Interest-aware Information Diffusion in Evolving Social Networks

Jiaqi Liu, Luoyi Fu, Zhe Liu, Xiao-Yang Liu and Xinbing Wang

***Abstract*—Many realistic wireless social networks are manifested to be evolving over time. While network evolution has its important influence on network performances, it is nevertheless overlooked in most existing studies on information diffusion. Motivated by this, in this paper we investigate the delivery accuracy of interest-aware information diffusion in evolving social networks. In doing so, we adopt a model, named affiliation networks, to characterize network evolution from three aspects, i.e., the arrival of new users, the generation of new interests and the creation of new links between them. Based on that, we consider a publishing based information diffusion mechanism that widely exists in wireless networking services such as Facebook, Twitter and Sina Weibo, where a user receives data items from his friends and then republishes the ones he is interested in to all his friends. Under the above network model, we study how the performance metric such as delivery accuracy is affected by the network evolution. The publishing based information diffusion mechanism is a blind targeting one that may suffer a low delivery accuracy. However, our analytical results demonstrate a contrary finding that the delivery accuracy is improved over time, and even more surprisingly, we disclose that with a sufficiently long evolving time, the delivery accuracy can achieve a perfect state where those who receive the data are exactly the ones that are interested in it. In addition, our theoretical findings are verified by experimental measurements through a social network dataset from Facebook.**

I. Introduction

EVOLUTION is manifested to be a common property of many realistic wireless social networks [1] [2], which are observed to exhibit both the arrival of new users and the generation of new interests. For example, Twitter was reported to have 300 billion monthly active users in the first quarter of 2016, with a 50% growth compared with that of 2012 [3]. In addition to the growth of users, there also emerge a wide span of new interests in people's communication, with a range from social events to personal preferences. As a result, social network users share an unprecedented amount of daily information, which imposes an urgent demand for effective information diffusion. In many wireless networking services, e.g., Facebook, Twitter and Sina Weibo, information diffusion happens in a spontaneous way. For example, in Twitter, a user receives a data from his friends' publications. Then, if he feels interested in the data, he will republish it and therefore all his friends can receive the data; otherwise, he will ignore it. Such a process is named as publishing based information diffusion mechanism in this work. Essentially, this is a blind targeting mechanism that may suffer a low delivery accuracy. Namely, data cannot flow to target users and meanwhile, users hardly access their interested data whereas being flooded by redundant data. Therefore, the following questions arise naturally: *In evolving social networks, is the publishing based information diffusion mechanism an effective one for achieving satisfactory performance metrics such as delivery accuracy? If so, what network properties are required to achieve good delivery accuracy? And how it evolves over time?*

We give positive answers to the questions above. Actually, network topology and information diffusion are co-evolving in realistic networks, resulting in a virtuous circle:

- The network construction is interest-guided, as the new users preferentially link to those who sharing interests.
- The links are therefore interest-based, strengthening the user relationships.
- Propagating through interest-based links, the blind targeting mechanism turns out to be interest-aware.

The main purpose of this work is to theoretically model and analyze the problem illustrated previously – delivery accuracy of the publishing based information diffusion mechanism in evolving social networks. By doing so, we can obtain a deeper understanding of how a data packet spreads among users in a spontaneous way, which could further provide guidance to the design of related information diffusion algorithms. The main challenge to analyze the delivery accuracy lies in two aspects: evolution of network structure and blind targeting property of the information diffusion mechanism. With network evolution, the arrival of new users expands the network size and the generation of new interests changes the relationships among users. Both factors result in the temporary dynamics of network structure and lead to an uncertain information flow. In addition, in the information diffusion process, user behavior follows a spontaneous manner, which is a blind targeting one that may appear to be inefficient.

To approach the problem, we start by capturing the evolution of social networks through adopting a model named affiliation networks [4]. The evolving process in this model incorporates two aspects, i.e., users and interests. At each time slot, a new user arrives with a certain probability, selects an existing user (someone who shares common interests with it) as the prototype and then establishes connections with him as well as some of his friends. A symmetrical process also occurs to the interests. Note that this kind of evolving process follows a natural and reasonable way and can well explain many network properties, such as power-law degree distribution, densification [4] and shrinking diameter [33]. The model has been verified to be a good capture of evolving social networks and consequently we adopt it as our theoretical model.

Based on the evolution model, we proceed to consider the publishing based information diffusion mechanism described as follows. We assume that the content of a data item includes

some sub-fields, according to which each user can be classified into either of the two types: a *viewer* that is interested in at least one sub-field of the data and a *redundant user* otherwise. Among all viewers, only some of them, referred as *publishers*, will republish the data since they exhibit stronger interest toward it. To evaluate the delivery accuracy of this information diffusion mechanism, we further adopt two previously proposed metrics, i.e., *recall* and *precision* [6] [7], with the former one measuring the fraction of receivers in all viewers, and the latter one characterizing the fraction of interested users in all receivers. At first glance, such a mechanism will still bring about some poor performances on recall and precision due to the concern that data will be inevitably delivered to some redundant users while some of the viewers will fail to receive it. However, as we will disclose subsequently in our results, such deficiency can potentially be eliminated with the network evolution, where the delivery accuracy may counter-intuitively reach a satisfactory state.

In this paper, we theoretically characterize the delivery accuracy of the interest-aware information diffusion in evolving social networks. We first present a special case to provide an intuitive view of this problem. The delivery accuracy in this case can achieve a perfect state, based on which we disclose necessary conditions to guarantee such a good performance. Then, we come to the general case where the evolving social networks are modeled by the affiliation networks, based on which we theoretically discuss how the information diffusion process behaves with the evolution of network and calculate the delivery accuracy in limit case where the evolving time approaches to infinity. We also conduct experimental measurements based on a dataset from Facebook [8]. Results show that recall and precision both achieve a better state with the evolution of social networks.

Our main contributions are summarized as follows:

- We disclose the necessary conditions to achieve perfect delivery accuracy in evolving social networks and demonstrate that they are jointly determined by network connectivity, user behavior and evolution of social networks.
- We theoretically analyze the publishing based information diffusion mechanism in evolving social networks and figure out that, the number of viewers increases and the number of publishers increases or remains the same over time.
- We prove that with a sufficiently long evolving time, the network structure evolves to satisfy all the necessary conditions and consequently the metrics recall and precision can achieve one simultaneously.

The rest of this paper is organized as follows. We give literature review in Section II. The system model and definitions are given in Section III and our main results are briefly introduced in Section IV. We provide theoretical analysis on delivery accuracy in evolving social networks in Section V. Our theoretical findings are verified through experimental results in Section VI. We conclude in Section VII.

II. Related Work

A flurry of previous studies [1] [2] [33] [9] have clarified that the structure of social networks evolves over time, including the arrival and departure of users [10] [11] and temporary dynamics of interest [12]. Regarding this, some models have been proposed to characterize the evolving process. Chung *et al.* [13] assume that either a vertex-arrival event or an edge-arrival event occurs at each time slot. Ghoshal *et al.* [14] establish a model that elucidates the role of the individual elementary mechanisms. Besides the aforementioned models, Lattanzi *et al.* [4] propose the affiliation networks, where two basic entries are related by affiliation of the former in the latter. The underlying evolving process follows a manner of preferential attachment and edge copying in a natural way.

As a hot topic in social networking, information diffusion is also under intensive study [15] - [22]. It is pointed out by some prior literature that the performance of information diffusion can be greatly improved by taking advantage of data content and user interests [23] [24] [25]. Matsubara *et al.* [26] propose a model for the rise and fall patterns of information diffusion, where the interest of event is taken into consideration. Wang *et al.* [27] describe the information propagation as a discrete Galton-Watson with Killing process, providing an explanation on the intrinsic interest of the message. Gao *et al.* [28] propose a probabilistic model of user interest and based on it develop a user-centric information diffusion approach. All the above works conduct the interest-aware information diffusion mechanism aiming to achieve a better performance. However, they do not take into accounts the feature of network evolution. As an exception, Fu *et. al* [29] investigate capacity scaling in an evolving network in terms of both geographic node distribution and traffic patterns. Though capacity is demonstrated to be significantly impacted by both social relations and network evolution, it remains unexplored in their work how the evolution of user interests may potentially affect some key performance metrics.

In order to better capture the delivery accuracy of information diffusion in realistic social networks, we jointly consider both the evolution of users and interests, and provide analysis of the effect they have on the delivery accuracy. To our best knowledge, this is the first attempt of performance derivation in the context of interest-aware evolving networks.

