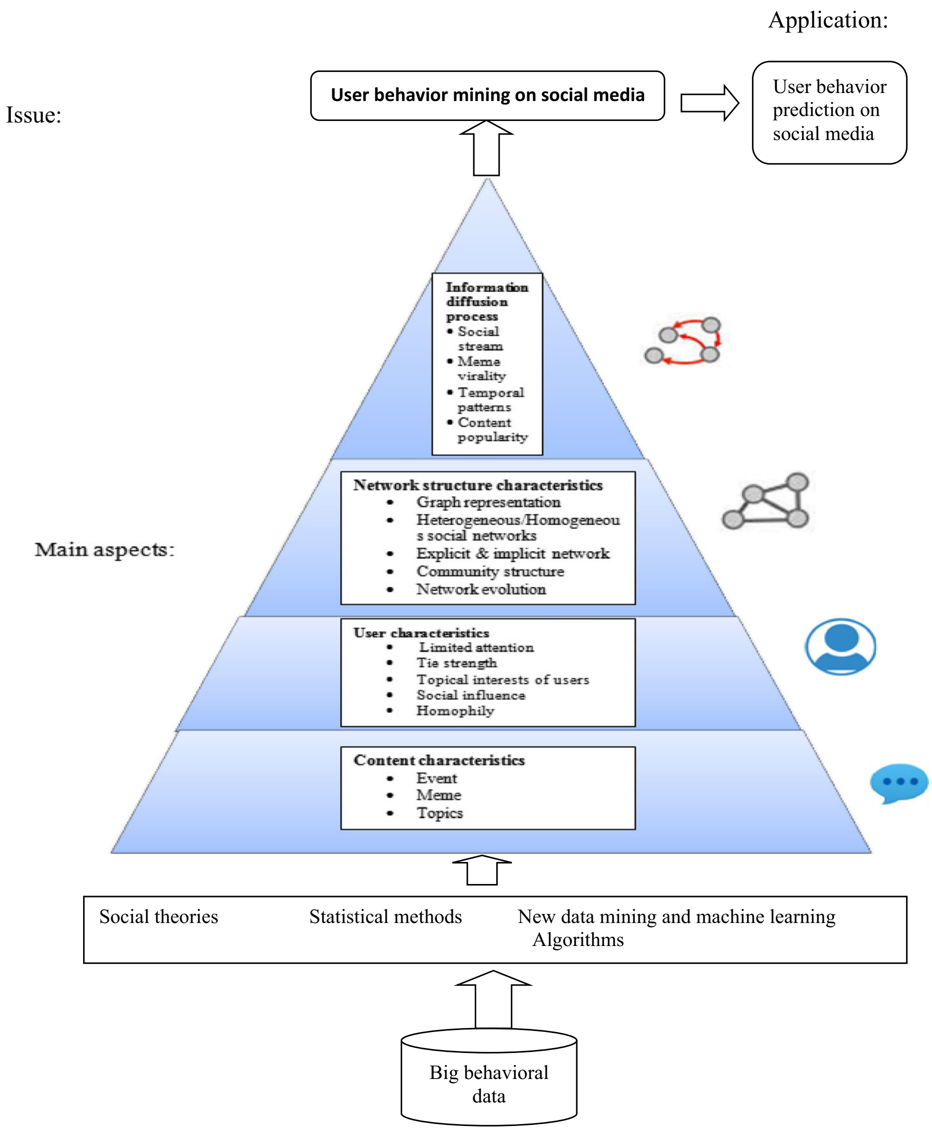


Fig. 8 A general scheme for user behavior mining on social media and its application

4.4 Effective characteristics of user aspect in UBMSM

Answer to question5:

For answering RQ5 in respect to user aspect, we can generally say, there are some characteristics of user such as limited attention of user, topical interests of user, social influence and homophily which can impact UBMSM process. First, we define these characteristics. Then we delve into the consequences and impacts of these characteristics in UBMSM.

4.4.1 Limited attention and its role in UBMSM

Nowadays, we live in a world with abundant information which is generated through various socio-technical systems and is exceeding our capacity to process it. Moreover, there is a cognitive limit to the number of relationships that one can maintain. These are relationships in which an individual knows who each person is and how each person relates to every other person. This number was first proposed in the 1990s by British anthropologist Robin Dunbar, who found a correlation between primate brain size and average social group size. [61] expressed, the amount of information an individual can produce and process is limited, whereas we are encountering with the vast amount of information which is exceeding our capacity to process it. This phenomenon named as limited attention that is a significant problem in information world [65]. In [42] book, attention has been defined as "new currency of business". Limited attention can affect UBMSM in two forms: attention to a piece of information that is received and consequently decision to propagate it, attention to other people due to creation of new social links. In former, limited attention causes an intensive competition among a vast quantity of items which are simultaneously available to attract attention. For instance, in Twitter social network the amount of attention on a hashtag is defined as the total number of receiving retweets which include that hashtag in a specific time [190]. [190] indicated the distribution of hashtag attention follows a power-law distribution. This distribution is an indicator that illustrates just a few pieces of information is popular and can achieve the high amount of attention and vice versa. In the latter, limited attention results in determining the amount of allocated user attention to other people due to creation of new social links/ties. As an indicator, the number of followers, subscribers, friends, determines how much attention a user can obtain from other users in a social network.

For decades, social science, has investigated how different types of relationships have the potential to efficiently impact users. The concept of tie strength was defined for first time by Mark Granovetter in his seminal paper as "The Strength of Weak Ties" [63]. He expressed the strength of a tie is a combination of the amount of time commitments, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie. He specified two types of ties: weak tie and strong tie. The different strength of Ties has various roles in the evolution of network structure and information sharing. The stronger tie between two users exhibits more similar they are and there is a tightly link overlap in their friendship circles and vice versa. By definition, stronger ties involve larger time commitments. [57] applied various dimensions of tie strength on social media and found that both intensity of communication and intimate factor are strong dimensions of relationship closeness. [92, 124] applied the concept of the dynamical tie strength and proposed a method that maps the dynamic nature of human interactions on to a static structure of the social network. They investigated two main dynamic factors such as the bursty nature of human interactions and the existence of group conversations affect the speed of information diffusion process in social networks. [16, 156, 179] investigated the activities such as sending private message, sharing wall posts, uploading photos and videos, liking, commenting, and etc. which users do on Facebook can be used as features for measuring of tie strength. [50] proposed a tool that uses a linear combination of 14 variables which had been collected from users' profile and posts to predict tie strength. [87] inferred tie strength from online user behavior and demographics. They showed private communications like as sending messages were not necessarily more informative than public communications such as wall posts. [164] proposed a model that uses classification techniques and related data mining methods on the users' behavior to predict their tie strength on LinkedIn social network. [19] illustrated strong ties are more influential in adaption behavior, whereas weak ties diffuses the novel information and connect users with the different communities. The presented studies showed the link overlap distribution, follows power-law distribution in twitter. Consequently, high amount of attention was observed on very weak and strong ties in social networks. This observed pattern is derived from two different competitive tendencies. On one hand, users interact with close friends due to social bonds and on the other hand, users search for information gathering due to necessary for novel information. Accordance to weak tie hypothesis defined by [63], users allocate their attention differently on strong and weak ties. Weak ties are as bridges between communities and are responsible for diffusion of novel information and do not carry much communication, while strong ties tend to be created close connections and carry the largest fraction of interactions and more traffic than weak ties. Besides, in terms of how active a user is and how traffic a link carries in the social network, the amount of individual attention assigned to a link varies. Active users (users with high out-degree, i.e., large number of follower) have more limited attention. As a result, the key role of limited attention in UBMSM is to explain the heterogeneity dynamics of global patterns such as content popularity, content lifetime, user popularity, and user activities. [195] considered the temporal evolution of popularity in Digg dataset. They used a single novelty factor to model the delay of collective attention. They represented novelty within groups follow a heavily-tailed exponential law distribution. [194] proposed an agent-based model to investigate the effect of the competitive dynamics of ideas, information and rumors on the popularity of different memes and the diversity of information for the specific topics in Twitter dataset. Some other studies noticed the competition for collective attention among multiple simultaneous spread of information [91, 103, 162, 190].

4.4.2 Topical interests of users and its role in UBMSM

A topical interest involves a set of features which are similar based on the properties of the topical interest. The topical interests affect user behavior in social media [149, 202]. [190] investigated the impact of user interests, which inferred from past behavior on predicting the future behavior of them. They defined a set of candidate posts as the n most recent posts of all users prior to the new post were appeared. Then, they used the Maximum Information Path (MIP) similarity measure to compute the similarity between the user interest and the candidate. User features include user biography, a wide variety of propagating behavioral actions, such as user' posts, comments, messages, likes, mentions, retweets, replies, and user activeness level in social media. Content features are related to the content of propagating user behavioral actions and receiving behavioral actions from her neighbors. In turn, graph features are features which exploit the structure of social media graphs such as social graph (follower/followee or friendship graph), activity graph (tweet/retweet, tweet/reply or mention/mentioned graphs) and propagation graph to infer topical interests of users in social media. Finally, hybrid features use a combination of above mentioned features. By identifying of the users' topical interests, we can construct better user profiles. [210] proposed Behavior Factorization (BF) as a way to build users' topic of interest's profiles in social media. They considered a diversity of behavioral actions have different weights. For instance, publishing actions such as posts or comments on the topics illustrate that these topics are more interesting for users who published them, thus more weight can be assigned to them in comparison to like actions which are less important to users. [172] proposed a tri-clustered tensor completion framework to improve the performance of social image tag refinement. They used user information (user interests and backgrounds) which often help to specify in (correct) tags of social images. A set of images uploaded by a specific user have close relations, if those share similar type of locations and events. Therefore, this fact can be benefit to refine image tags. They modeled the inter-relations and the intra-relations between users, images, tags. Obviously, we can claim that there are interplays between topical interests and behavioral actions of users. Therefore, it is necessary to identify topical interests and furthermore to quantify topical diversity of users. Topical diversity of users is referred to a variety of topics that each user pays attention to them throughout the propagation of his/her behavioral actions. By measuring topical diversity of users, one can distinguish users with diverse topical interests from those who focus on the one field. [192] investigated the role of topical diversity of users and messages as an effective factor to determine the influential users and popular hashtags. They found that topical diversity of users impacts on the influential behavior of users, such that the users who focus their attentions on the one field, are more influential than those who post a variety of topics. Moreover, they showed that general twitter hashtags have high popularity compared to novel ones. Many researchers applied a variety of users' topical interests modeling strategies on social media which consider different dimensions. We can broadly categorize these dimensions into the five groups: (1) Representation of user interests, (2) Semantics enrichment, (3) Weighting schema (4) Temporal dynamics of user interests, and (5) Cross sharing in multiple social networks.

Answer to question6:

For answering RQ6 in respect to the topical interests of users, a taxonomy of the applied dimensions in user interests modeling in social media is depicted in Fig. 9.

Among content features-based approaches for representation of user interests, previous studies applied Topic model, Bag of Words, Concept-based and tag-based [5, 70, 81]. In fact, Topic model and Bag of Words just focus on words and can't extract the semantics relationships among words. Moreover, because, posts are short, informal and noisy in social media, they involve a challenge for applying Latent Dirichlet Allocation (LDA) topic model in

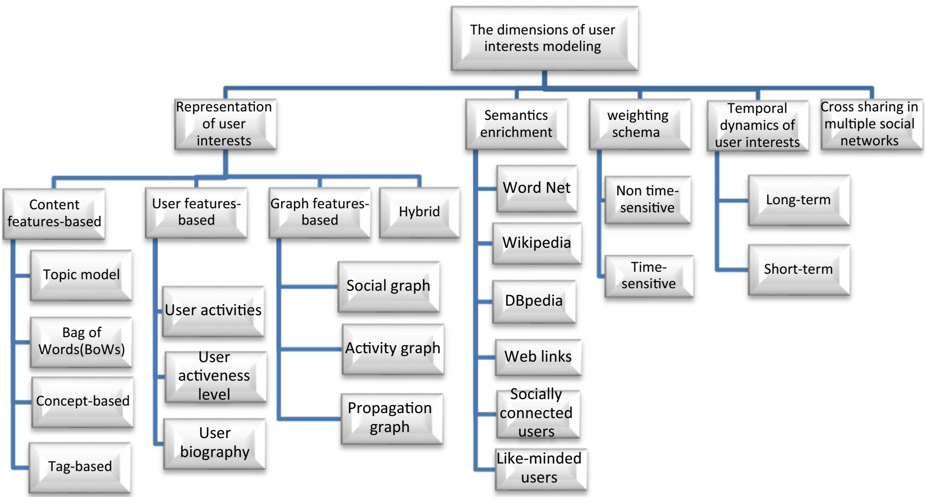


Fig. 9 A taxonomy of applied dimensions in user interests modeling in social media

comparison to traditional media. As a result, topic model is not the case in social media and decrease the quality of inferred user interests, thus it is necessary to be used a modified version of traditional topic model to discover topical interests of users on social media [32, 34, 209]. Topic models have been applied to estimate latent or implicit topics on social media. [118] improved the results of LDA topic modeling [134] used LDA model to automatically estimate users' interests. They indicated LDA model outperformed the TF-IDF baseline. [71] proposed a new approach to query log-based personalization which uses LDA to describe both the clicked URLs and the interests of the users over the same topic space. [200] introduced a modified author-topic model named twitter-user model, to discover users' topical interests on twitter. To filter out noise, they applied a tweet-level selection to distinguish tweets between user interest and social activities and achieved a better understanding of users' topics of interest. [132] focused on users who have accounts on both Twitter and Pinterest. They observed that users indicate different global patterns of activity across the twitter and Pinterest social networks. They used Labeled LDA to infer topics of tweets posted by a given user. Several studies used hybrid approaches that applied topic modeling in combination of user features or graph features. [141] proposed a scalable implementation of a partially supervised learning model (Labeled LDA) to model user interests in twitter social networks. They applied simultaneously content, user and graph features to represent user interests. [193] applied LDA model to automatically identify the topics that users are interested in it based on the tweets that they published. [9] modeled users' interests based on latent Dirichlet Allocation (LDA). They proposed TVUM (Time-Varying User Model) that divided user actions into epochs, where actions inside each epoch are modeled using fixed-dimensional hierarchical Polya-Urn model of LDA. [68] proposed an accurate hybrid approach that combines TF-IDF and topic models to estimate users' interests on social media. [211] built topic models for user behavior prediction using both the relationships from a composite social network and related behavior data. Their model is based on latent Dirichlet allocation (LDA) and model users' interests as latent topics. [74] proposed Co-Factorization Machines to address the problem of simultaneously predicting user decisions and modeling user interests in social media by analyzing rich information (tweets, co-followees, co-followers) gathered from Twitter. They used users' retweet behaviors to predict user decisions. [11] presented a probabilistic model named as the Hierarchical User Sentiment Topic Model (HUSTM), to mine the hidden structure of topics. Their work was an extension of the Hierarchical Aspect-Sentiment Model (HASM). They inserted a user-sentiment layer that captures the users' interest topics with different sentiment information. [157] modeled interests of users from their activities on social media. Their work reflects the behavioral change of users over a period based on their activities on social media over time. They divided their approach for inferring user interests into two phases which are Inferring niche interest of a user from the network and the activity graph and Inferring general interests by enriching niche interests with the semantic web using DBpedia ontology. [88] proposed a deep-neural-network-based approach for predicting user interests in social media. They used a word-embedding technique to map the words in social media content into vectors. They constructed a sentence matrix as input to a convolutional neural network (CNN) model to predict a user's interests. [176] studied the problem of inferring user interest profiles within a given time interval on Twitter. Their approach follows three steps which are (1) Inferring users' explicit interests by extracting information from the content of users; (2) Frequent pattern mining (FPM) based on the collective set of users'explicit interests to find the relation between topics in a given time interval; and, (3) Interest Profile Augmentation that augments the implicit interests into the explicit interest profile of the user based on the frequent topic patterns learnt. [187] proposed a systematic framework based on the textual and social network information for inferring users' interests on Sina Weibo microblog. They first discovered users'candidate interests based on the content of microblog, then presented a graph-based approach User-Networked Interest Topic Extraction (UNITE) based on social network information for ranking users'interests. [206] proposed an approach for inferring user interests on twitter microblog. They built an interest profile for a user over active topics by considering both explicit and implicit interests of the user. They applied three types of information for presenting their interest model which are user explicit contributions to topics, relationships between users, and the relatedness between topics. [207] proposed a framework that considers temporal evolution of user interests and uses semantic information from Wikipedia to predict user future interests. [135] studied temporal dynamics of user preferences. They proposed a temporal preference model to identify change of user preference over time. Also, they proposed strategies for changes of network structure based on centrality measure.