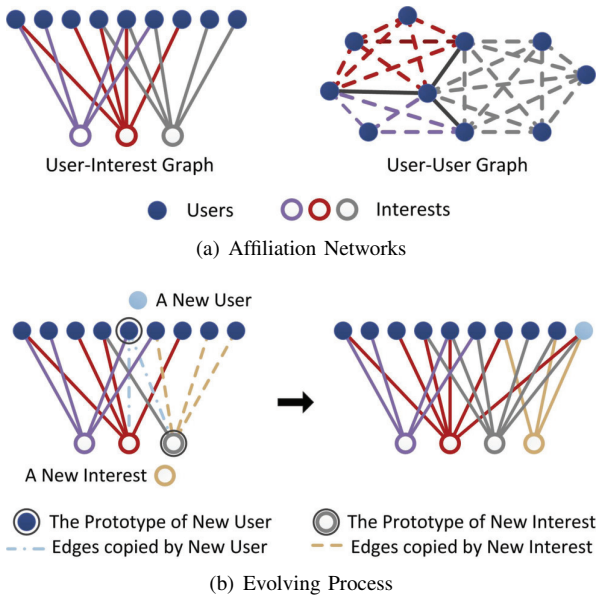
III. System Model

In this section, we first introduce the model of evolving social networks. Then, based on it we describe the publishing based information diffusion mechanism. Finally, we present definitions of two metrics, i.e., recall and precision.

A. Evolving Social Network Structure

We apply the affiliation networks model proposed by S. Lattanzi [4] in our model to characterize the evolving process of social networks. The affiliation in this model is employed to describe the relationships between users and interests, as in user-interest graph $B(U, I)$. Based on it the user-user graph $G(U, E)$ is generated, characterizing the relationships among users. An illustration of these two graphs is shown in Fig. 1(a).

1) User-Interest Graph: Let U denote the set of users and I denote the set of interests in the network. $B(U, I)$ is a simple bipartite graph composed of these two disjoint parts of nodes and a set of edges describing their relationships. In this graph,



(a) Affiliation Networks

(b) Evolving Process

Fig. 1. An illustration of the evolving social network structure. The color of each edge indicates the common interest shared by users. The edge between two users who share more than one common interests is colored by black.

an edge exists between user u and interest i if u is interested in i. For each user $u \in U$, it has a user profile recording all its interests, which is denoted by $I_u \subseteq I$.

2) User-User Graph: The relationships among users are characterized in the user-user graph $G(U, E)$. In this graph, U is the set of users, the same with that in $B(U, I)$, and E is the set of edges characterizing the relationships among users. $G(U, E)$ is generated from the user-interest graph $B(U, I)$ and thus builds the interest-based relationships among users. Let $d_i(t)$ denote the degree of interest i at time t. At each time slot t, if the common interest set of users u and v, i.e., $I_u \cap I_v$, is updated with a new interest i, an edge is added between nodes u and v with the probability $\frac{1}{c_0}\frac{1}{d_i(t)^\alpha}$, where c_0 is a sufficiently large constant to ensure that the probability is no greater than 1, and $\alpha \in [0, 1]$ is a constant. $G(U, E)$ is a multigraph.

3) Evolution of $B(U, I)$ and $G(U, E)$: In general, network structure evolves over time. For the convenience of modeling and analysis, we assume that the evolving time is slotted into time slots. The evolving process of $B(U, I)$ is proposed in [4] and for the convenience we reproduce it in Algorithm 1. Initial time of network evolution is set as $t = 0$. Line 1 to line 3 show the evolution of set U and line 4 to line 5 describe that of set I. An illustration of this evolving process is shown in Fig. 1(b). In addition, the evolving process of $G(U, E)$ is given in Algorithm 2. The intuition of this evolving process can be explained as follows. When a new user joins the network, he probably has some other users in mind, for example, friends of him who have joined the network, that may be the prototype, and he is likely to share common interests as this prototype. A symmetrical process happens to interest. Some more details on Affiliation Network Model can be found in [4].

B. Information Diffusion Process

Based on the evolving social network structure, we consider the publishing based information diffusion process.

Algorithm 1 Evolution of $B(U, I)$

Input: Parameters $c_u, c_i > 0$, and $\beta \in (0, 1)$; A bipartite graph $B_0(U, I)$ at time 0 with finite number of nodes and edges, where each node in U has at least c_u edges and each node in I has at least c_i edges.

Output: A temporary graph $B_t(U, I)$.

1: At time slot $t > 0$, a user u arrives with probability β and is added to the user set U.
2: *(Preferentially Chosen Prototype)* A node $v \in U$ is chosen as the prototype for the new node, with a probability proportional to its degree.
3: *(Edge Copying)* c_u edges are "copied" from v, that is, c_u neighbors of v, denoted by $i_1, ..., i_{c_u}$, are chosen uniformly and randomly (without replacement), and the edges $(u, i_1), ..., (u, i_{c_i})$ are added to the graph.
4: At time slot $t > 0$, an interest i arrives with probability $1 - \beta$ and is added to the interest set I.
5: Following a symmetrical process, add c_i edges to i.

Algorithm 2 Evolution of $G(U, E)$

Input: Parameters $c_u, c_i > 0$, $\alpha \in [0, 1]$ and $\beta \in (0, 1)$; A graph $G_0(U, E)$ at time 0 consisting of the node set U in $B_0(U, E)$, where each interest $i \in I_v \cap I_u$ generates an edge between u and v with the probability $\frac{1}{c_0}\frac{1}{d_i(0)^\alpha}$.

Output: A temporary graph $G_t(U, E)$.

1: At time slot $t > 0$, if a new user u arrives in $B_t(U, I)$, add u to the user set U.
2: *(Edges via Prototype)* An edge between u and the user v who has a common interest $i \in I_u \cap I_v$ in $B_t(U, I)$ is added in the graph with the probability $\frac{1}{c_0}\frac{1}{d_i(t)^\alpha}$. (Note that this is done after the edges of u are determined in $B_t(U, I)$.)
3: At time slot $t > 0$, if a new interest i arrives in $B_t(U, I)$, an edge is added between users u and v, who are neighbors of i in $B_t(U, I)$, with the probability $\frac{1}{c_0}\frac{1}{d_i(t)^\alpha}$. (Similarly, this is done after the edges of i are determined in $B_t(U, I)$.)

1) Generation Time of Data Item: Generation time of data item d, i.e., t_d, is set as the one when the first interest of the data d joins the network. We note that generation time of data item d, i.e., t_d, is different from the initial time of the network evolution, i.e., $t = 0$, where the former one is the time when the first interest of d joins the network while the latter one is the time before the first node $u \notin B_0(U, I)$ joins the network.

2) Structure of Data Item: In our model, we assume that a data item is related to M different interests, where M is a constant. This assumption is reasonable since the content of a data item may involve several different fields.

3) User Behavior: Let I_d denote the interests set of data item d. Whether a user u is a viewer of data item d or even a publisher depends on the number of common interests between it and the data, i.e., $|I_d \cap I_u|$. The exact definitions are given below. According to the definitions, all viewers are interested in the data, and the set of them is denoted by V_d. Among them only publishers are sufficiently interested to republish the data and we denote them by P_d. Obviously, $P_d \subseteq V_d$.

Definition 1 (Viewer)**.** *For a data item d with interests set I_d,*

a user u with user interests set I_u is a viewer if $I_d \cap I_u \neq \emptyset$.

Definition 2 (Publisher)**.** *For a data item d with interests set I_d, a user u with user interests set I_u is a publisher if u is a viewer and $|I_d \cap I_u| \geq m$, where $1 \leq m \leq M$ is an integer.*

4) Information Diffusion Mechanism: Once a data item d is generated, it proceeds along the edges in $G(U, E)$ following the publishing manner:

Initialization: A data item is generated by user p_0. Denoting the initial set of receivers by $R_d(0)$ we have $R_d(0) = \{p_0\}$.

Step i : At any given step $i \geq 1$, suppose that the data has been received by a set $R_d(i)$ of users. Among them only the users in $P_d \cap R_d(i)$ publish the data to their neighbors.

Step $i+1$: $R_d(i+1)$ is updated by adding all these neighbors. The process terminates when $R_d(i+1) = R_d(i)$, that is, the set of receivers does not increase from one step to the next. Let R_d denote the set of receivers at the end of process. We have $R_d = R_d(i+1)$.

The system model includes two time scales: one for network evolution, denoted by time slot t, and the other for information diffusion, denoted by step i. We assume that the former time scale is more coarser than the latter one, which satisfies

- The network structure remains the same during a whole information diffusion process.

This assumption makes sense since transmission time of a data item is much smaller than that of network evolution, where the former one often takes several days while an evident change of network structure often takes several months or even longer.

C. Delivery Accuracy

The intuition of a good delivery accuracy includes two aspects: most of the interested users can receive the data and few of other users are involved. At any given time t, we capture the delivery accuracy of the information diffusion process via two metrics, i.e., *recall* and *precision*, where recall measures the fraction of users who receive the data in all viewers, and precision measures the fraction of interested users in all receivers. Furthermore, we present a notion called RP-Perfect [7], which consolidates the above two metrics.

Definition 3 (Recall and Precision)**.** *Recall of an arbitrary data item d at time t is defined as $\frac{|R_d(t) \cap V_d(t)|}{|V_d(t)|}$, and precision of it is defined as $\frac{|R_d(t) \cap V_d(t)|}{|R_d(t)|}$, where $R_d(t)$ denotes the set of receivers at time t.*

Definition 4 (RP-Perfect)**.** *The delivery accuracy of an information diffusion process is called RP-Perfect if $\frac{|R_d(t) \cap V_d(t)|}{|R_d(t) \cup V_d(t)|} = 1$. Or equivalently, both recall and precision equal to one.*

Note that in random networks, the above metrics are defined by considering the expected values, where recall is defined as $\frac{E[|R_d(t) \cap V_d(t)|]}{E[|V_d(t)|]}$, precision is defined as $\frac{E[|R_d(t) \cap V_d(t)|]}{E[|R_d(t)|]}$ and RP-Perfect holds if $\frac{E[|R_d(t) \cap V_d(t)|]}{E[|R_d(t) \cup V_d(t)|]} = 1$.