Traditional Bag-of-Words is a simple approach to generate the keywords which will represent users' interests in the profile. The Bag-of-Words are collection of the words used in user's post. In this approach, the frequency of words used in user's post is weight or TF-IDF method is used as weighting schema. [69] presented a hybrid approach. They modeled twitter users by the tweets and follower/followee graphs. They used TF-IDF weighting schema to determine the significance of users' terms. Other researches that infer users' topics of interest from the content of social actions and social connections data have been described by [35, 125]. [35] built Bag-of-Words profiles from users' tweets for each Twitter user. [125] applied the attributes of some users in combination with the social graph network to infer the attributes of the remaining users. They used fine-grained data taken from two large online social networks to predict user profile attributes, and presented that users with similar attributes are more likely to be friends and form dense communities. Their method for inferring user attributes was based on previous community detection approaches in social networks.

Concept-based approach uses concepts to represent user interests. There are different ways to discover concepts from users' behavioral data. [2] proved entity-based user interest profiles outperform hashtag-based and topic-based user profiles on twitter in the context of news recommendations. In contrast to Bag-of-Words, concept-based method can exploit link structure of knowledge sources for semantic linking and enriching user interests. A variety of works has been proposed to apply concept-based representation of user interests using a knowledge base from Linked Data (e.g., Freebase, DBpedia) or an encyclopedia such as Wikipedia [89, 113, 122, 128, 142] for extending users' interest profiles. [3] analyzed strategies for linking Twitter messages with related news articles to enrich the semantics of microblogging activities for creating semantically rich user profiles on the Social Web. [4] used DBpedia to enrich users' interest profiles. Their results showed that extended profiles with rich information from DBpedia outperform the original profiles without considering any extension in the context of point of interests' recommendations. [130] used a hybrid approach. They investigated entity-based and category-based user profiles based on category information for entities from DBpedia. They applied DBpedia to extract related categories for concept expansion and to analyze the structure of the categories graph in order to understand the relevance of a category for representing a user interest. [136] proposed a method for extending user profiles that exploit different types of information from DBpedia. [139] used both Synsets from WordNet and concepts from DBpedia categories to represent user interests. Their results showed that using Synsets and concepts simultaneously for representing user interests, can improve the quality of user profiles instead of using concepts alone. [138] investigated various types of information from DBpedia such as categories, classes, connected entities via different properties and the combination of them, for extending user interests. [137] presented a mixed approach using entity-based and category based user profiles with a discounting strategy for extending user profiles using background knowledge from DBpedia. They proved that their mixed approach outperforms the entity-based or category-based approach. [113] applied concepts from Wikipedia, which is a large and inter-linked online knowledge base and graph features to represent user interest profile. They proposed a new recommendation model based on Wikipedia concepts and link structure. Explicit Semantic Analysis(ESA)algorithm [51] was used to extract relevant concepts of user's tweets. ESA, found a weighted Wikipedia concept vector to represent any given tweet of users. To expand user's profile, they run random walk on the Wikipedia concept graph. [122] proposed a knowledge-based approach to identify users' topics of interest based on the entities that they mentioned in their tweets without taking into account the effect of social network. Their approach leveraged Wikipedia as a knowledge base to disambiguate and categorize the entities in the tweets. [89] presented a Hierarchical Interest Graph for Twitter users by leveraging Wikipedia Category Graph. [142] presents an alternative method to construct a hierarchical user profile using Wikipedia as the vocabulary for describing the user interests. [174] introduced TUMS, a Twitter-based User Modeling Service, that discover semantic user profiles from the Twitter messages. TUMS provides the enrichment by linking tweets to news articles that describe the context of the tweets. [97] modeled user interests by using graph features and text mining which was consisted of extracting terms, mining frequent patterns, and pruning patterns. They enriched their personal user model with collaboration from other similar users. [56, 158] proposed a List-based methodology and a service for inferring topics of Twitter users. [23] proposed a new methodology to infer topics of interest of users in the Twitter social network. They built a system Who Likes What, which can infer the interests of many Twitter users. They indicated the List-topics (extracted by their methodology) are more accurate and more complete than topics inferred by Labeled LDA on the tweets that a given user either posted or received.

Approaches based on Graph and user features, potentially characterize how popular and how a user is well-connected. Intuitively, a popular user who has many friends and followers can be actively passing information by retweeting messages [74]. [177] proposed a personalized tweet ranking method. They ranked the users based on their likelihood of retweeting the tweets and concluded how retweet likelihood correlates with the interestingness of the tweets. [29] proposed an unsupervised probabilistic generative model that combines both complex interactions between various interests of the users, their activeness level in the network, and their friends' information propagation. They considered the fact that interests changes over time and provided an online inference algorithm that balances between the current estimate of Interests and the previous estimate.

Some approaches infer users' topic of interests from multiple social networks. [130] proposed a methodology for the automatic creation and aggregation of interoperable and multi-domain user profiles of interests using semantic technologies. Also, they investigated the effect of time decay functions on ranking user interests.

Several above-mentioned works rely on the assumption that user interests remains static over time and the significance of interests are equal, while this assumption is inconsistent with the real world situation. (B. [84]) focused on long-term tweets data to construct user interests as it allows for robust performance of method. They used the topic hierarchy tree model to show user interests change over time and have hierarchy. They classified user interests into primary interests and secondary interests such as the primary interests of user hold stability over time while the secondary interests, is more likely to keep up with hot topics or events over time. [79] proposed a probabilistic method based on topic modeling for dynamic user attribute discovery. They used time windows and decay function to discover dynamic user attributes and models. [1] defined time-sensitive user modeling strategies and analyzed the temporal dynamics of user profiles inferred from users' Twitter activities. A wide variety of approaches on inferring topical interests of users are shown in Table 6 which is included in appendix.

4.4.3 Social influence and its role in UBMSM

Social influence is the process that occurs when an individual's emotions, opinions, or behaviors are affected by others such that influenced individual becomes more similar to influential ones. [143] defined social influence as the change in an individual's thoughts, feelings, attitudes, and behaviors that results from interaction from other people or group. Figure 10 depicts the social influence process. Social influence appears when an individual changes his/her behavior after interacting with other individuals. [45] described two types of social influence: Informational influence and normative social influence. Informational influence is an influence to accept information from another as evidence about reality. Normative influence is an influence to conform to the positive expectations of others. [93] introduced three broad varieties of normative social influence which are compliance, identification and internalization respectively. Compliance is when an individual accepts the opinion of others but really maintains their dissenting opinions private. Identification occurs when an individual is influenced by a famous or honorable individual. Internalization is when an individual accepts the behavior or opinion of others both in public and private and this variation of social influence appears the strongest influence. Influence strength is measured by individual influence strength, relationship strength and influence diffusion strength. Social influence across the social media causes the diffusion of different behaviors, ideas and new technologies and is becoming a complex force that conducts the dynamics of human behavior in social networks. In the real world, the influencer is someone who is followed by many people and has the ability to make changes. In social media, we observe the same traits of real world in communications where individuals are interacting together and try to expand their relationships with others. Social influence in social media involves a variety of aspects such as user influence measuring, user influence models, user influence modeling, influence applications and test.

Answer to question6:

For answering RQ6 in terms of social influence, a taxonomy of social influence aspects in social media is shown in Fig. 11.

Fig. 10 Social influence process

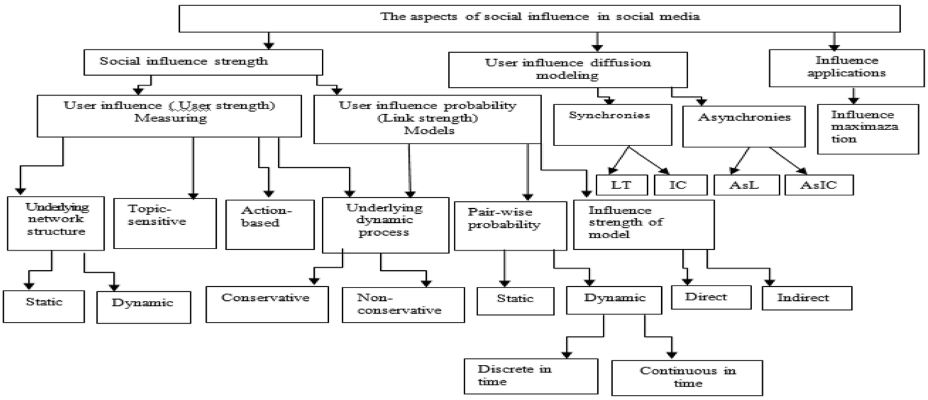


Fig. 11 A taxonomy of social influence aspects

User influence measuring The user influence measuring in social media determines the influential power of a user in the network that rises with the user relationship among other influential users. Many researchers apply various methods for measuring user influence in social media which consider underlying network structure, user content, activity of users and underlying dynamic process/flow or a combination of them. Some researchers measured the influence of users by addressing the structural properties of network such as closeness, betweenness, degree, graph centralities, PageRank and $\alpha - centrality$. The important drawback of these measures is that they are based on the static network structure whereas, in measuring user influence it is necessary to consider dynamics and changes that occur in the structure of social network over the time. [54] proposed an empirical estimate of user influence using the structure of the network and underlying dynamic process on the Digg social network. They classified different dynamic processes on the online social networks into conservative and non-conservative, based on the nature of the flow. Generally, the conservative process occurs when the initial mass or content of the network is equal to the final mass after the flow has taken place [27]. Further, the non-conservative process is when the initial mass of the network is not equal to the final mass after the flow has taken place. For example, a random walk on a network, information propagation in social networks like as Digg, Twitter, Facebook, etc., are examples of conservative and non-conservative dynamic process respectively. User influence in Twitter has been measured for the dynamic process of information propagation using number of followers, retweets and mentions[33, 203]. [33] found that the number of followers indicates user's popularity and users who have many followers do not necessarily implicate many retweets or mentions. [18] investigated influence of users in Twitter by studying the reposts of their URLs and found that the largest cascades tend to be generated by users who have been influential in the past and who have a large number of followers. Nowadays, many researchers identify influential users in the networks based on both the network structure and the topical similarity between users. In other words, influence score between users varies over different topics. [193] proposed to measure the influence of users in Twitter. They reported the phenomenon of homophily in a community of Twitter for the first time. TwitterRank is an extension of PageRank algorithms and consider underlying network structure and user topic-specific activity simultaneously. [148] proposed an algorithm that uses the concept of user passivity and information forwarding activity of users to determine user influence in Twitter. According to the large studies in information propagation, the majority of users acts as passive information consumers and does not forward the content to a Twitter network. [115] presented a method to detect influential users using their prior word usage in social media. They computed psycholinguistic category scores from word usage, and explained how people with different scores exhibited different influence behaviors on Twitter. Similar to [18, 33] they computed user influence using the average number of reposts from a user's tweets, but their research was complementary to [18]. [170] proposed a Topical Affinity Propagation (TAP) approach to describe topic-based social influence problem in social network. TAP applies both the results of the topic modeling and the underlying network structure to model topic-level user influence in social networks. [62] measured the influenceability of a user as the ratio between the number of actions for which the user was influenced, over the total number of actions performed by the user. Their measure is user activity-based and considers dynamic nature of network structure over time. [105] proposed two dynamic user influence measures named as time-window diffusion size and temporal closeness centrality on the temporal influence network that incorporate the temporal information. They indicated a user with a larger time-window diffusion size is more influential. [144] proposed a survey named as "Measuring user influence on Twitter" and investigated a variety of literatures with considering different aspects for measuring of user influence in Twitter social network. [95] reviewed the models to find influential bloggers in the blogosphere. They classified the existing models into feature-based and network-based categories. [215] proposed SIRank method to measure the spread influence of users in microblog. They applied the user interaction features, retweet intervals, location of users in information cascades and other relevant features in their work. [13] proposed Personalized PageRank (PPR) to infer topical influential users on Twitter. They used both the information obtained from network topology and the information obtained from user actions and activities. [166] introduced a model based on multi-features to find influential users in Sina Weibo. They applied the advantages of Bayesian network and PageRank algorithm in their model. They used features such as tweet content, retweet count, follower and following counts, comments, tweet count, user authentication professional background, and user interest to construct their model. [159] proposed a hybrid approach to find super spreaders based on k-shell measure. They considered the core number of a node, a node's degree and its friends' diversity in different shells in their method. They then used a method for reducing the overlap degree between selected nodes in order to maximize the spread of influence. Table 7 in appendix, illustrates a review on the applied user influence measures in different studies in terms of some aspects such as underlying network structure, topic-sensitive, activity-based and underlying dynamic process.

User influence probability models Measuring the user influence probabilities is another aspect of social influence studies that can be generally classified into the static and dynamic measuring. In the static measuring, the probabilities of user influence are constant and don't change over time but in the dynamic measuring, it is assumed that the probabilities of user influence change over time. The main difference of the static measuring models in compared to the dynamic measuring models is that they consider the current state of the network and most influential users in that state are measured whereas the dynamic measuring models maintain the history of the network in various times. To present dynamic user influence probability models, there are two types of probability models. The first one is continuous in time models which consider the evolution of a social network over time based on continuous functions of time. The experiments shows that continuous in time models are the most accurate, but they are very expensive to learn and take high computational time for large scale data because they are not incremental. Therefore, an approximation, known as Discrete in Time models have been proposed. The discrete in time models consider snapshots of a social network in the different points of time. These models are widely used and are more efficient because they are incremental and don't need to perform moment-by-moment measurements. [151] addressed the problem of predicting diffusion probabilities in complex networks for the first time. They focused on independent cascade(IC) model and defined the likelihood function for a sequence of newly active nodes and then, they used Expectation Maximization (EM) algorithm to predict diffusion probabilities which maximizes likelihood. They applied a greedy algorithm that uses IC model and learned influence probabilities based on EM algorithm to select high influential users. [62] investigated the problem of learning user influence probabilities from a log of past propagations under an instance of the General Threshold Model. They proposed both static and time-dependent models for measuring the probability with which a user is influenced by its neighbor. [170] studied topic-based social influence to learn the influence probabilities through mining past users' behavior. Given a topic distribution for each user in the social network and the network structure, they formalized the social influence problem in a topical factor graph model (TFG). Then, they proposed a Topical Affinity Propagation (TAP) on the topical factor graph to infer topic-specific social influence. [197] further investigate how to learn the influence probabilities from the history of user actions. [26] in his thesis proposed a new method based on agglomerative clustering techniques for learning influence probabilities between users. [94] assumed a uniform propagation probability and implemented information cascade model for information propagation. [15] estimated the propagation probability between two users using demographic and content characteristics. [108] proposed a generative graphical model which combines both heterogeneous link information and content associated with each user in the network to learn topics and influence strength simultaneously. Then based on the direct influence between users, they investigated conservative and non-conservative dynamic process mechanism to study indirect influence between users. They studied how friends influence each other and how influence propagates in heterogeneous social networks. They assumed the behaviors of an influenced user are based on either her/his interests or are influenced by his/her friends. [184] proposed a probabilistic model called latent influence and susceptibility(LIS) model for predicting the cascade dynamics by learning two low-dimensional user-specific vectors from observed cascades, capturing their influence and susceptibility respectively. The propagation probability of user is computed by his/her activated neighbors' influence vectors and his/her own susceptibility vector. [105] proposed an influence diffusion model based on Continuous-Time Markov Process (CTMP) to predict the influence dynamics of social network users. This model can predict the spreading coverage of a user in a given period of time. [116] presented a method for learning the diffusion probabilities in the Independent Cascade Model (ICM). They introduced linear decay and exponential decay weighting schemes and allocated a weight to diffusion samples of observed cascades. [86] introduced a model for information diffusion on Sina Weibo microblog. Their model consists of both direct diffusion and indirect diffusion. The number of reposts was considered as an important factor to measure the direct influence between users. Table 3 indicates some studies that are focus on measuring of Influence probability between users in social media.