D. Notations

For convenience, we present Table I to list all notations that will be used in later analysis, proofs and discussions.

TABLE I
NOTATIONS AND DEFINITIONS

Notation	Definition
I_u	Interests set of user u
I_d	Interests set of data d
α	Parameter of connecting probability
β	Probability that a new user arrives
$V_d(t)$	Viewers set of data item d at time t
$P_d(t)$	Publishers set of data item d at time t
$R_d(t)$	Receivers set of data item d at time t
M	Number of interests in I_d
m	Threshold on the number of interests in data item, larger than which the user is a publisher of the data

IV. Main Results and Intuitions

In this section we make a summary on our main results and present intuitions of them. The corresponding proofs are given in later theorems and lemmas in Section V.

A. Main Results

Our theoretical analysis is made as follows: We first offer a special case where the delivery accuracy is RP-Perfect, based on which we then make an analysis and disclose the necessary conditions to guarantee the RP-Perfect, as presented in the first result. Then, we figure out how network structure evolves over time and list two main conclusions in the second result. And finally, we prove that all the necessary conditions presented in the first result can be satisfied when $t \to \infty$ and thus the delivery accuracy can approach to RP-Perfect in a completely evolving network. The results are presented as follows:

1) Necessary conditions: For the publishing based information diffusion mechanism, the following necessary conditions should be satisfied to achieve RP-Perfect:

- The core graph, i.e., subgraph of $G(U, E)$ induced by all publishers, is connected with probability 1.
- Every user $u \in V_d(t) - P_d(t)$ connects to the core graph with probability 1.
- Each publisher has a limited number of redundant interests which can guarantee that $\frac{E[|R_d(t)| - |R_d(t) \cap V_d(t)|]}{E[|V_d(t)|]} \to 0$.

The proof of this result is provided in Section V-A. The first two conditions are given to guarantee the network connectivity among viewers and thus promise the metric recall, while the third one indicates that it should be satisfied that the number of redundant users who received the data item can be neglected compared to that of viewers and we achieve this requirement by controlling the number of redundant interests that each publisher has, which promises the metric precision. At first glance, the above conditions can hardly be satisfied in realistic networks and consequently, the publishing based information diffusion mechanism may suffer a low delivery accuracy. However, we obtain a contrary result in our theoretical analysis that the delivery accuracy is improved over time and finally achieves RP-Perfect, as presented in the next two results.

2) Evolution of viewers and publishers: For an arbitrary data item d that is generated at time t_d, denoting the number of

viewers at time t as $|V_d(t)|$, its expected value, i.e., $E[|V_d(t)|]$, has an asymptotic order that satisfies

$$E[|V_d(t)|] = \Theta\left(\left(\frac{t}{t_d}\right)^{c_1}\right)^1.$$

Similarly, denote the number of publishers at time t as $|P_d(t)|$, and then its expected value, i.e., $E[|P_d(t)|]$, has an asymptotic order that satisfies

$$E[|P_d(t)|] = \begin{cases} \Theta\left(t^{1-(1-c_1)m}\right), & 0 < (1-c_1)m < 1 \\ \Theta\left(\log t\right), & (1-c_1)m = 1 \\ \Theta\left(1\right), & (1-c_1)m > 1, \end{cases}$$

where $c_1 = \frac{\beta c_u}{\beta c_u + (1-\beta)c_i}$.

The proofs are provided in Lemma 2 and Lemma 3 respectively. The above two parameters are important in determining the delivery accuracy, which have direct influence on whether the first two necessary conditions hold or not and thus we calculate them first to provide guidance on the derivation of delivery accuracy. From the result we can observe that both the number of viewers and that of publishers increase over time, which improves the network connectivity and further provides a better delivery accuracy. In addition, we then show how the delivery accuracy performs in limit where $t \to \infty$.

3) RP-Perfect in theoretical model: For an arbitrary data item d with interests set I_d, the delivery accuracy of information diffusion in evolving social networks with time $t \to \infty$ can achieve RP-Perfect, that is,

$$\lim_{t\to\infty} \frac{E[|R_d(t) \cap V_d(t)|]}{E[|R_d(t) \cup V_d(t)|]} = 1,$$

if it is satisfied that $1 - (1-c_1)m - c_1\alpha > 0$ and $m > 2$.

The corresponding proof is given in Theorem 1. This result indicates that under the publishing based information diffusion mechanism, the delivery accuracy approaches to RP-Perfect in a completely evolving social network where $t \to \infty$. The reason behind this result lies in that the evolving process of network is actually interest-guided, which makes the network structure become more and more similar to the desired one and finally results in RP-Perfect delivery accuracy in our theoretical analysis. An intuitive description on this reason will be provided in the next subsection. Moreover, we further validate this conclusion through a real dataset in Section VI and results indicate that the delivery accuracy shows the same trend in realistic networks.

B. Intuitions

Our results show that the delivery accuracy of publishing based information diffusion mechanism approaches to RP-Perfect with the evolution of network. Before we prove this result rigorously, we first give an intuitive insight here.

The essential reason behind this result is that the formation of new links in evolving social networks is interest-guided. It should be noted that the delivery accuracy of publishing based information diffusion mechanism in an interest-guided network is significantly higher than that in a stochastic one, where users sharing common interests are more likely to be connected in the former one while in the latter one, the creation of new links has no relation with users' interests. This result is an intuitive one since in the interest-guided network, the viewers of an arbitrary data item are always connected together which promises a high recall, and in the other hand, the redundant users are unlikely to connect to the publishers and thus avoid error receiving, which results in a high precision. In our theoretical model, the creation of new links is interest-guided where the new viewer has a greater chance to be connected to other existing viewers and that's the same for the new redundant users. Therefore, the evolution of network enhances the interest-guided network structure over time and thus provides the increased recall and precision.

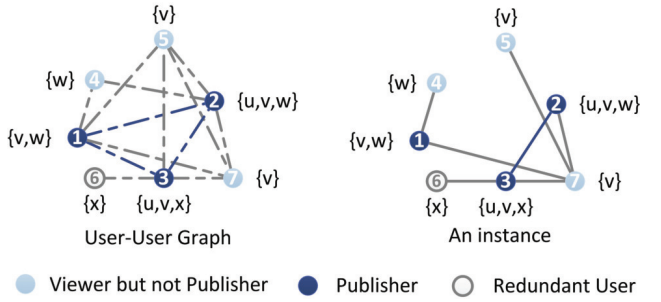


Fig. 2. An example of the publishing based information diffusion mechanism.

V. Delivery Accuracy of Information Diffusion Process in Evolving Social Networks

In this section, we first present a special case to show that the delivery accuracy can achieve RP-Perfect, based on which we figure out necessary conditions in order to guarantee such a good performance. Then, we expand the problem into general cases, and prove that under given parameter constrains, recall and precision can achieve one simultaneously in a completely evolving social network.

A. A Special Case

Before the description of this case, we first give a simple example here to show why some viewers cannot receive the data but some redundant users are involved.

Example (Understanding of the low delivery accuracy). As shown in Fig. 2, we assume that a data item d has interests set $I_d = \{u, v, w\}$ and each user has an interest profile as marked in the figure. The dashed lines in user-user graph denote edges between users, each of which exists with a certain probability. An instance of it is shown in the right subfigure. Assume node 2 is a source node. According to our information diffusion mechanism, only node 2, node 3, node 6 and node 7 can receive the data. In this case, since the core graph is disconnected, publisher 1 cannot receive the data. Viewer 5 cannot receive the data neither because it does not connect to the core graph. Though viewer 4 connects to a publisher, it still cannot receive the data because the publisher itself does

[1]We use standard asymptotic notations in our paper. Consider two nonnegative function $f(\cdot)$ and $g(\cdot)$: $f(n) = o(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) = 0$; $f(n) = O(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) < \infty$; $f(n) = \omega(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) = \infty$; $f(n) = \Omega(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) > 0$; $f(n) = \Theta(g(n))$ means $f(n) = \Omega(g(n))$ and $f(n) = O(g(n))$.

not receive the data. All these situations will result in a poor performance of recall. On the other hand, redundant user 6 will receive the data since it shares a redundant interest x with publisher 3, which has a negative influence on precision.

From the above example it seems that this kind of mechanism suffers a low delivery accuracy. While, we give a special case where the delivery accuracy can achieve RP-Perfect. Then based on it we make the analysis and answer the question:

What conditions (if any) are necessary to ensure that recall and precision can both achieve a good performance?

1) A RP-Perfect Network Structure: Now we give a special case where the delivery accuracy could achieve RP-Perfect. Some assumptions on network structure and user identifier are made more rigorously and defined as follows:

Connectivity: We assume that an edge exists between user u and user v if and only if they share at least one common interest, i.e., $I_u \cap I_v \neq \emptyset$. Note that this is a stronger assumption compared with that described in Section III.