Table 3 Influence probability models in social media

Journal/Conference	Study	Technique	Algorithm	Static	Dynamic	Influence strength	Network
Springer	[151]	• Based on a very rich set of past cascades • Based on the likelihood of a sequence of cascades as a function of the diffusion probabilities.	EM	–	continuous in time	Direct influence	Homogeneous
ACM	[62]	Bernoulli distribution Jaccard Index Bernoulli model with partial credit Jaccard model with partial credit Based on the notatin that influence probabilities decay exponentially with time	MLE	*	– Continuous in time Discrete in time	Direct influence	Homogeneous
ACM	[170]	Topic-based social influence The social influence problem is mapped in to a topical factor graph model (TFG) Learning probabilities based on user behavior mining	TAP algorithm distributed TAP learning algorithm	*	–	Direct influence	Homogenous
ACM	[197]	Based on the history of user's actions	MLE	–	Continuous in time	Direct influence	Homogeneous
Thesis of McGill University	[26]	Based on the local structure of a network Based on the past history of interactions between a specific user and its neighbors.	Agglomerative Clustering	*	–	Direct influence	Homogenous
Springer	[108]	Probabilistic generative model different types of influence propagation methods	Gibbs sampling algorithm a weighted linear combination	*	–	• Direct influence • Indirect influence • Global influence	Heterogonous
ACM	Wang et al. 2013c	Latent influence and susceptibility (LIS) model Based on learning two low-dimensional latent vectors for each user from observed cascades	An iterative algorithm using Projected Gradient (PG) method	–	Discrete in time	• Direct influence • Indirect influence • cumulative effect	Homogenous

Table 3 (continued)

Journal/Conference	Study	Technique	Algorithm	Static	Dynamic	Influence strength	Network
		Considering cumulative effect in influence propagation					
ACM	[169]	The learning model using the latent states of users' actions based on the assumption in HMM and Kalman Filters	HMM	–	Continuous in time	• Direct influence	Homogenous
ACM	[39]	Item-level social influence	A hybrid Factor Non-Negative Matrix Factorization (HF-NMF)	–	–	• Global influence	Homogenous
IEEE	[116]	learning diffusion probabilities of ICM based on Linear Decay and Exponential Decay Weighting schemes	EM with priors	–	Discrete in time	• Direct influence	Homogenous
Springer	[86]	Based on the number of reposts at a distance x and posting time t	The minimum sum of the squares	–	Discrete in time	• Direct influence • Indirect influence	Homogenous

4.4.4 Homophily and its role in UBMSM

"Homophily is a phenomenon that occurs on a daily basis in social media and is realized when similar individuals with more probability connect to each other than dissimilar ones" [98, 117]."Birds of a feather flock together," or "similarity breeds connection" are two common idioms which are continuously applied when describing homophily. Many researches on homophily area in social networks investigated dyadic measures, pre-defined social groups, or global patterns in the network by different experiments in the online settings [10, 14, 38, 44, 52, 155]. The empirical researches on homophily indicated similarity between individuals can be measured in terms of many different individual characteristics, including demographic characteristics, geographic locations, or topical interests expressed during interpersonal conversation [117]. [36] measured network structure and [7] considered network communications to examine homophily whereas, [19, 37, 189] applied combining of these two measures. [67] demonstrated the role of group size in social networks and the influence of homophily on the speed of information transmission. Some papers focused on the theoretical literature on the impact of homophily in communications in social media [24, 58, 82, 168]. Jackson and Lopez-Pintado showed the influence of the homophily on the spread of an idea throughout the entire network. There is a feedback loop between individuals both tend to connect with similar ones (homophily) and grow to be like as their friends (social influence) meanwhile, there are dissimilarities among close individuals. In a paper as *" Topic-Based Clusters in Egocentric Networks on Facebook"* the relationship between friendship structure and topical interests within local networks centered at individuals was been investigated and a feedback mechanism involving homophily, influence, or both was been suggested [191]. [110] proposed a sequential model of Bayesian social learning in networks and they found the density of network connections specifies the effect of preference diversity and homophily on learning. Figure 12 shows homophily process in the social network. Homophily affects UBMSM in various ways. The impact of homophily on relationship behavior in social media can be interpreted by link creation so that users tend to create a new link with similar others over time. As another role of homophily in UBMSM, we can refer to its impact on community behavior in social media. In general, social networks consist of densely connected subgraphs called communities so that the users in the same community are more likely to be similar and to adopt the same ideas. Hence, homophily may affect the adoption of ideas in communities such that social relationships are more likely created between the users which share similar ideas. In other words, communities use homophily mechanism to show the users with similar characteristics create more connections among themselves in natural. Besides, homophily potentially affects the diffusion of information and consequently we can observe the various patterns of propagation behavior within and across communities in social media. In fact, we expect more communication through intra-community links than inter-community links. Therefore, propagation pattern of content differs within and among communities.

4.5 The effective characteristics of content aspect in UBMSM

Social media content is generated through social media interactions done by the users through the site (Wikipedia). User-generated content on social media can be classified in to non-textual content such as, URL, videos or images and textual content. [173] proposed a novel Social anchor-Unit Graph Regularized Tensor Completion (SUGAR-TC) method to refine image tags, regardless of the data scale. Their work was a dominant study in image tag refinement which can efficiently assign high – quality tags to all non-anchor images by using the potential relationships between non-anchor units and anchor units. Users in any social media platform like Twitter, Facebook, Google plus and etc., generate content, messages, ideas or posts or consume them.

Answer to question5:

For answering RQ5 in respect to content aspect, we can generally say, there are some effective characteristics of content aspect such as Topic level network (Topic detection, Topic locality and Topic clusters) and Topical diversity which impact UBMSM process.

4.5.1 Topic level network and its role in UBMSM

There are a variety of topics which can be inferred from entities in content of social media. In other words, we can map information propagation process into the topic level network where memes of content are considered as nodes that are connected when they co-occur and a topic is assigned to a cluster of similar memes. In fact, Messages, ideas or posts in social media are semantically considered similar, when they discuss about same topics [108]. Obviously, in the real world, we focus more on the social network structure and information diffusion flows, to mine the user behavioral patterns in order to predict the user behavior over time whereas, topic formation and finding the correlation between the social network and the topic level network aids us to mine more precisely user behavioral patterns on social networks over time [191]. Therefore, it is necessary to consider topic space and investigate the aggregation of the structure level and topic level of the social network in UBMSM process over time.

Topic detection The ability to detect topics of online content can provide us with many applications and studies in the online scenario and social media. There are several questions which need a method to infer topics hidden behind the user-generated content in social media like as do the generality of user-generated content affect content popularity? Do users who discuss on a variety of subjects are more influential than those who focused on special ones? In response to these questions, we can say when generated content, messages, ideas or posts are general, many users tend to propagate and share them, and hence propagation behavior is appeared. As mentioned in section of 4.1.3, there are several approaches to infer topics of user-generated content in social media. For example, We can detect a topic by clustering a group of similar posts observed [191].

Topic locality In the context of social media, topic locality refers to the assumption that semantically similar entities are more likely to be appeared in the same posts and therefore to be close to each other in the entity specific co-occurrence network [192].

Fig. 12 The impact of homophily on the social network

Topic, meme and event clusters User-generated content clustering in social media has extending applications like as topic detection and content recommendation. For example, in twitter social network, users propagate entities created autonomously as explicit or implicit topic identifiers through social network interactions. There are several approaches for clustering user-generated content. [190] formed clusters of semantically similar hashtags as topic clusters in twitter social network. With the topic locality assumption, they indicated detected topic clusters are densely connected. They detected topic clusters by finding communities using the Louvain community detection method [25] in the hashtag co-occurrence network. [49] introduced the notion of protomemes to pre-cluster messages in real-time, streaming social media scenarios. Then, they defined several similarity measures between protomemes leveraging various content, metadata, and network features of tweets, and used a combination of these similarity measures. They showed that a simple combination based on pair wise maximization of similarity is as effective as a non-trivial optimization of parameters for robust performance. [41] proposed a framework for meme clustering in Reddit. They used Google Tri-gram method (GTM) semantic similarity to propose four semantic similarity scores and their combinations between submissions. Other research efforts related to meme clustering have illustrated in Table 4. A classification of the combination strategies for similarity measures are shown in Table 5.

4.5.2 Topical diversity and its role in UBMSM

By considering the correlation between the social network and the topic level network, each user or content involves a variety of topics and it is necessary to define topical diversity and its role in UBMSM. Topical diversity of content is a measure that determines how much content support a variety of topics in social media [192]. Topical diversity of a user's interests determines how diverse topical interests of a user are. Topical diversity affects content and user popularity in social media over time. On the one hand, there is a correlation between topical diversity of a hashtag's users or co-occurring hashtags and its future popularity. That means, the users who focus on the specific topics are connected to a few communities whereas, the users with diverse topics act as information diffusion channels and are connected to many communities in the social network. Therefore, according to the weak tie hypothesis a hashtag that the users with high topical diversity adopt it, will obtain more popularity in the future. Also, a hashtag is more likely to go viral and attract the attention of more users, if it co-occurs with many popular hashtags about diverse topics. Meanwhile, topical diversity of a user interests affects the social influence of a user, means that a user who is focused on a specific topic attracts more attention of other users and is more influential, because the concentration of user on a specific topic enhances content worthiness to be retweeted or diffused by other users. Furthermore, there is a positive correlation between topical diversity and tie strength so that we expect to observe the different response patterns in strong and weak ties. The results of earlier studies show strong ties reply to a wider variety of topics than weak ties. Hence, users with strong ties support high range of common interests in comparison to weak ties and follow similar behavioral patterns.

Table 4 Meme, topic and event clustering approaches

Journal/ Conference	Author	Features	Dynamic clustering method	Similarity measure	Evaluation metrics	Approach	Dataset
IEEE conference	[49]	Content, metadata and network features	Hierarchical clustering	The combination of four cosine similarity measures based on The pairwise maximization strategy	Normalized Mutual Information (NMI)	Unsupervised	Twitter
IEEE conference	[41]	Content features	Hierarchical clustering based on Google Tri-gram Method(GTM) semantic similarity	Internal Centrality-Based Weighting and similarity Score Reweighting with Relevance User Feedback strategy	Purity(the number of correctly assigned submissions over the total number of submissions)	unsupervised	Reddit
Springer Journal	(Jafari Asbagh et al. 2014)	Content, metadata and network features	Online K-means algorithm (Protomeme stream clustering (PSC))	The combination of four cosine similarity measures based on The pairwise maximization strategy	normalized mutual information (NMI)	Unsupervised	Twitter
SIAM	[8]	Content and network features	Online K- means clustering	a linear combination of the structural and content-based similarity values	Purity	Supervised /Unsupervised	Twitter and Enron Email
ACM conference	[22]	Content and metadata features	a single-pass incremental clustering algorithm with a threshold parameter	Ensemble-based technique for learning a similarity metric and classification-based similarity	Normalized Mutual Information (NMI) and B-Cubed	Supervised	Flicker and Last.fm

Table 5 A classification of the combination strategies for similarity measures

The combination strategies for similarity measures	Goals
Pairwise maximization strategy	The pairwise maximization strategy aims at choosing the measure that provides the highest value every time.
Pairwise average strategy	The pairwise average strategy computes the average value of the pairwise similarity measures. This strategy balances the scores among the similarities.
Linear combination strategy	In the linear combination strategy, users can assign different weighting values manually.
Internal centrality-based weighting	Internal Centrality-Based Weighting computes optimized weight factors for the linear combination strategy.
Similarity score reweighting with relevance user feedback	It manually specify the relationships between pairwise documents to guide the document clustering process and incorporate relevance user feedback by a visualization prototype

4.6 Effective characteristics of network structure aspect in UBMSM

Answer to question5:

For answering RQ5 in respect to network structure aspect, we can say there are some effective characteristics of network structure aspect in UBMSM which are dynamic or static network structure, explicit or implicit network structure, heterogeneous or homogeneous social network structure, global or local social network analysis and evaluation measures of social network structure.

4.6.1 Dynamic vs. static social network structure

Network structure specifies the underlying topology of linkages among individuals, which can be changed over time or be static. A dynamic network structure is considered with N nodes observed over time intervals $t = 1, 2, \ldots, T$ captured by a graph whose topology may arbitrarily switch between S discrete states. In other words, these networks consider the evolution of network over the time. Dynamics of the network topology directly affects information diffusion among users so that presents significant indicators to predict the future meme popularity in social media over time. Also, information diffusion process affects the network evolution through link creation behavior. In fact, there is the interplay between dynamics of the network and information diffusion process in social media over time. Understanding of this interplay helps UBMSM. Based on the classification of user link creation behavior there are several mechanisms to create new links which are triadic closure, traffic-based shortcuts, random browsing, mixture and causal friendship. For example, triadic closure is a strategy that forms new links based on shared friends. It means that two users with mutual friends are more likely to form a new link. It should be noted, a large portion of new links are created based on traffic-based shortcuts means users prefer to create a new link to others from whom or through whom they have received information. Traffic-base shortcut mechanism takes the temporal user activity patterns into consideration. By considering triadic closure mechanisms with traffic-based mechanisms we can make the network more efficient by shortening the path of information diffusion. Users apply different link creation mechanisms based on their behaviors over time. In other words, triadic closure mechanism applies when a user follows a small number of users while, traffic-based shortcut mechanism uses when a user follows many others or have been active for a long period of time. Moreover, users prefer to follow users who their posted contents are exposed multiple times. This phenomenon illustrates that all shortcuts are not equally probable.

There are several studies which account for evolution of the network structure in social media that we mention to some of them. [17] presented a switched dynamic structural equation model for tracking the evolution of dynamic networks from information cascades when topologies arbitrarily switch between discrete states. They assumed that the number of states is known. Their model inferred dynamic network such that captures sudden topology changes within a finite state-space. Moreover one can even identify the underlying network topologies with fewer cascade measurements by leveraging edge sparsity. [48] investigated network evolution of Twitter social network. They revisited Twitter graph to examine the changes exhibited in both the graph and the behavior of the users in it. They observed a network that gets denser through the years, with the number of edges between the users. Moreover, they indicated a "rich-get-richer" phenomenon, since the increased number of edges is mainly directed towards the most popular users. [119] applied a framework that considers the interplay between Individuals' evolving behavioral patterns and traits in dynamic heterogeneous social networks. Their framework was able to construct changes in peoples' social network patterns with time, measure changes in individuals' social network patterns with time, cluster the individuals with similar evolving patterns and connect similar clusters in individuals' behaviors. [199] extended the well-known static stochastic block model [73] for static networks to model time-evolving social networks. Their model used a set of unobserved time-varying states to characterize the dynamics of the networks. [131] applied a multi-layer graph concept for analyzing the dynamic social networks. They presented a generated hierarchical latent-variable model and methods for mining multi-layer networks. [112] explained social systems can include hierarchy and introduced multilevel social networks. A multilevel network is defined as distinct types of nodes taken place on multiple levels with links between all nodes in inter and intra levels. [163] in his book mentioned the analyzing of the social network is a multilevel task and depends on relations rather than individual attributes. For further study, different multilevel models for social networks have been proposed by [161, 186].

4.6.2 Explicit vs. implicit social network structure

An implicit network in contrast to an explicit network is a network where information diffuses along the network; however, unlike the explicit network, we cannot observe the individuals who are responsible for influencing others (who influenced whom), but only those who get influenced. In other words, in implicit network there is no knowledge of the structure of social network.

4.6.3 Heterogeneous vs. homogeneous social network structure

Networks can be treated as homogeneous or heterogeneous. In the former case, a single link type is studied; or multiple link types are treated equally and aggregated. In the latter case, different link types are explicitly distinguished, and often, their interplay is studied [119]. Many online social networks are heterogeneous and dynamic consisting of different types of object nodes. For example, Twitter is comprised of users and microblogs. Digg consists of users and website URL addresses. Citation network consists of authors and publication papers. Here, we use "document" to represent different types of associated content (e.g., microblog, website, and paper) to each user. Therefore, links in heterogeneous networks would contain friendships between users and authoring relationships between users and documents. The links can be directed or undirected. As an example of Heterogeneous domains we can mention to work by [160]. They proposed a new weakly-shared Deep Transfer Networks (DTNs) to translate cross-domain information from text domain to image domain. Their proposed network can represent complex representation of data in different domains.

4.6.4 Global vs. local social network analysis

Network analysis can be considered in global or local scale. In the former, one measures global network-level properties, such as the degree distribution [99]. Global properties might not be able to capture effectively complex topologies of real-world networks. Therefore, local network analysis that focuses on individuals might be more appropriate, via, e.g., centrality measures such as degree or betweenness.

4.6.5 Evaluation measures of social network structure

There is a variety of measures that use to evaluate social networks. For example, in order to capture network characteristics fully, we can apply multiple centralities such as degree centrality, clustering coefficient centrality, k-core centrality, betweenness centrality, closeness centrality and graph degree centrality [123, 196].

4.7 Effective characteristics of information diffusion aspect in UBMSM

Researchers have examined information diffusion from different points of view which are appropriate to predict the propagation behavior of users on social media.

Answer to question5:

For answering RQ5, based on the knowledge gained of reviewed studies, information diffusion process, information diffusion models, information diffusion types, popular information detection and identification of influential users are effective characteristics of information diffusion aspect in UBMSM.

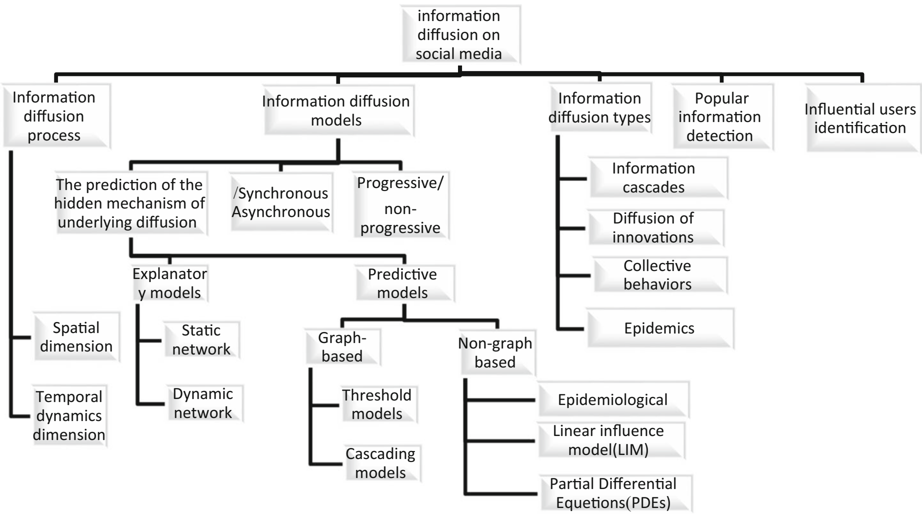


Fig. 13 A taxonomy of significant characteristics of information diffusion

Answer to question6:

For answering RQ6 in terms of information diffusion, a taxonomy of the significant characteristics of information diffusion on social media is presented in Fig. 13. In following, we take in account information diffusion process and information diffusion models in more detail.

4.7.1 Information diffusion process

Diffusion is the process by which an innovation is communicated through certain channels over time among the members of a social system [147]. This phenomenon has been widely studied in a vast area, such as viral marketing, information propagation on blogs, epidemiology and social behavior. Empirical studies about information diffusion in social networks started in the middle of the twentieth century. An information diffusion process that takes place in social media involves four main elements, a piece of information carried by messages, a transmitter node or a set of transmitter nodes who initiate information diffusion process, a receiver node or a set of receiver nodes who receive information and communication channel through which information diffusion takes place. A node in information diffusion process takes place only in one of two states of active or inactive over time. When a node receives a piece of information becomes active. We can describe the diffusion process by a sequential activation of nodes throughout network named activation sequence. In other words, an activation sequence consists of an ordered set of nodes which adopt a piece of information over time. Activation sequences provide knowledge about where and when a piece of information is diffused. Information diffusion process in social media is characterized by two significant dimensions which are spatial dimension and temporal dynamics dimension. Spatial dimension is a concern to structure of diffusion such as diffusion graph (who influenced whom) and specifies the effect of influential users in the information diffusion modeling. Temporal dynamics dimension is concern to the evolution of the diffusion rate over time due to interaction of a large amount of users in social media. The evolution of diffusion rate indicates how many users are adopted with a piece of information diffused over time. There are three levels for studies done in information diffusion area over online social networks which are empirical study, macroscopic modeling and microscopic modeling levels. The existed empirical studies on different social networks indicate intrinsic patterns of information diffusion dynamics in them. For example, [104] investigated an Empirical Study of Spread of information on Digg and Twitter Social Networks. [171] indicated a large-scale empirical study on content diffusion process, user features and network structure on Digg social network. [165] reviewed the factors that show the epidemic transmission of popular news posted on Digg. The macroscopic modeling deals with modeling information diffusion from global viewpoint. [55, 167, 201] used mathematical modeling for information diffusion in implicit networks. [153] applied SIS (susceptible, infected, and susceptible) epidemic model to identify influential users in social networks during a given time period. [181] presented first Partial Differential Equation (PDE) based diffusion model called diffusive logistic model that focuses on both spatial and temporal dimensions. The microscopic modeling focuses on understanding microscopic level of user interactions in social networks. In other words, information diffusion prediction and identification of influential users depend on precision of interactions. [94] used two basic diffusion models namely linear threshold and Independent Cascade models to find influential users in social networks. Information diffusion is occurred in social networks due to the effects of internal or external influential sources. Hereafter, we consider a social network as a closed world that ignores the effect of external influence and focuses on internal influential sources which are social platform diffusivity, information popularity and influential users. Therefore, three main questions in information diffusion area are raised which are as follows:

1- How, why and through which paths information is diffusing, and will be diffused in the future?
2- Which topics or pieces of information are more popular?
3- Which users of network are more influential users and play a major role in information diffusion process?

In response to above questions, it is necessary to explain diffusion models that can capture and predict the hidden mechanism underlying diffusion.

4.7.2 Information diffusion models

Information diffusion models aim to model information diffusion process in social media. They provide knowledge about how and why information diffuses. Generally, we can distinguish between different information diffusion models from the various perspectives which are progressive/non-progressive, Synchronous/Asynchronous and the prediction of hidden mechanism underlying diffusion process perspectives.

Progressive/non-progressive perspective The substantially difference between progressive and non-progressive version of diffusion models is that in the progressive version when a node becomes active remains active forever while in the non-progressive version a node may deactivate-activate again and again [126].

Synchronous/asynchronous perspective Diffusion process in the real world involves time evolution and information spreads along the continuous time axis with asynchronous time delays. There are two various types of time delay called as link delay and node delay [152]. Link delay corresponds to information propagation delay and node delay corresponds to the delay associated with user action. Further, two types of node activation scheme from the multiple parents associated with time delay taken into account: non-override and override. Non-override updating scheme depends on the initial decision when to activate and override ones decides to activate multiple times when one of the parents becomes newly activated. The presented basic models for information diffusion do not consider time delay and nodes change their states synchronously at discrete in time steps. Therefore, it is necessary to deal with asynchronous time delay for modeling of information diffusion in the real world. [154] extended two basic models called AsIC and AsLT by incorporating asynchronous time delay to make them realistic enabling effective use of observed information diffusion data. Finally they applied these methods to behavioral analysis of topic propagation using the real blog propagation data.

The prediction of hidden mechanism underlying diffusion process perspective The diffusion models developed in social networks, consider users are influenced by actions taken by their neighbors. In other words, information diffuses due to informational cascades. Informational cascade shapes a directed tree that the first node of the activation sequence is as a root node of tree. The tree indicates who influence whom and spreads information according to activation sequence. In the recent studies, information diffusion models are generally categorized into explanatory and predictive models.

The main objective of explanatory models is to retrace the spreading path of the information in social networks and aim at understanding how information is propagated. These models try to infer spreading properties of social media such as pairwise transmission rate, pairwise transmission probability and cascade properties. Further, the network over which propagations take place is usually unknown and unobserved. [60] proposed NETINF, a scalable algorithm for inferring networks of diffusion and influence. They used a generative probabilistic model of cascades and indicated how, on a fixed hypothetical network, information spread as directed trees through the network. They infer the network connectivity using submodular optimization by considering only the most probable directed tree supported by each cascade. [146] developed a flexible model, NETRATE, of the spatiotemporal structure underlying diffusion processes. They introduced continuous temporal dynamics, allowing variable transmission rates across edges, and avoided further assumptions dramatically simplified the problem compared with CONNIE model introduced by [127] and NETINF model. NETRATE provides a unique solution to the network inference problem with high recall, precision and accuracy. They illustrated that NETRATE performs comparably to NETINF and it outperforms CONNIE on precision and recall. NETRATE and CONNIE infer not only the network connectivity but also transmission rates of infection or prior probabilities of infection using convex optimization by considering all possible directed trees supported by each cascade. Most of these existing models consider that the underlying network structure is static and don't change over time but it is not a satisfying assumption in the real world, because the topology of networks evolves over time. [145] proposed an efficient approximation algorithm, MultiTree, with provable near-optimal performance that solve an open problem on network inference from diffusion cascades first introduced by [60]. They indicated the network inference problem as a submodular optimization problem in which they did not consider only the most probable directed tree as in NETINF but all directed trees supported by each cascade as in CONNIE and NETRATE. By considering all trees, MultiTree algorithm was able to infer the network of diffusion more accurately than NETINF when the number of observed cascades was small compared to the network size. Moreover, in MultiTree algorithm, was applied the greedy algorithm for submodular maximization in contrast to convex optimization, that was considerably faster than CONNIE and NETRATE. In terms of precision and recall MultiTree was able to gain higher recall values in comparison to NETINF, CONNIE and NETRATE. In terms of accuracy MultiTree outperformed NETINF and accuracies of CONNIE and NETRATE were typically significantly lower. [59] proposed an on-line dynamic network inference algorithm, INFOPATH, that allowed to infer daily networks of information diffusion between online media sites over a 1 year period using more than 179 million different contagions diffusing over the underlying media network. This model allows information to propagate at different rates across different edges by adopting a data driven approach, where only the recorded temporal diffusion events are used. In comparison, INFOPATH is faster than NETRATE and as fast as NETINF, while achieving the same accuracy as NETRATE. Importantly, INFOPATH and NETRATE always improve accuracy with the running time, until convergence. [127] presented a general solution to the problem of inferring latent social networks from the information diffusion. They applied a maximum likelihood approach based on convex programming for optimizing of the problem. They illustrated their approach was more general and robust in comparison to NETINF. [6] proposed a simple and efficient network inference method, First-Edge, that was focused on the number of traces that network inference tasks require, called as trace complexity of the problem. [150] investigated the problem of missing data in information cascades. By considering only a fraction of the complete cascade, their goal was to estimate the properties of the complete cascade such as its size or depth. They Proposed an analytical k-tree model of cascades and rigorously derived a number of their important properties. [40] investigated the network structure inference problem for a general family of continuous-time diffusion models using likelihood maximization framework. Table 8 in appendix, shows a classification of surveyed explanatory models in terms of underlying network structure, optimization method, inferred properties and ability to work with incomplete data.

The second category of models, predictive models, predicts how a specific diffusion process would unfold in a given social network, based on temporal and spatial dimensions by learning from past diffusion traces. These models are generally classified into graph based and non-graph based models. Graph based models consider that the underlying network structure is known and predict who influence whom but these models don't work in implicit networks. There are two basic graph-based predictive models which are Linear Threshold (LT) and Independent Cascades(IC) (Goldenberg et al., 2001) models. These models consider the structure of network is static and they focus on the structure of the diffusion process based on a directed graph where each node can be activated or deactivated. Both models depend on the non-realistic assumption that information diffuses in a synchronous way along a discrete in time-axis. Further, these models have parameters such as diffusion probabilities for the IC model and an influence degree defined on each edge and an influence threshold of each node for the LT model which must be estimated. These models are able to predict who will influence whom. Threshold models were applied as early as 1970s by Granovetter et al. A LT model is an example of a threshold model. In the LT model at each time-step, an inactive node is activated when the summation of influence degree of its active parent neighbors is more than influence threshold of that node. In IC model when a node becomes active at a given time-step it has a single chance to activate each of its inactive child neighbors with a probability in the next time-step. Moreover, if multiple parent nodes at a given time-step be able to activate a given node, then their activation attempts are sequenced in an arbitrary order. The process continues until no further activations are possible. IC model is sender-centric that is sender node decides to activate receiver node while LT model is receiver-centric, means that the receiver nodes get activated by multiple sender nodes. Non-graph based models ignore the topology of the network and just predict the evolution of information diffusion rate. There are three different types of non-graph based predictive models namely epidemic model, Linear Influence Model (LIM) and Partial Differential Equations (PDEs). In epidemic models nodes take place into different states. The aim of these models is that illustrate the evolution of the proportions of nodes in each state over time. SIS and SIR are two main epidemic models where S stands for "susceptible", I for "infected" and R for recovered. Nodes in state S switch to state I with a fixed probability for each model. Also in the SIS model nodes in state I switch to state S with a fixed probability but in the SIR model, nodes always switch to state R. In these models it is assumed that each node connects to another in the population randomly with same probability. [201] proposed a Linear Influence Model that focused on modeling the global influence of a node on information diffusion rate in implicit network over time. This model doesn't require any knowledge of the network topology and subsequently, doesn't formulate a problem of predicting which node will influence which other nodes in the network. They modeled the number of newly influenced nodes as a function of the times when other nodes been influenced in the past. They modeled influence functions in a non-parametric way and their method of model parameter estimation was based on simple least squares procedure. [185] modified Linear Threshold model and proposed a new model called LT-MLA model that considers multi-level attitude. They considered the positive and negative effect of user influence in interaction relationship between users. Also, [72] extended Linear Threshold model to model negative influence considering distrust relationships between users. They studied Positive Opinion Influential Node Set (POINS) selection problem and presented a greedy algorithm called POINS-GREEDY to solve POINS problem. [75] studied the effect of trust and distrust relationships on influence of users in the signed social networks. They presented a sign-aware cascade model (SC) to model information diffusion. [76] proposed some new models based on Independent cascade and Threshold models for information diffusion. They considered distrust relationships in their models and classified information diffusion models based on three schemes. [204] introduced a new dynamical information diffusion model considering social tie strength and self-confirming mechanism. They Extended SIR model. They divided individuals into four groups which are ignorants (susceptible), hesitators (uncertain), spreaders (infected), and stiflers (removed). [198] proposed a dynamic model for information diffusion based on multidimensional complex network space and social game. Their model divided into three parts namely static driving mechanism quantification, dynamic driving mechanism quantification, and model simulation. [180] presented the game choice information propagation model based on page rank (GCIP-Page Rank). They used social factors such as social trust, opportunity and game choice motivation based on triadic closure principle in social network for constructing their model. [208] proposed Interaction-Aware Diffusion (IAD) framework to understand contagion adaption behaviors of users in social networks. They applied a set of interactions between users, contagions' contents and sentiments. Their proposed framework considers three kinds of interactions which are user-user interaction, contagion-contagion interaction and user-contagion interaction. They presented a LDA-S model to discover the sentiment distribution and topic distribution from contagions. [86] introduced an information diffusion model based on superposition theory on Sina Weibo microblog. They analyzed data collected from Sina Weibo based on the network structure, the characteristics of the information diffusion and users' behaviors. The results showed that there is a relationship between the number of reposts, hop count and time of post. They improved the accuracy of their model considering the impact of users' activity behaviors. [109] presented a stochastic model of the information diffusion process namely Susceptible View Forward Removed (SVFR) model based on the WeChat information diffusion mechanisms. Their model is based on classic SIR model and captures three key features of cascade trees such as the average path length, the degree variance and the distribution of the sizes of the cascade trees. They applied the Random Recursive Trees (RRTs) to model the cascade trees, without considering the underlying dynamics of users. [212] proposed an Adoption Behavior-based Graphical Model (ABGM) to model the dynamic process of diffusion. They applied two types of correlation between users' adoption behaviors namely Homophily and heterophily to promote the accuracy of innovation diffusion. Table 9 in appendix, illustrates a comparison of predictive models in terms of different features in detail.