Viewer and Publisher: The definition of viewer in this case is the same as that in Definition 1. The definition of publisher is a stronger one: For an arbitrary data item d, a user u is one of its publishers if and only if $I_u = I_d$, that is, every publisher has exactly the same interests set with that of data item d.

The proof of this result is presented here briefly. For any user $u \in V_d$, we have $I_u \cap I_d \neq \emptyset$. Since the source node v is a publisher with $I_v = I_d$, we have $I_u \cap I_v \neq \emptyset$. Thus, there must be an edge existing between them, and viewer u can obtain the data from publisher v. This result indicates that every user interested in the data item d can receive it, and thus we have that recall equals to one. On the other hand, any user u who received the data item d must connect to at least one publisher v, which means that $I_u \cap I_v \neq \emptyset$. According to the definition of publisher, we have that $I_v = I_d$ and thus $I_u \cap I_d \neq \emptyset$, that is, user u is a viewer of the data item d. This result indicates that every receiver is interested in the data item d, and thus we have that precision in this case equals to one. Since both recall and precision can achieve one in this case, we have proved that the performance of it is RP-Perfect.

2) Necessary Conditions: Although the requirements in this case are difficult to satisfy, they can help us to figure out how to guarantee a RP-Perfect delivery accuracy.

To guarantee that recall equals to one, we should make sure that all the viewers can receive the data. Since that publishers can transmit the data among themselves while other viewers can only receive the data from publishers, the following two conditions should be satisfied:

- The core graph, i.e., subgraph of $G(U, E)$ induced by all publishers, is connected with probability 1.
- Every user $u \in V_d(t) - P_d(t)$ connects to the core graph with probability 1.

On the other hand, to guarantee that precision equals to one, no redundant users can be involved. Thus, each redundant user in the network can not connect to the publisher, which requires that the interests of each publisher are exactly the same with that of the data since any redundant interest may result in the neighbor of redundant users and thus followed by a loss of precision. The requirement is difficult to satisfy in most cases. However, we could relax it in large scale networks as long as the number of involved redundant users can be neglected compared with that of viewers. The necessary condition after relaxation is presented as follows:

- Each publisher has a limited number of redundant interests which can guarantee that $\frac{E[|R_d(t)|-|R_d(t)\cap V_d(t)|]}{E[|V_d(t)|]} \to 0$.

TABLE II
INTEREST AND SEARCH INDEX IN TWO DATA ITEMS

Interest in d_1	Search Index	Interest in d_2	Search Index
"Oscar"	478	"TOEFL"	4,700
"Spotlight"	658	"IELTS"	5,876

[1] d_1: "The movie Spotlight won Best Picture Oscar."
[2] d_2: "TOEFL v.s. IELTS: which is better?"

At first glance, the aforementioned conditions are difficult to satisfy in realistic evolving social networks. However, we find that under certain parameter constraints, the network structure characterized by the affiliation networks model becomes more and more similar to the desired one with the evolution of network and even more surprisingly, both recall and precision converge to one when $t \to \infty$. In the following parts, we will verify our results through both theoretical analysis and experimental measurements respectively.

B. General Case

In this subsection, we will present a general analysis on the delivery accuracy of publishing based information diffusion mechanism. The analysis approach is organized as: pick an arbitrary data item d and analyze the information diffusion process in terms of this chosen data item d as time goes on. Denote the excepted value of $d_i(\cdot)$ as $E[d_i(\cdot)]$. We first show how $E[d_i(t)]$ and $E[d_u(t)]$ evolve with time t (Lemma 1).

Then, we come to the expected number of viewers and that of publishers separately. Results show that $E[|V_d(t)|]$ increases with the same rate as that of $E[d_i(t)]$ (Lemma 2); $E[|P_d(t)|]$ increases or remains the same in order sense under different parameters (Lemma 3). Based on the above results, we further analyze the evolution of recall and precision. We confirm that under certain parameter constrains, the delivery accuracy in a completely evolving network can achieve RP-Perfect, i.e., both recall and precision approach one (Theorem 1). And finally, we make a detailed discussion on our results.

1) Assumptions: An assumption used in our later theoretical analysis and proofs is given here.

- For an arbitrary data item d with M interests, the expected value of degrees of all these M interests are in the same order, i.e., $\Theta\left(E[d_{i_k}(t)]\right) = \Theta\left(E[d_i(t)]\right), \forall i_k \in I_d$.

We note that the assumption may can not fully characterize all types of data structure, which, however, still holds in many cases. Intuitively, interests in a same data item are often related with each other and therefore their degrees are similar in the order sense. For example, consider two data items: *The movie Spotlight won Best Picture Oscar.* and *TOEFL v.s. IELTS: which is better?*, and denote them by d_1 and d_2 respectively. There are two major interests included in d_1, i.e., *Spotlight* and *Oscar*, and two included in d_2, i.e., *TOEFL* and *IELTS*.

Then, we utilize Baidu Search Index[2] to measure an interest's search times in Baidu, which can be interpreted as its degree since the more times an interest was searched in the Internet, the more people who are interested in it. Table II shows the results. We can observe that the indexes of interests in a same data item are similar while vary a lot in different data items, which verifies the rationality of our assumption.

2) Useful Properties of Information Diffusion in Evolving Social Networks: Before the analysis of delivery accuracy, we first show some useful properties of information diffusion process in evolving social networks, i.e., the evolving process of $E[d_u(t)]$, $E[d_i(t)]$, $E[|V_d(t)|]$ and $E[|P_d(t)|]$.

***Lemma* 1.** *For an interest i (a user u) in $B_t(U,I)$, with the initial condition that i (u) arrives at the network at time t_i (t_u), if $t \to \infty$, w.h.p., the expected value of $d_i(t)$ ($d_u(t)$) has an asymptotic order that satisfies*

$$E[d_i(t)] = \Theta\left(\left(\frac{t}{t_i}\right)^{c_1}\right) \quad (E[d_u(t)] = \Theta\left(\left(\frac{t}{t_u}\right)^{1-c_1}\right),$$

where $c_1 = \frac{\beta c_u}{\beta c_u + (1-\beta)c_i}$.

Proof. In order to calculate the expected value of $d_i(t)$, i.e., $E[d_i(t)]$, we consider the randomness of variable $d_i(t)$ in two aspects: the randomness in the update of $d_i(t-1)$, for a given $B_{t-1}(U,I)$; the randomness over all the possible $B_{t-1}(U,I)$.

Firstly, let us consider the update of $d_i(t-1)$. According to the evolving process of $B(U,I)$, we know that the degree of interest i increases only if a new user is added to $B(U,I)$ and one of its edges points to i. Note that in the random process the endpoint of any edge in I is chosen with equal probability as a destination of the new edge. The probability that a new added edge points to interest i is $\frac{d_i(t-1)}{e(t-1)}$, where $e(t-1)$ denotes the number of edges in $B_{t-1}(U,I)$. Therefore, given a certain $B_{t-1}(U,I)$, $d_i(t) = d_i(t-1)+1$ when a user arrives and points to i, which happens with probability $\beta c_u \frac{d_i(t-1)}{e(t-1)}$; $d_i(t) = d_i(t-1)$, otherwise. Under a particular $B_{t-1}(U,I)$, the conditional expected value of $d_i(t)$ is

$$\begin{aligned}&E[d_i(t)|B_{t-1}(U,I)]\\ =&E[d_i(t-1)|B_{t-1}(U,I)] + \beta c_u \frac{E[d_i(t-1)|B_{t-1}(U,I)]}{e(t-1)}.\end{aligned}$$

Secondly, sum over all the possible $B_{t-1}(U,I)$ and we have

$$E[d_i(t)] = E[d_i(t-1)]\left(1 + \frac{\beta c_u}{e(t-1)}\right). \tag{1}$$

Before the further calculation of Equation (1), we first make a discussion on $e(k)$ – the number of edges in $B_k(U,I)$ at time k. Using the Chernoff Bound [30] we have

$$e(k) = \begin{cases} E[e(k)] \pm o(k), & k = \omega(1) \\ E[e(k)] \pm \Theta(k), & k = \Theta(1), \end{cases}$$

where $E[e(k)] = (\beta c_u + (1-\beta)c_i)\,k + e(0)$.