5 Conclusion and future works

In this paper, we have presented a systematic literature review based on existing studies of UBMSM. The results illustrate the real history of UBMSM. First, we identified the distributions over popular publisher in this field. We observed the majority of studies were published by famous publisher, ACM. Also, we presented the greatest fraction of studies were published by well-known journals that was about 47%. Next, we classified the studies of UBMSM based on their focus area into four main aspects namely user aspect, content aspect, network structure aspect and information diffusion aspect. Based on the presented classification, we illustrated distribution of studies per category. The results show most studies were concentrated on user aspect that specifies the importance of this aspect in UBMSM. Moreover, we introduced separately the effective characteristics of each aspect and realized the appropriate classifications and taxonomies for them that affect UBMSM, aiming to eliminate barriers to future research efforts, such as predicting of the user behaviors on social media. With respect to the findings in this paper, investigating the interplays between these four aspects is of utmost importance for the planning of future work.

Appendix

Table 6 Approaches on inferring topical interests of users

Journal/ Conference	Author(s)	Types of data used	Method	Representation of user interests	Weighting Schema	Enrichment source	Extending strategies	Temporal dynamics	Datasets	Multiple social media
Springer	[209]	Explicit data	Twitter-LDA model	Topic model (content-based)	TF	User tweets	–	–	Twitter traditional news media	–
ACM	[34]	Social Connections and Explicit data	Incorporating the popularity component in to LDA and using Three refined LDA models.	Topic model+ graph features based (Hybrid)	TF Polya-Urn, Two path and weight representation of LDA	Popularity component	Considering reader-side popularity as well as the writer-side popularity	–	Twitter	–
ACM	[118]	Explicit data	without modifying the basic machinery of LDA through various pooling schemes	Topic model (Content-based)	TF-IDF	User tweets, hashtag-based pooling schema	pooling schemes and automatic TF similarity-based hashtag assignment	–	Twitter	–
ACM	(Ramage et al. 2009)	Explicit, implicit, Social Connections data	Scalable labeled LDA (Tweet-level LDA)	Topic model+ user features-based+ graph features-based	TF-IDF	User tweets	–	–	Twitter	–
ACM	(Pennacchiotti et al. 2011)	Explicit data	User-level LDA	Topic model	TF-IDF	Tweets posted by users	–	–	Twitter	–
Other	(Xu et al. 2011)	Explicit data	A modified author-topic model	Topic model	TF	Tweets posted by users	–	–	Twitter	–
ACM	[193]	Social connections, and Explicit data	using LDA to discovery his latent topics of interest	Topic model+ graph features-based	TF	Tweets posted by users	–	–	Twitter	–
ACM	[140]	Explicit and Social connections data	Unigrams are extracted from list meta-data and Wikipedia categories	User features-based	The weighted average of both list meta-data and Wikipedia scores	List meta-data	By leveraging the category structure of Wikipedia	–	Twitter	–
ACM	[9]	Explicit, implicit data	A Bayesian approach (a time-varying LDA	Topic model+ user features-based	Time-sensitive	–	–	An exponential time decay function	Yahoo	–
ACM				Topic model +TF-ICF	TF-ICF		–	–	Twitter	–

Table 6 (continued)

Journal/ Conference		Author(s)	Types of data used	Method	Representation of user interests	Weighting Schema	Enrichment source	Extending strategies	Temporal dynamics	Datasets	Multiple social media
		(Han et al. 2016)	Explicit and implicit data	A hybrid approach by integrating TF-ICF and topic models.			News media categories				
	ICWSM	[132]	Explicit data	L-LDA	Topic model	–	Tweets posted by a given user	–	Long-term	Twitter and Pinterest	*
	ACM	[211]	Explicit, implicit and social connections data	Composite social topic model (ComSoc)	Topic model+ user features-based+ graph features-based	Representation of transferring the relationship knowledge	Interactions between users and items in different networks	–	–	Tencent and Douban	*
ACM	ACM	[56] [158]	Explicit data	Extracting frequently occurring topics from the List meta-data and associating Topics with the listed users	User features-based	Frequency of occurrence of the topic	The meta-data of crowd sourced lists that contain the user	–	–	Twitter	–
	ACM	[23]	Explicit, Implicit and Social Connections data	Twitter Lists and inferring user interests based on number of topical experts	User features-based+ graph features-based	Ranked list of topics	Topical expertise of a given user's followings	–	–	Twitter	–
	ACM	[69]	Explicit& Social Connections data	Bag of Words from tweets	Bag of Words + graph features-based	TF-IDF	Follower and followee tweets	–	–	Twitter	–
	ACM	[35]				TF	Received tweets from user's followee	–	–		
	ACM	[125]	Explicit and Social connections data	a method based on community detection	Bag of Words + graph features-based+ user features-based	TF	The Content of user social actions in combination with the social network graph	–	–	Facebook	–
	Springer	[2]	Explicit data	Using the TF × IDF method to measure similarity between user tweet and news articles	Concept-based	TF-IDF	Linked data(semantics extracted from news articles)	–	–	Twitter	–
	Springer	[4]	Explicit data	DBpedia to extend user profiles with respect to Point of Interests	Concept-based	TF	Linked data (DBpedia)	–	–	Twitter	–
	Springer	[3]	Explicit & implicit data	Extracting Entity, Hashtag and category-based	Concept-based	Time-sensitive	News articles	–	By considering the temporal patterns	Twitter	–

Table 6 (continued)

Journal/ Conference	Author(s)	Types of data used	Method	Representation of user interests	Weighting Schema	Enrichment source	Extending strategies	Temporal dynamics	Datasets	Multiple social media
			user profiles with semantic enrichment and temporal dynamics					(short-term, long-term) profiles		
ACM	(Abel et al. [1]	Explicit & implicit data	Entity and Hashtag-based user profiles	Concept-based	Time-sensitive (a temporal decay for the occurrence frequency)	News articles	–	interest decay function (Short-term and Long-term adopters)	Twitter	–
ACM	[130]	Explicit and social connections data	entity extraction algorithm on every message and social activity, using DBpedia to link to its entities and to extract related categories	Concept-based+ graph features-based	Time-sensitive	Linked data(DBpedia)	Category-based user profiles using DBpedia with the discounting strategy and time decay function	Exponential time decay function(Short time and Long time)	Twitter & Facebook	*
ACM	(Piano et al. 2016c)	Explicit data	The combination of category, class and property-based extension strategies using DBpedia	Concept-based	CF-IDF	DBpedia	lever-aging different types of information from DBpedia and using a discounting strategy	interest decay function (Long-term (Ahmedα)a modified version of Long-term(Ahmed))	Twitter	–
ACM	(Piano et al. 2016a)	Explicit data	aggregating strategy of user profiles from different OSNs	Concept-based	CF-IDF& aggregated global weight	DBpedia	combined user profiles of entity-and category-based profiles using DBpedia with the discounting strategy	–	Twitter& Google+	*
ACM	(Piano et al. 2016b)	Explicit data	Category-based extension strategy	Concept-based	TF-IDF	synsets from WordNet and concepts from DBpedia	WordNet synsets, category information from DBpedia and a discounting strategy	–	Twitter	–
ACM	[113]					Wikipedia		–	Twitter	–

Table 6 (continued)

Journal/ Conference	Author(s)	Types of data used	Method	Representation of user interests	Weighting Schema	Enrichment source	Extending strategies	Temporal dynamics	Datasets	Multiple social media
		Explicit and social connection data	Use of Explicit Semantic Analysis algorithm	Concept-based + graph features-based	Explicit Semantic Analysis		random walk on the Wikipedia concept graph			
ACM	(Michelson et al. 2010)	Explicit data	Classifying entities in the tweets by leveraging Wikipedia as a knowledge base	Concept-based	CF-IDF	By querying Wikipedia	–	–	Twitter	–
ACM	[210]	Explicit and implicit data	A behavior factorization approach	Topic model+ user feature-based (Hybrid)	by the number of occurrences of topic i in all user u's behaviors, divided by the sum of occurrences of all items.	Google's Knowledge Graph	–	–	Google+	–
Springer	[89]	Explicit data	Hierarchical Interest Graph	Concept-based	TF-IDF	Wikipedia	–	–	Twitter	–
HP labs	(Ramanathan et al. 2009)	Explicit data	A hierarchical user profile using Wikipedia.	Concept-based	CF	Wikipedia	An Exponential decay function	–	web pages from the Internet Explorer	–
Springer	(Tao et al. 2011)	Explicit data	A Twitter-based User Modeling Service (TUMS)	Concept-based	TF-IDF or TF	News articles	–	Considering the Temporal patterns	Twitter	–
Elsevier	[97]	Explicit, implicit and social connection data	extracting terms, mining frequent patterns, and pruning patterns	Concept-based+ graph features-based	TF-IDF	Personalized term patterns of like-minded users	–	–	NSF (National Science Foundation) and Movie Lens	–
ACM	(Uysal et al. 2011)	Explicit and Social connections data	Generating user profiles based on information regarding users' followings/followers on Twitter.	Content-based+ Graph features-based+ user features-based	TF-IDF	Users' Follower/-Followings	–	–	Twitter	–
Springer	[47]	Social connections, implicit	PGPI algorithm to infer user profiles under the constraint of a partial social graph and without training.	Content-based	–	group and publication information (views and likes)	–	–	Facebook	–
Elsevier		Explicit data		Concept-based	CF		–		Twitter	–

Table 6 (continued)

	Journal/ Conference	Author(s)	Types of data used	Method	Representation of user interests	Weighting Schema	Enrichment source	Extending strategies	Temporal dynamics	Datasets	Multiple social media
		(Jiang et al. 2015)		A semantic enrichment method and employ the topic hierarchy tree (THT) model			multi-sources (dbpedia, freebase, yago, website)		Short-term and long-term		
Other		[29]	Explicit, social connections data	A probabilistic generative model of user utterances	Topic model+ graph features-based+ user features-based	Time-sensitive	A taxonomy derived from the Open Directory Project (ODP)	–	The prior distribution over interests at every time step is updated	Twitter	–
	Springer	[80]	Explicit data	based on Biterm Topic Model	Topic model	Time-sensitive	Using a Biterm	–	Using time windows and decay function	Sina Weibo	
	ACM	[106]	Explicit data	classifying the relative interests of Twitter users using Wikipedia	Graph features-based	Weighting in relation to his/her interest in other catego-ries.	Wikipedia	–	–	Twitter	–
	Springer	[46]	Explicit data	Sequential labeling model trained with automatically constructed labeled data	Tag-based+User features-based	TF-IDF	the extracted interest tags from biographies	–	–	Twitter	–
	Springer	[183]	Implicit, Explicit and social connections data	Random-walk based mutual reinforcement model combining both text and link information	Topic model+User features--based+graph features based	Based on text information and link information	Text information and behavior information	–	–	Tencent weibo	
Springer		[5]	Explicit, implicit data	Cross-system user modeling strategies that interweave user profiles from diverse Social Web systems	Tag-based+ user features-based	TF	Wordnet	–	–	Twitter Facebook LinkedIn, Delicious, Flicker and Google	*
	Springer	[70]	Explicit and social connections data	Categorizing users and their social ties according to a collection of curated topical categories	Tag-based+graph features-based	TF	Followers and followees tags	–	–	Twitter	*
	IEEE	[120]	Explicit, implicit, social connections data	Relying on combining information tags, users and resources	Tag-based+ graph features-based+ user features-based	Time-sensitive	metadata of the resources	–	Detecting relevant interests in each time Period	Delicious	–
	ACM	[74]	Explicit, implicit, social	Co-Factorization Machines (CoFM)	Topic model+ graph and user features	TF		–	–	Twitter	–

Table 6 (continued)

Journal/ Conference	Author(s)	Types of data used	Method	Representation of user interests	Weighting Schema	Enrichment source	Extending strategies	Temporal dynamics	Datasets	Multiple social media
		connections data				Users' Tweets, co-followees, co-followers				
Elsevier	[11]	Implicit data	Hierarchical User Sentiment Topic Model (HUSTM)	Hierarchical topic model+users' sentiment information	TF	users' sentiment information	–	–	LAPTOPS, DIGITALSLRS and Smartphone	–
Springer	[157]	Explicit, and social connections data	Using network, activity graph and semantic web	weighted aggregation of entities and RDF graph of DBpedia ontology	Time-sensitive	DBpedia	–	Short-term	Twitter	–
Springer	[88]	Explicit data	biGRU-CNN method	*k*-dimensional word-embedding vector	Time- sensitive	External knowledge base- Sentiment analysis	–	Short-term	Facebook an Twitter	–
Springer	[206]	Explicit data	Topic-Association Mining	LDA topic model+Wikipedia concepts	Time-sensitive	Collective knowledge	–	Short-term	Twitter	–
Intelligent data analy-sis	[187]	Explicit and social connections data	User-Networked Interest Topic Extraction	LDA topic model	TF and the influences of all followees of a given user	–	–	–	Sina Weibo	–
Elsevier	[206]	Explicit, Implicit and social connections data	graph-based link prediction schema	LDA topic model+ Semantic concept+ Social graph	Weighted Path Count (WPC)	Wikipedia	–	–	Twitter	–
Springer	[207]	Explicit and Implicit data	Twitter-LDA	LDA topic model+Wikipedia concepts+users' historical activities	Time-sensitive	Wikipedia	–	Short-term	Twitter	–
Springer	[135]	Explicit, Implicit and social connections data	a temporal preference model	LDA topic model+centrlity measure	Time-sensitive	centrlity measure	–	Short-term and Long-term	Twitter	–
arXiv pre-print arXiv	[78]	Explicit, implicit and social connections data	Probabilistic longitudinal model, CosTot	Topic model+ graph features-based	Time-sensitive	User posts, time, and network information	–	Multi-scale	Sina Weibo	–