Now, Equation (1) could be calculated into two cases: when $t_i = \omega(1)$, we have

$$\begin{aligned}E[d_i(t)] =& d_i(t_i) \prod_{t_i \le k \le t-1} \left(1 + \frac{\beta c_u}{(\beta c_u + (1-\beta)c_i)k \pm o(k) + e(0)}\right)\\ =& d_i(t_i) \prod_{t_i \le k \le t-1} \left(1 + \frac{c_1}{k}\right) = d_i(t_i)\frac{\Gamma(t-1+c_1)}{\Gamma(t-1)}\frac{\Gamma(t_i)}{\Gamma(t_i+c_1)}\\ =& d_i(t_i)\left(\frac{t}{t_i}\right)^{c_1},\end{aligned}$$

and when $t_i = \Theta(1)$, we have

$$\begin{aligned}E[d_i(t)] =& d_i(t_i) \prod_{\substack{t_i \le k \le t-1\\ k=\omega(1)}} \left(1 + \frac{\beta c_u}{(\beta c_u + (1-\beta)c_i)k \pm o(k) + e(0)}\right)\\ &\cdot \prod_{\substack{t_i \le k \le t-1\\ k=\Theta(1)}} \left(1 + \frac{\beta c_u}{(\beta c_u + (1-\beta)c_i)k \pm \Theta(k) + e(0)}\right)\\ =& d_i(t_i) \prod_{t_i \le k \le t-1} \left(1 + \frac{c_1}{k}\right) \prod_{\substack{t_i \le k \le t-1\\ k=\Theta(1)}} \Delta_k\\ =& d_i(t_i)\frac{\Gamma(t-1+c_1)}{\Gamma(t-1)}\frac{\Gamma(t_i)}{\Gamma(t_i+c_1)} \prod_{\substack{t_i \le k \le t-1\\ k=\Theta(1)}} \Delta_k\\ =& d_i(t_i)t^{c_1}\frac{\Gamma(t_i)}{\Gamma(t_i+c_1)} \prod_{\substack{t_i \le k \le t-1\\ k=\Theta(1)}} \Delta_k,\end{aligned}$$

where $\Gamma(\cdot)$ denotes Gamma Function that has a useful limit for asymptotic approximation, i.e., $\lim_{n\to\infty} \frac{\Gamma(n+\alpha)}{\Gamma(n)} = n^\alpha$, and the factor $\Delta_k = \left(1 + \frac{\beta c_u}{(\beta c_u + (1-\beta)c_i)k \pm \Theta(k) + e(0)}\right) / \left(1 + \frac{c_1}{k}\right)$ satisfies that $\Delta_k = \Theta(1)$ when $k = \Theta(1)$. Moreover, note that $\Gamma(t_i+c_1)$, $\Gamma(t_i)$, $d_i(t_i)$ and $\prod_{t_i \le k \le t-1, k=\Theta(1)} \Delta_k$ are all finite ones and we could obtain a uniform expression of $d_i(t)$ in the two cases as given in the Lemma. The proof of user degree follows a symmetrical method and thus we omit it here. □

Lemma 1 shows that the expected value of degree of interest grows with time t. Then based on the above result, we show how $E[|V_d(t)|]$ and $E[|P_d(t)|]$ evolve with time t.

***Lemma* 2.** *For an arbitrary data item d that is generated at time t_d, denoting the number of viewers at time t as $|V_d(t)|$, its expected value, i.e., $E[|V_d(t)|]$, has an asymptotic order that satisfies*

$$E[|V_d(t)|] = \Theta\left(\left(\frac{t}{t_d}\right)^{c_1}\right),$$

where $c_1 = \frac{\beta c_u}{\beta c_u + (1-\beta)c_i}$.

Proof. In evolving social networks, the degree of interest increases with time t, which indicates that the earliest joined interest has the largest degree. Based on this observation, letting i_0 denote the earliest joined interest, we have

$$d_{i_0}(t) = \max_{i \in I_d} d_i(t).$$

Since the data generation time t_d is set as the one when node i_0 joins the network, we have

$$E[d_{i_0}(t)] = \Theta\left(\left(\frac{t}{t_d}\right)^{c_1}\right).$$

For any interest $i \in I_d$, each of its neighbors in $B(U,I)$ is a

[2]Baidu Search Index (https://zhishu.baidu.com/) is a high search volume keywords and top keywords tool to see the searching times of a particular keyword in Baidu search engine (https://www.baidu.com).

viewer of data item d. Thus, we have $E[|V_d(t)|] \geq E[d_{i_0}(t)]$. In addition, every viewer of data item d links to at least one interest $i \in I_d$, i.e., $E[|V_d(t)|] \leq \sum_{i \in I_d} E[d_i(t)] \leq M \cdot E[d_{i_0}(t)]$. Since M is a constant, we have that $E[|V_d(t)|]$ is in the same order of $E[d_{i_0}(t)]$, and therefore we complete the proof. □

As expected, Lemma 2 shows that $E[|V_d(t)|]$ is in the same order of its interest degree. The definition of graph $B(U, I)$ is an explanation of the result, by noting the degree of interest i indicates the number of users interested in it. Since that only publishers can transmit the data to viewers while involve some redundant users at the same time, they play a very important role in information diffusion process. Lemma 3 shows how the number of publishers evolves with time t.

***Lemma* 3.** *For an arbitrary data item d that is generated at time t_d, when $t = \omega(t_d)$, denoting the number of publishers at time t as $|P_d(t)|$, its expected value, i.e., $E[|P_d(t)|]$, has an asymptotic order that satisfies*

$$E[|P_d(t)|] = \begin{cases} \Theta\left(t^{1-(1-c_1)m}\right), & 0 < (1-c_1)m < 1 \\ \Theta\left(\log t\right), & (1-c_1)m = 1 \\ \Theta\left(1\right), & (1-c_1)m > 1, \end{cases}$$

where $c_1 = \frac{\beta c_u}{\beta c_u + (1-\beta)c_i}$.

Proof. According to the definition of $P_d(t)$, a user u may become a publisher of the data item d only at the time when it is added to the network; or it will never become a publisher in the following network evolving process.

For a new added user u with c_u edges, at least m edges of which should point to the interests $i_k \in I_d$. Note that there are $\binom{c_u}{m}$ possible combinations of these m edges, and each of them connects to one of interests $i_k \in I_d$ with probability $\frac{d_{i_k}(t-1)}{e(t-1)}$. Under a particular $B_{t-1}(U, I)$, the conditional expected value of $|P_d(t)|$ is

$$\begin{aligned} &E[|P_d(t)|\,|B_{t-1}(U,I)] \\ =&E[|P_d(t-1)|\,|B_{t-1}(U,I)] + \beta\binom{M}{m}\binom{c_u}{m}\prod_{k=1}^{m}\frac{E[d_{i_k}(t-1)|B_{t-1}(U,I)]}{e(t-1)} \end{aligned}$$

Sum over all the possible $B_{t-1}(U, I)$ and we have

$$E[|P_d(t)|] = E[|P_d(t-1)|] + \beta\binom{M}{m}\binom{c_u}{m}\prod_{k=1}^{m}\frac{E[d_{i_k}(t-1)]}{e(t-1)}$$

Then, $E[|P_d(t)|]$ can be calculated using an iterative method.

$$E[|P_d(t)|] = \left|P_d(t_g)\right| + \beta\binom{M}{m}\binom{c_u}{m}\sum_{j=t_g}^{t-1}\prod_{k=1}^{m}\frac{E[d_{i_k}(j)]}{e(j)},$$

where t_g denotes the time when the M-th interest in I_d enters the network and therefore we have $|P_d(t_g)| \leq (c_i + 1)(t_g - t_d) = \Theta(1)$. Since that $E[d_{i_k}(t)] = \Theta\left(E[d_{i_0}(t)]\right), \forall i_k \in I_d$, we have $E[d_{i_k}(t)] = \Theta\left(\left(\frac{t}{t_d}\right)^{c_1}\right)$. We assume that t_p is a variable satisfying $t_p > t_g$, $t_p \to \infty$ and $t_p = o(t)$. Using lemma 1, the above equation can be simplified to

$$\begin{aligned} &E[|P_d(t)|] \\ =&\Theta\left(\beta\binom{M}{m}\binom{c_u}{m}\sum_{j=t_p}^{t-1}\left(\frac{1}{j}\left(\frac{j}{t_d}\right)^{c_1}\right)^m\right) = \Theta\left({t_d}^{-c_1 m}\sum_{j=t_p}^{t-1}\frac{1}{j^{m(1-c_1)}}\right). \end{aligned}$$

Then, since that $t = \omega(t_d)$ and $t = \omega(t_p)$, the mathematical result of this equation can be calculated using the sum of *p-series*, that is,

$$\lim_{n\to\infty}\sum_{x=1}^{n}\frac{1}{x^p} = \begin{cases} \Theta\left(n^{1-p}\right), & 0 \leq p < 1 \\ \Theta\left(\log n\right), & p = 1 \\ \Theta\left(1\right), & p > 1. \end{cases}$$

Thus we complete the proof. □

In order to ensure that most viewers can receive the data, there have to exist enough publishers. An intuitive thought is that the information diffusion has poor performance in case $E[|P_d(t)|] = \Theta(1)$, since the number of publishers is limited while the number of viewers approaches to infinity when $t \to \infty$. In this case, it is hard to guarantee a high recall. Actually, this thought is proved to be true in the following derivations. And we will show that the information diffusion also performs negatively in terms of recall even under the case $E[|P_d(t)|] = \Theta(\log t)$. Only in the case $E[|P_d(t)|] = \Theta\left(t^{1-(1-c_1)m}\right)$ with certain parameter constrains, it can achieve one.

3) Evolution of Recall and Precision: In the following part, we will come to the performance of recall and precision in a completely evolving network, i.e., the one with $t \to \infty$. Before that, we present three useful Lemmas on network connectivity, average number of common interests between users and degree distribution in $B(U, I)$, respectively.

***Lemma* 4.** *Let $G(n, p)$ denote the random graph with n nodes where any two nodes in it are connected with probability p. When $p = \frac{\log n + c_n}{n}$, we have*

$$\lim_{n\to\infty} Pr\{G(n,p) \text{ is connected}\} = \begin{cases} 0, & c_n \to -\infty \\ e^{-e^{-c}}, & c_n \to c \\ 1, & c_n \to +\infty, \end{cases}$$

where c is a constant. [31]

***Lemma* 5.** *For any two users u and v in the network, denoting the number of common interests shared by them as $|I_u \cap I_v|$, its expected value, i.e., $E_i[|I_u \cap I_v|]$, has an asymptotic order that satisfies*

$$E_i[|I_u \cap I_v|] = \Theta(1).$$

Proof. Note that the common interests shared by user u and user v may be generated in an arbitrary time slot. We calculate $E_i[|I_u \cap I_v|]$ by dividing the interests into two parts: the ones generated in time period $t = \Theta(1)$ and that generated in time period $t = \omega(1)$. Denoting the expected number of the former part of interests as $E[|i \in I_u \cap I_v|\,|t_i = \Theta(1)]$ and that that of the latter part as $E[|i \in I_u \cap I_v|\,|t_i = \omega(1)]$, we have

$$\begin{aligned} &E_i[|I_u \cap I_v|] \\ =&E[|i \in I_u \cap I_v|\,|t_i = \Theta(1)] + E[|i \in I_u \cap I_v|\,|t_i = \omega(1)]. \end{aligned}$$

Note that in each time slot, at most one interest can be added. Thus we have

$$E[|i \in I_u \cap I_v|\,|t_i = \Theta(1)] = \Theta(1).$$

Let t_0 denote the beginning time of the time period $t = \omega(1)$. We have that $t_0 = \omega(1)$ and $t \in [t_0, +\infty)$. There are at most

$t - t_0$ interests generated during the time period $t = \omega(1)$. In this time period, the expected number of new added common interests can be obtained by using the law of total probability.

$$E[|i \in I_u \cap I_v| \,|t_i = \omega(1)] = \sum_{x=1}^{t-t_0} x \cdot Pr\{|I_u \cap I_v| = x | t = \omega(1)\} = \sum_{x=1}^{t-t_0} x \prod_{k=1}^{x} p_{i_k}(u,v) \prod_{j=x+1}^{t-t_0} \left(1 - p_{i_j}(u,v)\right),$$

where $p_{i_k}(u,v) = \frac{E[d_u(t_k)]}{e(t_k)} \cdot \frac{E[d_v(t_k)]}{e(t_k)}$ denotes the probability that interest i_k added at time t_k connects to user u and user v simultaneously. The values of $E[d_u(t_k)]$ and $E[d_v(t_k)]$ can be obtained through Lemma 1. Note that $p_{i_k}(u,v) \leq \Theta\left(\frac{1}{{t_d}^{2c_1}}\right)$ and $1 - p_{i_k}(u,v) \leq 1 - \Theta\left(\frac{1}{t^{2c_1}}\right)$. The above equation can be simplified to

$$\begin{aligned}
&E[|i \in I_u \cap I_v| \,|t_i = \omega(1)] \\
\leq &\Theta\left(\sum_{x=1}^{t-t_0} x \left(\frac{1}{{t_d}^{2c_1}}\right)^x \left(1 - \frac{1}{t^{2c_1}}\right)^{t-t_0-x}\right) \\
\leq &\Theta\left(\left(1 - \frac{1}{t^{2c_1}}\right)^{t-t_0} \cdot \sum_{x=1}^{\infty} x \left(\frac{t^{2c_1}}{{t_d}^{2c_1}\left(t^{2c_1}-1\right)}\right)^x\right) \\
\leq &\Theta(1) \cdot \Theta\left(\sum_{x=1}^{\infty} x \left(\frac{t^{2c_1}}{{t_d}^{2c_1}\left(t^{2c_1}-1\right)}\right)^x\right).
\end{aligned}$$

Since that $\sum_{x=1}^{\infty} x\left(\frac{t^{2c_1}}{{t_d}^{2c_1}\left(t^{2c_1}-1\right)}\right)^x$ is convergent, we have

$$E[|i \in I_u \cap I_v| \,|t_i = \omega(1)] \leq \Theta(1).$$

Combining the two part of results, we complete the proof. □

***Lemma* 6.** *For the bipartite graph $B(U, I)$ generated after t time slots, almost surely, when $t \to \infty$, the degree sequence of nodes in I follows a power-law distribution with exponent $-2 - \frac{c_i(1-\beta)}{c_u\beta}$. [4]*

Theorem 1. *For an arbitrary data item d with interests set I_d, the delivery accuracy of information diffusion in a completely evolving network, i.e., $t = \omega(t_d)$, can achieve RP-Perfect*

$$\lim_{t\to\infty} \frac{E[|R_d(t) \cap V_d(t)|]}{E[|R_d(t) \cup V_d(t)|]} = 1,$$

if it is satisfied that $1 - (1 - c_1)m - c_1\alpha > 0$ and $m > 2$.

Proof. First, we calculate the connecting probability between two users. For any two users u and v, the probability $p_{u,v}$ that they are connected depends on the common interests between them, which can be calculated as,

$$\begin{aligned}
p_{u,v} &= \sum_{k=1}^{\infty} \left(\frac{1}{c_0} \sum_{j \in I_u \cap I_v} \frac{1}{E[d_j(t)]^\alpha}\right) Pr\{|I_u \cap I_v| = k\} \\
&\geq \sum_{k=1}^{\infty} \frac{k}{c_0 E[d_{max}(t)]^\alpha} Pr\{|I_u \cap I_v| = k\} \\
&= \frac{E_i[|I_u \cap I_v|]}{c_0 d_{max}(t)^\alpha} \overset{(a)}{=} \Theta\left(\frac{1}{E[d_{max}(t)]^\alpha}\right),
\end{aligned} \tag{2}$$

where $E[d_{max}(t)] = \max\{E[d_i(t)], i \in I_u \cap I_v\}$. Equality (a) holds according to Lemma 5. Then based on this result, we consider the problem in the following three cases: $(1-c_1)m > 1$, $(1 - c_1)m = 1$ and $0 < (1 - c_1)m < 1$.

Case 1: $(1 - c_1)m > 1$. In this case, $E[|P_d(t)|] = \Theta(1)$ according to Lemma 3. Assume that there are c publishers who received the data. Obviously, $c \leq |P_d(t)|$. Then, for any user $u \in V_d(t) - P_d(t)$, the probability $p_{u,con}$ that user u can receive the data is equivalent to the one that it connects to at least one of c publishers. Letting p_{u,v_i} denote the probability that user u connects to publisher v_i, we have

$$p_{u,con} = 1 - \prod_{i=1}^{c} (1 - p_{u,v_i}) < 1$$

Since it can not be ensured that user u could receive the data, we have proved that in this case, recall can not achieve one.

Case 2: $(1 - c_1)m = 1$. In this case, $E[|P_d(t)|] = \Theta(\log t)$. We calculate the connectivity of core graph using Lemma 4. Since that every pair of users shares different amounts of interests, the probability that they are connected varies from pair to pair. For the convenience of calculations, we introduce a lower bound and a upper bound of probability $p_{u,v}$, denoted by p_{min} and p_{max} respectively, satisfying that $p_{min} \leq p_{u,v} \leq p_{max}, \forall u, v \in U$. Then, there is an apparent conclusion that, if the core graph is connected when $p = p_{min}$, it can be connected when $p = p_{u,v}$. Based on this observation, we give a lower bound $p_{min} = \Theta\left(t^{-c_1\alpha}\right)$ and show that under this bound the core graph is connected with probability 1. And thus we can obtain the actual connectivity of the core graph. The core graph can be modeled as a random graph $G(n, p)$ with parameters $n = E[|P_d(t)|]$ and $p = p_{min}$. The probability p can be expressed as

$$p = \frac{\log E[|P_d(t)|] + c_n}{E[|P_d(t)|]}.$$

Thus, we have that in a completely evolving network,

$$c_n = p_{min}E[|P_d(t)|] - \log E[|P_d(t)|] = \Theta\left(\frac{\log t}{t^{c_1\alpha}} - \log\log t\right).$$

Since $c_n \to -\infty$ when $t \to \infty$, the connectivity of the core graph can not be guaranteed according to Lemma 4.

Case 3: $0 < (1 - c_1)m < 1$. In this case, $E[|P_d(t)|] = \Theta\left(t^{1-(1-c_1)m}\right)$. Following the same method as in case 2, we will prove that the core graph in this case is connected with probability 1. The probability p can be expressed as

$$p = \frac{\log E[|P_d(t)|] + c_n}{E[|P_d(t)|]}.$$

Thus, we have that in a completely evolving network,

$$\begin{aligned}
c_n &= p_{min}E[|P_d(t)|] - \log E[|P_d(t)|] \\
&= \Theta\left(\frac{t^{1-(1-c_1)m}}{t^{c_1\alpha}} - \log t^{1-(1-c_1)m}\right) \\
&= \Theta\left(t^{1-(1-c_1)m-c_1\alpha} - \log t\right).
\end{aligned}$$

In this case, $c_n \to +\infty$ if $1 - (1 - c_1)m - c_1\alpha > 0$. Then using Lemma 4, we can prove that the core graph is connected with probability 1, if it is satisfied that $1 - (1 - c_1)m - c_1\alpha > 0$.