Table 6 (continued)

Journal/ Conference	Author(s)	Types of data used	Method	Representation of user interests	Weighting Schema	Enrichment source	Extending strategies	Temporal dynamics	Datasets	Multiple social media
ACM	[205]	Explicit, implicit and social connections data	A dynamic temporal context-aware mixture model (DTCAM)	Topic model+ graph features-based	Time-sensitive	Users posts, social information in a social network	–	Decay factor	Digg, MovieLens,-Douban Movie, and Delicious	–
Springer	[114]	Explicit and implicit data	multi-source personal information fusion and semantic reasoning	Concept-based	Time-sensitive	hierarchical relations between users' explicit interests and implicit	–	Considering the information update frequency of each source	Twitter, Facebook and LinkedIn	*
Springer	[96]	Explicit and implicit data	Merging interests across different sources and users	Concept-based+ user features-based	CF	using synonyms of words or metadata of URLs (The categories from the ODP directory)	–	–	online social network profiles and web browsing traces	*
ACM	[133]	Explicit and social connections data	The Authority Learning Framework (ALF)	User features-based	the z-score with the topical popularity of the account	Wikipedia	–	–	Instagram	–
Springer	[108]	Explicit and social connections data	A probabilistic generative model	Topic model+ Graph features-based	TF	–	–	–	Renren, Twitter, Digg	–

Table 7 A review on a variety of user influence measures

Journal/Conference	Author(s)	Measure	Underlying network structure	Topic-sensitive-based (yes/no)	Activity-based(yes/no)	Underlying dynamic process(yes/no)	Dataset
ACM	(Ghosh et al. 2010)	• empirical user influence estimation	Static	No	No	Yes	Digg
ICWSM	[33]	• Number of followers • Number of retweets • Number of mentions	Static	No	Yes	No	Twitter
	(Kwak et al. 2010)	• Number of Followers • Number of retweets • Page Rank	Static	No	Yes	No	Twitter
ACM	[18]	• Number of URL reposts	Static	No	Yes	No	Twitter
ACM	[193]	• TwitterRank • Page Rank • Number of followers	Static	Yes	Yes	No	Twitter
Springer	[148]	• User passivity • Number of retweets	Static	No	Yes	Yes	Twitter
ICWSM	[115]	• User's LIWC category score • user reposts	Static	Yes	Yes	No	Twitter
ACM	[170]	• Topical Affinity Propagation(TAP)	Static	Yes	No	No	Twitter
ACM	[62]	• Influenceability score	Dynamic	No	Yes	Yes	Twitter
	(Jingxuan [105])	• Time-window diffusion size • Temporal closeness centrality	Dynamic	No	Yes	Yes	Twitter
PloS one journal	(Weng et al. 2015)	• by measuring topical diversity of a user's interests	Static	Yes	No	No	Twitter
IEEE	[12]	• Users' structural location in a network • Users' attributes	Static	No	Yes	No	Flicker
Elsevier	[214]	• the number of boards created by user • How active the user is in posting pins to a board he/she created	Static	Yes	Yes	No	Pinterest

Table 7 (continued)

Journal/Conference	Author(s)	Measure	Underlying network structure	Topic-sensitive-based (yes/no)	Activity-based(yes/no)	Underlying dynamic process(yes/no)	Dataset
		• number of followers as well as the influence of each of the followers • How engaged the follower of a board is,					
Elsevier	[213]	• user trust network • user interest degree	Static	No	Yes	Yes	Epinions
Elsevier	[101]	• Number of following • longer tweets • sentiment (either positive or negative) on • tweets • mentions on tweets • embedded links on tweets	Static	No	Yes	No	Twitter
Elsevier	[30]	• Semantic profiles	Static	Yes	Yes	Yes	Twitter
Springer	[215]	• SIRank	Static	No	Yes	Yes	Sina Weibo
Elsevier	[13]	• Personalized PageRank(PPR)	Static	Yes	Yes	No	Twitter
Pattern recognition and artificial intelligence	[166]	• Based on Bayesian network and PageRank algorithm	Static	Yes	Yes	No	Sina Weibo
Information science	[159]	• A hybrid method based on k-shell measure	Static	No	No	No	NetHEPT, Gnutella and Brightkite

Table 8 A classification of surveyed explanatory models

Journal/ Conference	Author/year	Model	Optimization method	Network(static/ dynamic)	Inferred properties			Ability to work with incomplete data
					Cascade properties	Pairwise transmission probability	Pairwise transmission rate	
ACM	[146]	NETRATE	Convex optimization	Static	*	*	*	–
ACM	[59]	INFOPATH	Convex optimization	Static, Dynamic	*	*	*	–
ACM	[150]	K-tree model	–	Static	*	–	–	*
ACM	[60]	NETINF	Submodular optimization	Static	*	*	–	–
ACM	[60]	MultiTree	Submodular optimization	Static	*	*	*	–
ACM	(Myers et al. 2010)	CONNIE	Convex optimization	Static	*	*	*	–
ACM	[6]	First-Edge	–	Static	*	–	*	–
Elsevier	[175]	A novel cascade model	–	static	*	–	–	–

Table 9 The comparison of predictive models

Journal/ Conference	(Author, year)	Model	Model type	Network structure	Network(Explicit/ Implicit)	Temporal dynamics(evolution of the diffusion rate)	Activation time
Journal	[64]	LT	Graph-based	Static	Explicit	–	Discrete in time
ACM conference	[94]	IC	Graph-based	Static	Explicit	–	Discrete in time
ACM conference	[152]	AsLT	Graph-based	Static	Explicit	–	Continuous-time
ACM conference	[152]	AsIC	Graph-based	Static	Explicit	–	Continuous-time
ACM journal	[66]	T-BASIC	Graph-based	Static	Explicit	*	Continuous -time
ACM journal	[129]	SIR,SIS	Non-graph-based	–	Implicit	*	Discrete in time
IEEE conference	[201]	LIM	Non-graph-based	–	Implicit	*	Discrete in time
IEEE conference	[181]	PDE	Non-graph-based	–	Implicit	*	
IEEE conference	[77]	Improved Hydro-IDP	Graph-based	Dynamic	Explicit	*	Continuous time
IEEE journal	[20]	Component-based IC	Graph-based	Static	Explicit	–	Discrete in time
IEEE conference	[121]	TS-IDM	Graph-based	Static	Explicit	–	Continuous time
IEEE journal	[85]	Based on graphical evolutionary game theory	Graph-based	Dynamic	Explicit	*	Discrete in time
IEEE conference	[43]	PDE-based diffusion model	Non-graph based	–	Implicit	*	Discrete in time
IEEE conference	[182]	Linear Diffusive Model	Non-graph based	–	Implicit	*	Discrete in time
Elsevier journal	[107]	Dynamic model based on SIR model	Non-graph based	–	Implicit	*	Discrete in time
Springer conference	(Wang et al. 2016a)	LT-MLA model	Graph-based	Static	Explicit	–	Discrete in time
Springer journal	[72]	Extended LT	Graph-based	Static	Explicit	–	Discrete in time
	[75]	Extended IC	Graph-based	Static	Explicit	–	Discrete in time

Table 9 (continued)

Journal/ Conference	(Author, year)	Model	Model type	Network structure	Network(Explicit/ Implicit)	Temporal dynamics(evolution of the diffusion rate)	Activation time
Intelligent data analysis							
Elsevier journal	[76]	Threshold-based Model and Cascade-based model	Graph-based	Static	Explicit	–	Discrete in time
Elsevier journal	[204]	Susceptible-Uncertain-Infected-Removed(SUIR)	Graph-based	Static	Explicit	*	Discrete in time
Elsevier journal	[198]	Extended SIR model	Graph-based	Static	Explicit	*	Continuous in time
Elsevier journal	[180]	GCIP-Page Rank model	Graph-based	Static	Explicit	–	Discrete in time
IEEE journal	[208]	IAD framework	Graph-based	Static	Explicit	–	Discrete in time
Springer journal	[86]	superposition theory based	Graph-based	Static	Explicit	–	Discrete in time
Elsevier journal	[109]	SVFR model	Non-graph based	–	Implicit	*	Discrete in time
Springer conference	[212]	Adoption Behavior-based Graphical Model	Graph-based	Static	–	–	Discrete in time

Journal/ Conference	Extension strategy	Mathematical modeling(parametric/ Non-parametric)	Synchronous/ Asynchronous	Time delay	Parameter setting	Spatial dimension	Topic-sensitive
Journal	–	Parametric	Synchronous	–	Fixed	*	–
ACM conference	–	Parametric	Synchronous	–	Fixed	*	–
ACM conference	–	Parametric	Asynchronous	*	Fixed	*	–
ACM conference	–	Parametric	Asynchronous	*	Fixed	*	–
ACM journal	–	Parametric	Asynchronous	*	Dynamic	*	–
ACM journal	–	Parametric	Synchronous	–	Fixed	–	–
IEEE conference	Time decay	Non-parametric	Synchronous	–	dynamic	–	*
IEEE conference	Time decay	parametric	Synchronous	–	dynamic	*	–
IEEE conference	–	Non-parametric	Synchronous	–	Fixed	*	–
IEEE journal	Time decay factor, structural diversity factor	Parametric (EM)	Synchronous	–	dynamic	*	–

Table 9 (continued)

Journal/ Conference	Extension strategy	Mathematical modeling(parametric/ Non-parametric)	Synchronous/ Asynchronous	Time delay	Parameter setting	Spatial dimension	Topic-sensitive
IEEE conference	–	Parametric	Asynchronous	*	Fixed	*	*
IEEE journal	–	Parametric	Synchronous	–	Fixed	–	–
IEEE conference	–	Parametric	Synchronous	–	Dynamic	*	–
IEEE conference	Time decay and heterogeneity of spreading power of users	Parametric	Synchronous	–	Fixed	*	–
Elsevier journal	Link rewiring	Parametric	Synchronous	*	Fixed	–	–
Springer conference	Distrust relationships	Parametric	Synchronous	–	Fixed	*	–
Springer journal	Distrust relationships	Parametric	Synchronous	–	Fixed	*	–
Intelligent data analysis	Distrust relationships	Parametric	Synchronous	–	Fixed	*	–
Elsevier journal	Distrust relationships	Parametric	Synchronous	–	Fixed	*	–
Elsevier journal	–	Parametric	Synchronous	–	Dynamic	*	–
Elsevier journal	user’s social behavior and psychological characteristics	Parametric	Synchronous	–	Dynamic	*	*
Elsevier journal	Key social factors such as opportunity, trust and motivation	Parametric	Synchronous	–	Fixed	*	–
IEEE journal	Three kinds of interactions	Parametric	Synchronous	–	Fixed	*	*
Springer journal	–	Parametric	Synchronous	*	Fixed	*	–
Elsevier journal	Non-liner and non-homogeneous probability for a user	Parametric	Synchronous	–	Fixed	*	–
Springer conference	Homophily and heterophily	Parametric	Synchronous	–	Fixed	*	–

References

1. Abel F, Gao Q, Houben G-J, Tao K (2011a) Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web. Paper presented at the proceedings of the 3rd international web science conference
2. Abel F, Gao Q, Houben G-J, Tao K (2011b) Analyzing user modeling on twitter for personalized news recommendations. Paper presented at the international conference on user modeling, adaptation, and personalization
3. Abel F, Gao Q, Houben G-J, Tao K (2011c). Semantic enrichment of twitter posts for user profile construction on the social web. Paper presented at the Extended Semantic Web Conference
4. Abel F, Hauff C, Houben G-J, Tao K (2012) Leveraging user modeling on the social web with linked data. Paper presented at the International Conference on Web Engineering
5. Abel F, Herder E, Houben G-J, Henze N, Krause D (2013) Cross-system user modeling and personalization on the social web. *User Modeling and User-Adapted Interaction*:1–41
6. Abrahao B, Chierichetti F, Kleinberg R, Panconesi A (2013) Trace complexity of network inference. Paper presented at the proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining
7. Adamic LA, Glance N (2005) The political blogosphere and the 2004 US election: divided they blog. Paper presented at the proceedings of the 3rd international workshop on link discovery
8. Aggarwal CC, Subbian K (2012) Event detection in social streams. Paper presented at the proceedings of the 2012 SIAM international conference on data mining
9. Ahmed A, Low Y, Aly M, Josifovski V, Smola AJ (2011) Scalable distributed inference of dynamic user interests for behavioral targeting. Paper presented at the proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining
10. Aiello LM, Barrat A, Schifanella R, Cattuto C, Markines B, Menczer F (2012) Friendship prediction and homophily in social media. ACM Trans Web 6(2):9
11. Almars A, Li X, Zhao X (2019) Modelling user attitudes using hierarchical sentiment-topic model. Data Knowl Eng 119:139–149
12. Almgren K, Lee J (2015) A hybrid framework to predict influential users on social networks. Paper presented at the digital information management (ICDIM), 2015 tenth international conference on
13. Alp ZZ, Öğüdücü ŞG (2018) Identifying topical influencers on twitter based on user behavior and network topology. Knowl-Based Syst 141:211–221
14. Aral S, Muchnik L, Sundararajan A (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. Proc Natl Acad Sci 106(51):21544–21549
15. Artzi Y, Pantel P, Gamon M (2012) Predicting responses to microblog posts. Paper presented at the proceedings of the 2012 conference of the north American chapter of the Association for Computational Linguistics: human language technologies
16. Backstrom L, Bakshy E, Kleinberg JM, Lento TM, Rosenn I (2011) Center of attention: how facebook users allocate attention across friends. ICWSM 11:23
17. Baingana B, Giannakis GB (2017) Tracking switched dynamic network topologies from information cascades. IEEE Trans Signal Process 65(4):985–997
18. Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: quantifying influence on twitter. Paper presented at the Proceedings of the fourth ACM international conference on Web search and data mining
19. Bakshy E, Rosenn I, Marlow C, Adamic L (2012) The role of social networks in information diffusion. Paper presented at the proceedings of the 21st international conference on world wide web
20. Bao Q, Cheung WK, Zhang Y, Liu J (2017) A component-based diffusion model with structural diversity for social networks. IEEE T Cybernetics 47(4):1078–1089
21. Barabasi A-L (2005) The origin of bursts and heavy tails in human dynamics. Nature 435(7039):207–211
22. Becker H, Naaman M, Gravano L (2010) Learning similarity metrics for event identification in social media. Paper presented at the Proceedings of the third ACM international conference on Web search and data mining
23. Bhattacharya P, Zafar MB, Ganguly N, Ghosh S, Gummadi KP (2014) Inferring user interests in the twitter social network. Paper presented at the proceedings of the 8th ACM conference on recommender systems