Now, let us consider the user $u \in V_d - P_d$. Since this part of users can only receive the data from publishers, they must

connect to the core graph. For any user $u \in V_d - P_d$, let $p_{u,core}$ denote the probability that it connects to the core graph. By noting that the probability that user u connects to an arbitrary publisher satisfies $p_{u,v} \geq t^{-c_1\alpha}$, we have

$$p_{u,core} = 1 - \prod_{i=1}^{|P_d(t)|} (1 - p_{u,v_i}) \geq 1 - \left(1 - \frac{1}{t^{c_1\alpha}}\right)^{t^{1-(1-c_1)m}}.$$

Since that $\lim_{x\to+\infty}\left(1-\frac{1}{x}\right)^x = \frac{1}{e}$, we have

$$p_{u,core} = \begin{cases} c, & 1-(1-c_1)m \leq c_1\alpha \\ 1, & 1-(1-c_1)m > c_1\alpha, \end{cases}$$

where $0 < c < 1$ is a constant.

Combining this result and the one in order to ensure the connectivity of core graph, we know that every viewer of the data can receive it if the aforementioned conditions are satisfied. Thus we conclude that in a completely evolving network, recall can achieve one if $1-(1-c_1)m - c_1\alpha > 0$.

We consider the performance of precision in this case. As defined, precision is expressed as $\frac{E[|R_d(t)\cap V_d(t)|]}{E[|R_d(t)|]}$. In order to achieve high precision, the number of redundant users should be as small as possible. Consequently, we will first show how to calculate the expected number of involved redundant users, i.e., $E[|R_d(t)| - |R_d(t)\cap V_d(t)|]$.

In the information diffusion process, every publisher who received the data may transmit it to some redundant users, and we denote this part of publishers as $\hat{P}_d(t)$. Then, for each $u \in \hat{P}_d(t)$, it has $|I_u| - |I_u \cap I_d|$ interests that may involve redundant users. Let $E[N_i(t)]$ denote the expected number of redundant users involved by an arbitrary interest i and we have

$$E[|R_d(t)| - |R_d(t)\cap V_d(t)|] \leq \sum_{u\in\hat{P}_d(t)} \sum_{i\in I_u - I_u\cap I_d} E[N_i(t)]. \quad (3)$$

The equality holds only when all redundant interests introduced by different publishers are not overlapped. For the interest i with degree $d_i(t)$, the average number of redundant users it involves is $d_i(t)\cdot\frac{1}{d_i(t)^\alpha}$. Then, the value of $E[N_i(t)]$ for an arbitrary interest i can be calculated using the law of total probability, that is,

$$E[N_i(t)] = \sum_{x=1}^{\infty} x\frac{1}{x^\alpha}\cdot Pr\{d_i(t) = x\}.$$

According to Lemma 6, we have

$$E[N_i(t)] = \Theta\left(\sum_{x=1}^{\infty} x^{-1-\frac{c_i(1-\beta)}{cu\beta}-\alpha}\right) = \Theta(1),$$

where the second equality holds due to the sum of *p-series*. Then, the Equation (3) can be calculated as

$$\begin{aligned} E[|R_d(t)| - |R_d(t)\cap V_d(t)|] &\leq \left|\hat{P}_d(t)\right| E[d_u(t)]E[N_i(t)] \\ &= \left|\hat{P}_d(t)\right| \Theta\left(t^{1-c_1}\right). \end{aligned} \quad (4)$$

Note that the number of redundant interests introduced by publisher u is definitely less than its degree, i.e., $d_u(t)$. Thus the number of items in the second summation is no greater than $d_u(t)$. Then using Lemma 1, we have the above result.

According to the result on recall, we know that in this case, $\frac{E[|R_d(t)\cap V_d(t)|]}{E[|V_d(t)|]} = 1$ if it is satisfied that $1-(1-c_1)m-c_1\alpha > 0$. Then under this condition, we will show that precision can also achieve one if $m > 2$. According to Equation (4), we have

$$\begin{aligned} E[|R_d(t)| - |R_d(t)\cap V_d(t)|] &\leq \left|\hat{P}_d(t)\right| \Theta\left(t^{1-c_1}\right) \\ &= \Theta\left(t^{2-c_1-(1-c_1)m}\right). \end{aligned}$$

Since that $\frac{E[|R_d(t)\cap V_d(t)|]}{E[|V_d(t)|]} = 1$ holds in this case, precision can be expressed as

$$\begin{aligned} \frac{E[|R_d(t)\cap V_d(t)|]}{E[|R_d(t)|]} &= \frac{E[|R_d(t)\cap V_d(t)|]}{E[|V_d(t)|] + E[|R_d(t)|] - E[|R_d(t)\cap V_d(t)|]} \\ &= \frac{\frac{E[|R_d(t)\cap V_d(t)|]}{E[|V_d(t)|]}}{1 + \frac{E[|R_d(t)|-|R_d(t)\cap V_d(t)|]}{E[|V_d(t)|]}} \\ &\geq \frac{1}{1+\Theta\left(\frac{t^{2-c_1-(1-c_1)m}}{t^{c_1}}\right)}. \end{aligned}$$

Obviously, it equals to 1 if $2-c_1-(1-c_1)m < c_1$, i.e., $m > 2$.

Combining the above results, we have that in a completely evolving network, recall and precision can achieve one simultaneously if $1-(1-c_1)m-c_1\alpha > 0$ and $m > 2$. □

C. Discussion on Results

In this part, we make a detailed discussion on our results.

1) Number of Viewers and Publishers: As shown in Lemma 2 and Lemma 3, the number of viewers at time t is $\Theta\left(t^{c_1}\right)$ and that of publishers is $\Theta\left(t^{1-(1-c_1)m}\right)$ when $0 < (1-c_1)m < 1$. By noting the physical meanings of parameters c_1 and m, we can conclude that the number of viewers mainly depends on the evolution of network and the number of publishers depends on both evolving process and user behavior. Obviously, both the number of viewers and that of publishers grow with parameter c_1. Moreover, a larger m leads to a more rigorous definition of publishers, which results in a smaller number of publishers.

2) Recall and Precision: To get a high recall, the number of publishers in viewers should maintain a certain ratio. The result in Theorem 1 indicates that the number of publishers should be no less than the α-power of viewers. Then, in order to achieve a high precision, the definition of publishers should be a rigorous one, i.e., $m > 2$. This is because that though there are many redundant users sharing the redundant interests with publishers, the relationship between them is weak. Since the redundant users just share interests with several publishers randomly, the probability that there exists a real link is small. However, a viewer always share interests, i.e., the ones of data, with most publishers, and thus the relationship between them is a strong one. Consequently, though the number of redundant users is larger than that of viewers, the ones among them that actually receive the data are not so many. Thus it can achieve a high precision. According to Theorem 1, we conclude that in the publishing based information diffusion mechanism, the following conditions should be satisfied to achieve RP-Perfect.

- The network for the data item should be a completely evolving one, i.e., evolving time $t \to \infty$.
- The publishers should have more than 2 interests related to the data item.
- In the evolving process, the number of edges added by users should be greater than that by interests, i.e., $c_1 > 0.5$, such that $0 < m(1-c_1) < 1$.

TABLE III
DATASET STATISTICS

Statistical Item	Value	Statistical Item	Value
# of Nodes	4,039	# of Common Attributes	10
# of Edges	88,234	Diameter	8

TABLE IV
SIMULATION SETTINGS

Parameter	Value	Parameter	Value
c_u	5	M	5
c_i	100	m	2
β	0.9975		

- Parameter α should be sufficiently small in order to ensure the network connectivity.

We note that the above conditions hold in certain real social networks. The first condition is easy to be satisfied since many real social networks have been observed to grow rapidly over time. The second condition is also a reachable one since that in certain social networks, users are only willing to republish the data that is closely related to their interests, i.e., data with more than $m = 2$ interests that the user feels interested in. The third condition $c_1 = \frac{\beta c_u}{\beta c_u + (1-\beta)c_i} > 0.5$, or equally $\beta c_u > (1-\beta)c_i$, indicates the average number of connections in $B(U, I)$ added due to user arrival is larger than that of connections added due to interest arrival, which holds in certain social networks such as lifestyle sharing networks where the number of users grows much faster than that of interests. And finally, the fourth condition is included to ensure the network connectivity, which generally holds since most real social networks are connected ones. Last but not least, as we will validate in our experimental measurements, even for a diffusion where the above conditions are not fully satisfied, it owns a good delivery accuracy.

VI. EXPERIMENTAL MEASUREMENT

In this section, we first give a description on the dataset, based on which we then validate our theoretical model and finally show how recall and precision evolve over time.

A. Dataset Description and Simulation Settings

Our experimental measurements are conducted on a dataset [8] collected from Facebook, which includes 4,039 users and 88,234 edges. The statistical properties of it are summarized in Table III. This dataset has a small diameter 8 and a power-law degree distribution as shown in Fig. 7. These properties are consistent with the widely observed ones and thus results in this dataset can well represent that in realistic networks.