24. Bloch, F., Demange, G., & Kranton, R. (2018). Rumors and social networks. International Economic Review 59(2), 421-448.
25. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. (JSTAT) 2008(10):P10008
26. Bordianu GC (2012) Learning influence probabilities in social networks. McGill University
27. Borgatti SP (2005) Centrality and network flow. Soc Networks 27(1):55–71
28. Brereton P, Kitchenham BA, Budgen D, Turner M, Khalil M (2007) Lessons from applying the systematic literature review process within the software engineering domain. J Syst Softw 80(4):571–583
29. Budak C, Kannan A, Agrawal R, Pedersen J (2014). Inferring user interests from microblogs. Technical Report, MSR-TR-2014-68
30. Cano AE,Mazumdar S, Ciravegna F (2014) Social influence analysis in microblogging platforms–a topic-sensitive based approach. Semantic Web 5(5):357
31. Cao L (2010) In-depth behavior understanding and use: the behavior informatics approach. Inf Sci 180(17):3067–3085
32. Cha Y, Cho J (2012) Social-network analysis using topic models. Paper presented at the proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval.
33. Cha M, Haddadi H, Benevenuto F, Gummadi PK (2010) Measuring user influence in twitter: the million follower fallacy. ICWSM 10(10–17):30
34. Cha Y, Bi B, Hsieh C-C, Cho J (2013) Incorporating popularity in topic models for social network analysis. Paper presented at the proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval
35. Chen J, Nairn R, Nelson L, Bernstein M, Chi E (2010) Short and tweet: experiments on recommending content from information streams. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems
36. Colleoni E, Rozza A, Arvidsson A (2014) Echo chamber or public sphere? Predicting political orientation and measuring political homophily in twitter using big data. J Commun 64(2):317–332
37. Conover M, Ratkiewicz J, Francisco MR, Gonçalves B, Menczer F, Flammini A (2011) Political polarization on twitter. ICWSM 133:89–96
38. Crandall D, Cosley D, Huttenlocher D, Kleinberg J, Suri S (2008) Feedback effects between similarity and social influence in online communities. Paper presented at the proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining
39. Cui P, Wang F, Liu S, Ou M, Yang S, Sun L (2011) Who should share what?: item-level social influence prediction for users and posts ranking. Paper presented at the Proceedings of the 34th international ACM SIGIR conference on Research and development in Information retrieval
40. Daneshmand H, Gomez-Rodriguez M, Song L, Schoelkopf B (2014) Estimating diffusion network structures: recovery conditions, Sample Complexity & Soft-thresholding Algorithm. Paper presented at the ICML
41. Dang A, Moh'd A, Gruzd A, Milios E, Minghim R (2015) A visual framework for clustering memes in social media. Paper presented at the advances in social networks analysis and mining (ASONAM), 2015 IEEE/ACM international conference on
42. Davenport TH, Beck JC (2013) The attention economy: Understanding the new currency of business. Harvard Business Press
43. Davoudi A, Chatterjee M (2016) Prediction of information diffusion in social networks using dynamic carrying capacity. Paper presented at the big data (big data), 2016 IEEE international conference on.
44. De Choudhury M (2011) Tie formation on twitter: homophily and structure of egocentric networks. Paper presented at the privacy, security, risk and trust (PASSAT) and 2011 IEEE third International conference on social computing (SocialCom), 2011 IEEE third international conference on
45. Deutsch M, Gerard HB (1955) A study of normative and informational social influences upon individual judgment. J Abnorm Soc Psychol 51(3):629
46. Ding Y, Jiang J (2014) Extracting interest tags from twitter user biographies. Paper presented at the Asia Information Retrieval Symposium
47. Dougnon RY, Fournier-Viger P, Nkambou R (2015) Inferring user profiles in online social networks using a partial social graph. Paper presented at the Canadian Conference on Artificial Intelligence
48. Efstathiades H, Antoniades D, Pallis G, Dikaiakos MD, Szlávik Z, Sips R-J (2016) Online social network evolution: revisiting the Twitter graph. Paper presented at the big data (big data), 2016 IEEE international conference on
49. Ferrara E, JafariAsbagh M, Varol O, Qazvinian V, Menczer F, Flammini A (2013). Clustering memes in social media. Paper presented at the advances in social networks analysis and mining (ASONAM), 2013 IEEE/ACM international conference on

50. Fogués RL, Such JM, Espinosa A, Garcia-Fornes A (2014) BFF: a tool for eliciting tie strength and user communities in social networking services. Inf Syst Front 16(2):225–237
51. Gabrilovich E, Markovitch S (2007) Computing semantic relatedness using wikipedia-based explicit semantic analysis. Paper presented at the IJcAI
52. Gallos LK, Rybski D, Liljeros F, Havlin S, Makse HA (2012) How people interact in evolving online affiliation networks. Physical Review X 2(3):031014
53. Garousi V, Zhi J (2013) A survey of software testing practices in Canada. J Syst Softw 86(5):1354–1376
54. Ghosh R, Lerman K (2010) Predicting influential users in online social networks. arXiv preprint arXiv: 1005.4882
55. Ghosh R, Lerman K (2011) A framework for quantitative analysis of cascades on networks. Paper presented at the Proceedings of the fourth ACM international conference on Web search and data mining
56. Ghosh S, Sharma N, Benevenuto F, Ganguly N, Gummadi K (2012) Cognos: crowdsourcing search for topic experts in microblogs. Paper presented at the proceedings of the 35th international ACM SIGIR conference on research and development in information retrieval
57. Gilbert E, Karahalios K (2009) Predicting tie strength with social media. Paper presented at the Proceedings of the SIGCHI conference on human factors in computing systems
58. Golub B, Jackson MO (2012) How homophily affects the speed of learning and best-response dynamics. Q J Econ 127(3):1287–1338
59. Gomez Rodriguez M, Leskovec J, Schölkopf B (2013) Structure and dynamics of information pathways in online media. Paper presented at the Proceedings of the sixth ACM international conference on Web search and data mining
60. Gomez-Rodriguez M, Leskovec J, Krause A (2012) Inferring networks of diffusion and influence. ACM Trans Knowl Discov Data 5(4):21
61. Gonçalves B, Perra N, Vespignani A (2011) Modeling users' activity on twitter networks: validation of Dunbar's number. PLoS One 6(8):e22656
62. Goyal A, Bonchi F, Lakshmanan LV (2010) Learning influence probabilities in social networks. Paper presented at the Proceedings of the third ACM international conference on Web search and data mining
63. Granovetter MS (1973) The strength of weak ties. Am J Sociol 78(6):1360–1380
64. Granovetter M (1978) Threshold models of collective behavior. Am J Sociol 83(6):1420–1443
65. Greenberger M (1971) Computers, communications, and the public interest. Johns Hopkins University Press
66. Guille A, Hacid H, Favre C, Zighed DA (2013) Information diffusion in online social networks: a survey. ACM SIGMOD Rec 42(2):17–28
67. Halberstam Y, Knight B (2016) Homophily, group size, and the diffusion of political information in social networks: evidence from twitter. J Public Econ 143:73–88
68. Han J, Lee H (2016) Characterizing the interests of social media users: refinement of a topic model for incorporating heterogeneous media. Inf Sci 358:112–128
69. Hannon J, Bennett M, Smyth B (2010) Recommending twitter users to follow using content and collaborative filtering approaches. Paper presented at the Proceedings of the fourth ACM conference on Recommender systems
70. Hannon, J., McCarthy, K., O'Mahony, M. P., & Smyth, B. (2012). A multi-faceted user model for twitter. Paper presented at the International Conference on User Modeling, Adaptation, and Personalization
71. Harvey M, Crestani F, Carman MJ (2013) Building user profiles from topic models for personalised search. Paper presented at the proceedings of the 22nd ACM international conference on conference on information & knowledge management
72. He JS, Kaur H, Talluri M (2016) Positive opinion influential node set selection for social networks: considering both positive and negative relationships. In: Wireless communications, networking and applications. Springer, p 935-948
73. Holland PW, Laskey KB, Leinhardt S (1983) Stochastic blockmodels: first steps. Soc Networks 5(2):109–137
74. Hong L, Doumith AS, Davison BD (2013) Co-factorization machines: modeling user interests and predicting individual decisions in twitter. Paper presented at the Proceedings of the sixth ACM international conference on Web search and data mining
75. Hosseini-Pozveh M, Zamanifar K, Naghsh-Nilchi AR, Dolog P (2016) Maximizing the spread of positive influence in signed social networks. Intell Data Anal 20(1):199–218
76. Hosseini-Pozveh M, Zamanifar K, Naghsh-Nilchi AR (2019) Assessing information diffusion models for influence maximization in signed social networks. Expert Syst Appl 119:476–490
77. Hu Y, Chen M (2016) Information diffusion prediction in mobile social networks with hydrodynamic model. Paper presented at the communications (ICC), 2016 IEEE international conference on

78. Hu Z, Wang C, Yao J, Xing E, Yin H, Cui B (2013) Community specific temporal topic discovery from social media. arXiv preprint arXiv:1312.0860
79. Huang, A. H., Lehavy, R., Zang, A. Y., & Zheng, R. (2017). Analyst information discovery and interpretation roles: A topic modeling approach. Management Science, 64(6), 2833-2855.
80. Huang, X., Yang, Y., Hu, Y., Shen, F., & Shao, J. (2016). Dynamic user attribute discovery on social media. In Asia-Pacific Web Conference (pp. 256-267). Springer, Cham.. Paper presented at the Asia-Pacific Web Conference
81. Hung C-C, Huang Y-C, Hsu JY-j, Wu DK-C (2008) Tag-based user profiling for social media recommendation. Paper presented at the Workshop on Intelligent Techniques for Web Personalization & Recommender Systems at AAAI
82. Jackson MO, López-Pintado D (2013) Diffusion and contagion in networks with heterogeneous agents and homophily. Network Science 1(01):49–67
83. JafariAsbagh, M., Ferrara, E., Varol, O., Menczer, F., & Flammini, A. (2014). Clustering memes in social media streams. Soc Netw Anal Min 4(1): 237.
84. Jiang B, Sha Y (2015) Modeling temporal dynamics of user interests in online social networks. ProcediaComput Sci 51:503–512
85. Jiang C, Chen Y, Liu KR (2014) Evolutionary dynamics of information diffusion over social networks. IEEE Trans Signal Process 62(17):4573–4586
86. Jin D, Ma X, Zhang Y, Abbas H, Yu H (2018) Information diffusion model based on social big data. Mobile NETW APPL 23(4):717–722
87. Jones JJ, Settle JE, Bond RM, Fariss CJ, Marlow C, Fowler JH (2013) Inferring tie strength from online directed behavior. PLoS One 8(1):e52168
88. Kang J, Choi H, Lee H (2019) Deep recurrent convolutional networks for inferring user interests from social media. J Intell Inf Syst 52(1):191–209
89. Kapanipathi P, Jain P, Venkataramani C, Sheth A (2014) User interests identification on twitter using a hierarchical knowledge base. Paper presented at the European Semantic Web Conference
90. Kaplan AM, Haenlein M (2010) Users of the world, unite! The challenges and opportunities of social media. Bus Horiz 53(1):59–68
91. Karrer B, Newman ME (2011) Competing epidemics on complex networks. Phys Rev E 84(3):036106
92. Karsai M, Kivelä M, Pan RK, Kaski K, Kertész J, Barabási A-L, Saramäki J (2011) Small but slow world: how network topology and burstiness slow down spreading. Phys Rev E 83(2):025102
93. Kelman, H. C. (2017). Further thoughts on the processes of compliance, identification, and internalization *Social power and political influence* (pp. 125-171): Routledge.
94. Kempe D, Kleinberg J, Tardos É (2003) Maximizing the spread of influence through a social network. Paper presented at the Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining
95. Khan HU, Daud A, Ishfaq U, Amjad T, Aljohani N, Abbasi RA, Alowibdi JS (2017) Modelling to identify influential bloggers in the blogosphere: a survey. Comput Hum Behav 68:64–82
96. Khemmarat S, Saha S, Song HH, Baldi M, Gao L (2014) On understanding user interests through heterogeneous data sources. Paper presented at the International Conference on Passive and Active Network Measurement
97. Kim H-N, Ha I, Lee K-S, Jo G-S, El-Saddik A (2011) Collaborative user modeling for enhanced content filtering in recommender systems. Decis Support Syst 51(4):772–781
98. Kossinets G, Watts DJ (2009) Origins of homophily in an evolving social network 1. Am J Sociol 115(2): 405–450
99. Kumar R, Novak J, Tomkins A (2010) Structure and evolution of online social networks. In: Link mining: models, algorithms, and applications. Springer, p 337-357
100. Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In Proceedings of the 19th international conference on World wide web (pp. 591-600). AcM.
101. Lahuerta-Otero E, Cordero-Gutiérrez R (2016) Looking for the perfect tweet. The use of data mining techniques to find influencers on twitter. *Computers in Human behavior* 64:575–583
102. Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, … Gutmann M (2009) Life in the network: the coming age of computational social science. *Science (New York, NY) 323*(5915):721
103. Lehmann J, Gonçalves B, Ramasco JJ, Cattuto C (2012) Dynamical classes of collective attention in twitter. Paper presented at the proceedings of the 21st international conference on world wide web
104. Lerman K, Ghosh R, Surachawala T (2012) Social contagion: an empirical study of information spread on Digg and twitter follower graphs. arXiv preprint arXiv:1202.3162
105. Li J, Peng W, Li T, Sun T, Li Q, Xu J (2014) Social network user influence sense-making and dynamics prediction. Expert Syst Appl 41(11):5115–5124

106. Lim KH, Datta A (2013) Interest classification of Twitter users using Wikipedia. Paper presented at the proceedings of the 9th international symposium on open collaboration
107. Liu C, Zhang Z-K (2014) Information spreading on dynamic social networks. Commun Nonlinear Sci Numer Simul 19(4):896–904
108. Liu L, Tang J, Han J, Yang S (2012) Learning influence from heterogeneous social networks. Data Min Knowl Disc 25(3):511–544
109. Liu L, Qu B, Chen B, Hanjalic A, Wang H (2018) Modelling of information diffusion on social networks with applications to WeChat. Physica A 496:318–329
110. Lobel I, Sadler E (2015) Information diffusion in networks through social learning. Theor Econ 10(3):807–851
111. Locher DA (2002) Collective behavior. Prentice Hall, Upper Saddle River
112. Lomi, A., Robins, G., & Tranmer, M. (2016). Introduction to multilevel social networks. *Social networks*(44), 266-268.
113. Lu C, Lam W, Zhang Y (2012) Twitter user modeling and tweets recommendation based on wikipedia concept graph. Paper presented at the Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence
114. Ma Y, Zeng Y, Ren X, Zhong N (2011) User interests modeling based on multi-source personal information fusion and semantic reasoning. Paper presented at the International Conference on Active Media Technology
115. Mahmud J (2014) Why do you write this? Prediction of Influencers from Word Use. Paper presented at the ICWSM
116. Mashayekhi Y, Meybodi MR, Rezvanian A (2018) Weighted estimation of information diffusion probabilities for independent cascade model. Paper presented at the 2018 4th international conference on web research (ICWR)
117. McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: Homophily in social networks. Annu Rev Sociol 27(1):415–444
118. Mehrotra R, Sanner S, Buntine W, Xie L (2013) Improving lda topic models for microblogs via tweet pooling and automatic labeling. Paper presented at the proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval
119. Meng L, Hulovatyy Y, Striegel A, Milenković T (2016) On the interplay between individuals' evolving interaction patterns and traits in dynamic multiplex social networks. IEEE Trans Netw Sci Eng 3(1):32–43
120. Mezghani M, Zayani CA, Amous I, Péninou A, Sedes F (2014) Dynamic enrichment of social users' interests. Paper presented at the research challenges in information science (RCIS), 2014 IEEE eighth international conference on
121. Michelle GG, Kumaran P, Chitrakala S (2016) Topic sensitive information diffusion modelling in online social networks. Paper presented at the advances in electrical, electronics, information, communication and bio-informatics (AEEICB), 2016 2nd international conference on
122. Michelson M, Macskassy SA (2010) Discovering users' topics of interest on twitter: a first look. Paper presented at the Proceedings of the fourth workshop on Analytics for noisy unstructured text data
123. Milenković T, Memišević V, Bonato A, Pržulj N (2011) Dominating biological networks. PLoS One 6(8): e23016
124. Miritello G, Moro E, Lara R (2011) Dynamical strength of social ties in information spreading. Phys Rev E 83(4):045102
125. Mislove A, Viswanath B, Gummadi KP, Druschel P (2010) You are who you know: inferring user profiles in online social networks. Paper presented at the Proceedings of the third ACM international conference on Web search and data mining
126. MUSCILLO, A. (2014). Discrete Models of Information Diffusion in Networks.
127. Myers S, Leskovec J (2010) On the convexity of latent social network inference. Paper presented at the Advances in Neural Information Processing Systems
128. Narducci F, Musto C, Semeraro G, Lops P, de Gemmis M (2013) Exploiting big data for enhanced representations in content-based recommender systems. Paper presented at the International Conference on Electronic Commerce and Web Technologies
129. Newman ME (2003) The structure and function of complex networks. SIAM Rev 45(2):167–256
130. Orlandi F, Breslin J, Passant A (2012) Aggregated, interoperable and multi-domain user profiles for the social web. Paper presented at the proceedings of the 8th international conference on semantic systems
131. Oselio B, Kulesza A, Hero AO (2014) Multi-layer graph analysis for dynamic social networks. IEEE J Sel Topics Signal Process 8(4):514–523
132. Ottoni R, Las Casas DB, Pesce JP, Meira Jr W, Wilson C, Mislove A, Almeida VA (2014) Of pins and tweets: investigating how users behave across image-and text-based social networks. Paper presented at the ICWSM

133. Pal A, Herdagdelen A, Chatterji S, Taank S, Chakrabarti D (2016) Discovery of topical authorities in instagram. Paper presented at the proceedings of the 25th international conference on world wide web
134. Pennacchiotti M, Gurumurthy S (2011) Investigating topic models for social media user recommendation. Paper presented at the proceedings of the 20th international conference companion on world wide web
135. Pereira FS, Gama J, de Amo S, Oliveira GM (2018) On analyzing user preference dynamics with temporal social networks. Mach Learn 107(11):1745–1773
136. Piao G (2016) Towards comprehensive user modeling on the social web for personalized link recommendations. Paper presented at the proceedings of the 2016 conference on user modeling adaptation and personalization
137. Piao G, Breslin JG (2016a) Analyzing aggregated semantics-enabled user modeling on Google+ and Twitter for personalized link recommendations. Paper presented at the proceedings of the 2016 conference on user modeling adaptation and personalization
138. Piao G, Breslin JG (2016b) Exploring dynamics and semantics of user interests for user modeling on Twitter for link recommendations. Paper presented at the proceedings of the 12th international conference on semantic systems
139. Piao G, Breslin JG (2016c). User modeling on Twitter with WordNet Synsets and DBpedia concepts for personalized recommendations. Paper presented at the proceedings of the 25th ACM international on conference on information and knowledge management
140. Pochampally R, Varma V (2011) User context as a source of topic retrieval in Twitter. Paper presented at the workshop on enriching information retrieval (with ACM SIGIR)
141. Ramage D, Hall D, Nallapati R, Manning CD (2009) Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora. Paper presented at the proceedings of the 2009 conference on empirical methods in natural language processing: volume 1-volume 1
142. Ramanathan K, Kapoor K (2009) Creating user profiles using wikipedia. Paper presented at the International Conference on Conceptual Modeling
143. Rashotte L (2007) Social influence. The Blackwell Encyclopedia of Social Psychology 9:562–563
144. Riquelme F, González-Cantergiani P (2016) Measuring user influence on twitter: a survey. Inf Process Manag 52(5):949–975
145. Rodriguez MG, Schölkopf B (2012) Submodular inference of diffusion networks from multiple trees. arXiv preprint arXiv:1205.1671
146. Rodriguez MG, Balduzzi D, Schölkopf B (2011). Uncovering the temporal dynamics of diffusion networks. arXiv preprint arXiv:1105.0697
147. Rogers EM, Cartano DG (1962) Methods of measuring opinion leadership. Public Opin Q:435–441
148. Romero DM, Galuba W, Asur S, Huberman BA (2011a) Influence and passivity in social media. Paper presented at the Joint European Conference on Machine Learning and Knowledge Discovery in Databases
149. Romero DM, Tan C, Ugander J (2011b) On the interplay between social and topical structure. arXiv preprint arXiv:1112.1115
150. Sadikov E, Medina M, Leskovec J, Garcia-Molina H (2011) Correcting for missing data in information cascades. Paper presented at the Proceedings of the fourth ACM international conference on Web search and data mining
151. Saito K, Nakano R, Kimura M (2008). Prediction of information diffusion probabilities for independent cascade model. Paper presented at the Knowledge-based intelligent information and engineering systems
152. Saito K, Kimura M, Ohara K, Motoda H (2010) Generative models of information diffusion with asynchronous Timedelay. Paper presented at the ACML
153. Saito K, Kimura M, Ohara K, Motoda H (2012a) Efficient discovery of influential nodes for SIS models in social networks. Knowl Inf Syst 30(3):613–635
154. Saito K, Kimura M, Ohara K, Motoda H (2012b) Learning asynchronous-time information diffusion models and its application to behavioral data analysis over social networks. arXiv preprint arXiv:1204.4528
155. Schifanella R, Barrat A, Cattuto C, Markines B, Menczer F (2010) Folks in folksonomies: social link prediction from shared metadata. Paper presented at the Proceedings of the third ACM international conference on Web search and data mining
156. Servia-Rodríguez S, Díaz-Redondo R, Fernández-Vilas A, Pazos-Arias JJ (2013) Mining facebook activity to discover social ties: towards a social-sensitive ecosystem. Closer 367:71–85
157. Shah B, Verma AP, Tiwari S (2018) User interest modeling from social media network graph. Enriched with semantic web. Paper presented at the Proceedings of International Conference on Computational Intelligence and Data Engineering
158. Sharma NK, Ghosh S, Benevenuto F, Ganguly N, Gummadi K (2012) Inferring who-is-who in the twitter social network. ACM Sigcomm Comp Com 42(4):533–538

159. Sheikhahmadi A, Nematbakhsh MA (2017) Identification of multi-spreader users in social networks for viral marketing. J Inf Sci 43(3):412–423
160. Shu X, Qi G-J, Tang J, Wang J (2015) Weakly-shared deep transfer networks for heterogeneous-domain knowledge propagation. Paper presented at the Proceedings of the 23rd ACM international conference on Multimedia
161. Slaughter AJ, Koehly LM (2016) Multilevel models for social networks: hierarchical Bayesian approaches to exponential random graph modeling. Soc Networks 44:334–345
162. Sneppen K, Trusina A, Jensen MH, Bornholdt S (2010) A minimal model for multiple epidemics and immunity spreading. PLoS One 5(10):e13326
163. Snijders TA (2016) The multiple flavours of multilevel issues for networks Multilevel network analysis for the social sciences. Springer, p 15-46
164. Sohrabi MK, Akbari S (2016) A comprehensive study on the effects of using data mining techniques to predict tie strength. Comput Hum Behav 60:534–541
165. Steeg GV, Ghosh R, Lerman K (2011) What stops social epidemics? arXiv preprint arXiv:1102.1985
166. Sun Q, Wang N, Zhou Y, Luo Z (2016) Identification of influential online social network users based on multi-features. Int J Pattern Recognit Artif Intell 30(06):1659015
167. Szabo G, Huberman BA (2010) Predicting the popularity of online content. Commun ACM 53(8):80–88
168. Tabasso N (2015) Diffusion of multiple information: on information resilience and the power of segregation
169. Tan C, Tang J, Sun J, Lin Q, Wang F (2010) Social action tracking via noise tolerant time-varying factor graphs. Paper presented at the proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining.
170. Tang J, Sun J, Wang C, Yang Z (2009) Social influence analysis in large-scale networks. Paper presented at the proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining
171. Tang S, Blenn N, Doerr C, Van Mieghem P (2011) Digging in the digg social news website. IEEE Trans Multimedia 13(5):1163–1175
172. Tang J, Shu X, Qi G-J, Li Z, Wang M, Yan S, Jain R (2017) Tri-clustered tensor completion for social-aware image tag refinement. IEEE Trans Pattern Anal Mach Intell 39(8):1662–1674
173. Tang J, Shu X, Li Z, Jiang Y-G, Tian Q (2019) Social anchor-unit graph regularized tensor completion for large-scale image retagging. IEEE Trans Pattern Anal Mach Intell
174. Tao K, Abel F, Gao Q, Houben G-J (2011) Tums: twitter-based user modeling service. Paper presented at the Extended Semantic Web Conference
175. Tong C, He W, Niu J, Xie Z (2016) A novel information cascade model in online social networks. Physica A 444:297–310
176. Trikha AK, Zarrinkalam F, Bagheri E (2018) Topic-association mining for user interest detection. Paper presented at the European Conference on Information Retrieval
177. Uysal I, Croft WB (2011) User oriented tweet ranking: a filtering approach to microblogs. Paper presented at the proceedings of the 20th ACM international conference on information and knowledge management
178. Vespignani A (2009) Predicting the behavior of techno-social systems. Science 325(5939):425–428
179. Viswanath B, Mislove A, Cha M, Gummadi KP (2009) On the evolution of user interaction in Facebook. Paper presented at the proceedings of the 2nd ACM workshop on online social networks
180. Wan J, Chen X, Du Y, Jia M (2019) Information propagation model based on hybrid social factors of opportunity, trust and motivation. Neurocomputing 333:169–184
181. Wang F, Wang H, Xu K (2012) Diffusive logistic model towards predicting information diffusion in online social networks. Paper presented at the distributed computing systems workshops (ICDCSW), 2012 32nd international conference on.
182. Wang F, Wang H, Xu K, Wu J, Jia X (2013a) Characterizing information diffusion in online social networks with linear diffusive model. Paper presented at the distributed computing systems (ICDCS), 2013 IEEE 33rd international conference on
183. Wang T, Liu H, He J, Du X (2013b) Mining user interests from information sharing behaviors in social media. Paper presented at the Pacific-Asia Conference on Knowledge Discovery and Data Mining
184. Wang Y, Shen H-W, Liu S, Cheng X-Q (2013c) Learning user-specific latent influence and susceptibility from information cascades. arXiv preprint arXiv:1310.3911
185. Wang F, Wang G, Xie D (2016a) Maximizing the spread of positive influence under LT-MLA model. Paper presented at the Asia-Pacific Services Computing Conference
186. Wang P, Robins G, Pattison P, Lazega E (2016b) Social selection models for multilevel networks. Soc Networks 44:346–362
187. Wang H, Huang X, Li L (2018) Microblog oriented interest extraction with both content and network structure. Intell Data Anal 22(3):515–532

188. Watts DJ (2007) A twenty-first century science. Nature 445(7127):489–489
189. Weber I, Garimella VRK, Batayneh A (2013) Secular vs. Islamist polarization in Egypt on twitter. Paper presented at the proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining
190. Weng L (2014) Information diffusion on online social networks. Citeseer
191. Weng L, Lento TM (2014) Topic-based clusters in egocentric networks on Facebook. Paper presented at the ICWSM
192. Weng L, Menczer F (2015) Topicality and impact in social media: diverse messages, focused messengers. PLoS One 10(2):e0118410
193. Weng J, Lim E-P, Jiang J, He Q (2010) Twitterrank: finding topic-sensitive influential twitterers. Paper presented at the Proceedings of the third ACM international conference on Web search and data mining
194. Weng L, Flammini A, Vespignani A, Menczer F (2012) Competition among memes in a world with limited attention. Sci Rep 2
195. Wu F, Huberman BA (2007) Novelty and collective attention. Proc Natl Acad Sci 104(45):17599–17601
196. Wuchty S, Stadler PF (2003) Centers of complex networks. J Theor Biol 223(1):45–53
197. Xiang R, Neville J, Rogati M (2010) Modeling relationship strength in online social networks. Paper presented at the proceedings of the 19th international conference on world wide web
198. Xiao Y, Wang Z, Li Q, Li T (2019) Dynamic model of information diffusion based on multidimensional complex network space and social game. Physica A 521:578–590
199. Xu KS, Hero AO (2014) Dynamic stochastic blockmodels for time-evolving social networks. J Sel Topics Signal Process 8(4):552–562
200. Xu T, Oard DW (2011) Wikipedia-based topic clustering for microblogs. Proc Am Soc Inf Sci Technol 48(1):1–10
201. Yang J, Leskovec J (2010) Modeling information diffusion in implicit networks. Paper presented at the data mining (ICDM), 2010 IEEE 10th international conference on
202. Yang L, Sun T, Zhang M, Mei Q (2012) We know what@ you# tag: does the dual role affect hashtag adoption? Paper presented at the proceedings of the 21st international conference on world wide web
203. Ye S, Wu SF (2010) Measuring message propagation and social influence on Twitter. com. Paper presented at the international conference on social informatics
204. Yi Y, Zhang Z, Gan C (2018) The effect of social tie on information diffusion in complex networks. Physica A 509:783–794
205. Yin H, Cui B, Chen L, Hu Z, Zhou X (2015) Dynamic user modeling in social media systems. ACM Trans Inf Syst 33(3):10
206. Zarrinkalam F, Kahani M, Bagheri E (2018) Mining user interests over active topics on social networks. Inf Process Manag 54(2):339–357
207. Zarrinkalam F, Kahani M, Bagheri E (2019) User interest prediction over future unobserved topics on social networks. Inform Retrieval J 22(1–2):93–128
208. Zhang X, Su Y, Qu S, Xie S, Fang B, Yu P (2018) IAD: interaction-aware diffusion framework in social networks. IEEE Trans Knowl Data Eng
209. Zhao WX, Jiang J, Weng J, He J, Lim E-P, Yan H, Li X (2011) Comparing twitter and traditional media using topic models. Paper presented at the European Conference on Information Retrieval
210. Zhao Z, Cheng Z, Hong L, Chi EH (2015) Improving user topic interest profiles by behavior factorization. Paper presented at the proceedings of the 24th international conference on world wide web
211. Zhong E, Fan W, Wang J, Xiao L, Li Y (2012) Comsoc: adaptive transfer of user behaviors over composite social network. Paper presented at the proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining
212. Zhou E, Li D, Madden A, Chen Y, Ding Y, Kang Q, Su H (2019) Modeling adoption behavior for innovation diffusion. Paper presented at the International Conference on Information
213. Zhu Z (2013) Discovering the influential users oriented to viral marketing based on online social networks. Physica A 392(16):3459–3469
214. Zhu Z, Su J, Kong L (2015) Measuring influence in online social network based on the user-content bipartite graph. Comput Hum Behav 52:184–189
215. Zhuang K, Shen H, Zhang H (2017) User spread influence measurement in microblog. Multimed Tools Appl 76(3):3169–3185

Amir Masoud Rahmani received his BS in Computer Engineering from Amir Kabir University, Tehran, in 1996, the MS in Computer Engineering from Sharif University of Technology, Tehran, in 1998 and the PhD degree in Computer Engineering from IAU University, Tehran, in 2005. Currently, he is a Professor in the Department of Computer Engineering at the IAU University. He is the author/co-author of more than 150 publications in technical journals and conferences. His research interests are in the areas of distributed systems, ad hoc and wireless sensor networks and evolutionary computing.

Affiliations

Rahebeh Mojtahedi Safari[1] • **Amir Masoud Rahmani**[2] • **Sasan H. Alizadeh**[3]

[1] Faculty of computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran

[2] Computer Engineering Department, Science and Research Branch, Islamic Azad University, Tehran, Iran

[3] Department of Information Technology, ICT Research Institute (Iran Telecommunication Research Center), Tehran, Iran