1) User Profile: In this dataset, we pick up 10 nonoverlapping interests with various degrees as given in Table III. Each user has a profile recording its own interests within them. To protect user privacy, the interpretation of features is obscured. For example, where the original dataset may have contained a feature "political = Democratic Party", the new one would simply contain "political = anonymized feature 1". Thus, using the anonymized data, it is possible to determine whether two users have the same features, instead of their specific privacy. From the anonymized data, we can still get some implications of interests. For example, the No. 8 interest in Table V with the largest degree 3,279 is labeled as "location = anonymized feature 128", which is possible to be a common city among all the 4,039 users in the network.

2) Evolving Time: Note that in our model, we study the evolving process of a specific data item and set the initial time $t = 0$ as the time when the first interest of it joins the network. Therefore, given a social network, whether it is a completely evolving one varies with data items. For example, assume that the network is established at time $t = 0$, the interests of data item d_1 is generated at time t_1 and that of data item d_2 is generated at time t_2. The evolving time of d_1 and that of d_2 are different, which are $t-t_1$ and $t-t_2$, respectively. Therefore, by analyzing the data items with different evolving time, we can figure out the evolution of recall and precision indirectly. According to the evolution model, the earlier joined interest always has a larger degree. Thus we can deduce the evolving time of interest based on its degree. For example, the No. 8 interest in Table V with largest degree 3,279 is the earliest one joins the network, and consequently the data item related to it has a long evolving time. Using this method, we can figure out the evolution of recall and precision by evaluating data items with different average degrees. Consequently in the following measurements, we deduce the evolving time of network through the average degree of interests in a data item. Our simulation includes three sample traces as given in Table VI. For sample trace 1, we employ six interest sets with average degrees ranging from 394.3 to 2262.7 to represent different evolving times, each of which corresponds to a data item containing 3 interests, i.e., $M = 3$. In addition, sample traces 2 and 3 are generated in a similar way with $M = 4$ and $M = 5$, respectively.

B. Simulation Results

1) Validation of Theoretical Model: We then come to the validation of the applicability of affiliation networks model [4] in the interest-aware information diffusion by comparing the evolution of number of viewers (publishers) in real dataset and that in our theoretical model. Simulation settings are listed in Table IV. Simulation curves are empirical averages over 500 instances. Real dataset curves are obtained from Sample Trace 3. In Fig. 3 and Fig. 4, the results of datasets are represented by the blue dashed lines and that of our theoretical model are represented by the red lines. We can observe that increase rates of the number of viewers and that of publishers are well fitted by our theoretical ones. From the above results we demonstrate that the adopted model can well characterize the interest-aware information diffusion in real evolving networks.

2) Performance of Recall and Precision: As observed in Fig. 5 and Fig. 6, for all the three sample traces, both the metrics recall and precision of publishing based information diffusion mechanism are greatly improved with the evolution of network. In particular, the two metrics both approximately achieve 1 after time 2,000, which exactly verifies our theoretical results in section V. Moreover, we fix a certain sample trace (Sample Trace 3) to see the effect of parameter m on

TABLE V
INTEREST DEGREE

No.	Degree	No.	Degree
1	368	6	1,484
2	2,520	7	549
3	749	8	3,279
4	2,025	9	440
5	956	10	375

TABLE VI
STATISTICAL PROPERTIES OF SAMPLE TRACES

Sample Trace 1 (M = 3)		Sample Trace 2 (M = 4)		Sample Trace 3 (M = 5)	
Avg. Degree	Interests Set	Avg. Degree	Interests Set	Avg. Degree	Interests Set
394.3	{1, 9, 10}	433.0	{5, 3, 7, 9}	487.2	{1, 3, 7, 9, 10}
579.3	{3, 7, 9}	673.5	{5, 3, 7, 9}	627.0	{5, 3, 7, 9, 10}
751.3	{5, 3, 7}	934.5	{6, 5, 3, 7}	834.4	{6, 5, 3, 7, 9}
1419.3	{4, 6, 3}	1303.5	{4, 6, 5, 3}	1167.8	{4, 6, 5, 3, 7}
2009.7	{2, 4, 6}	1746.3	{2, 4, 6, 5}	1559.2	{2, 4, 6, 5, 3}
2262.7	{4, 6, 8}	2327.0	{2, 4, 6, 8}	2074.2	{2, 4, 5, 6, 8}

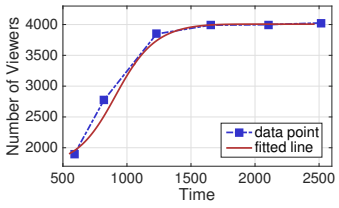


Fig. 3. Evolution of # viewers.

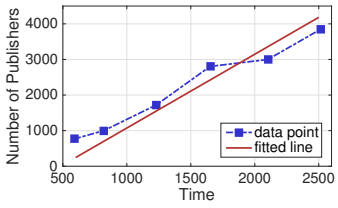


Fig. 4. Evolution of # publishers.

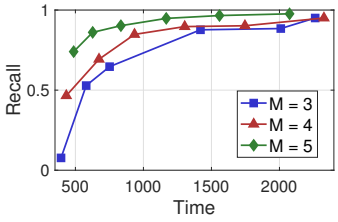


Fig. 5. Evolution of recall.

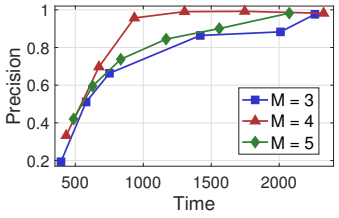


Fig. 6. Evolution of precision.

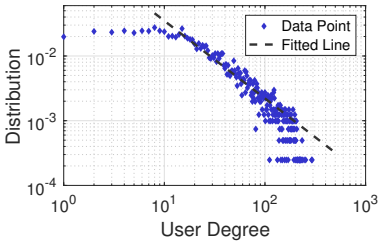


Fig. 7. Distribution of user degree.

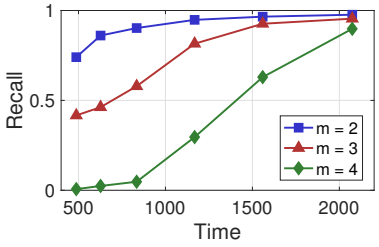


Fig. 8. Recall with varying values of m.

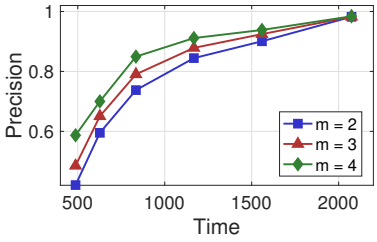


Fig. 9. Precision with varying values of m.

recall and precision. From Fig. 8 and Fig. 9 we can observe that with the increase of parameter m, recall increases and precision decreases. This is because that with a large amount of publishers, viewers are more likely to receive the data but at the same time, redundant users are also more likely to be involved. This result confirms to our initial concern that it is difficult to guarantee good performances of recall and precision simultaneously. However, since these two metrics both perform better with the evolution of network, the delivery accuracy can still achieve RP-Perfect.

VII. CONCLUSION AND FUTURE WORK

In this paper, we study the delivery accuracy of an interest-aware information diffusion in evolving social networks. Firstly, we find that to achieve an RP-Perfect delivery accuracy, the network composed by viewers should be connected, and the number of redundant interests of each publisher should be limited within a certain range. Then, we demonstrate that the number of viewers increases with the evolution of network and the number of publishers increases or remains the same in different cases. Finally, We prove that in a completely evolving network, recall and precision can achieve one simultaneously under some given parameter constraints. Besides the theoretical analysis, we also verify our conclusions through experimental measurements based on a Facebook dataset.

There remains some future directions that can be explored. For example, this work assumes that all the users (interests) in the network grow with a same rate. It is a desirable future work to study the performances of the publishing based information diffusion mechanism with a heterogeneous degree growth rate – a time dependent growth rate that varies among nodes, which provides a better capture of real social networks.

ACKNOWLEDGMENT

This work was supported by NSF China (No. 61532012, 61325012, 61521062, 61602303, 61428205).

REFERENCES

[1] M. Atzmueller, A. Ernst, F. Krebs, C. Scholz, and G. Stumme, "On the Evolution of Social Groups During Coffee Breaks," in *Proc. ACM WWW*, pp. 631-636, 2014.
[2] N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar and D. Song, "Evolution of Social-Attribute Networks: Measurements, Modeling, and Implications using Google+," in *Proc. ACM IMC*, pp. 131-144, 2012.
[3] Twitter: Number of Active Users, http://www.statista.com/, 2016.
[4] S. Lattanzi and D. Sivakumar, "Affiliation Networks," in *Proc. ACM STOC*, pp. 427-434, 2009.
[5] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel and S. Geirhofer, "Network Densification: The Dominant Theme for Wireless Evolution into 5G," in *IEEE Communications Magazine*, vol. 52, no. 2, pp. 82-89, 2014.
[6] R. Baeza-Yates and B. Ribeiro-Neto, "Modern information retrieval," in *New York: ACM Press*, 1999.
[7] R. B. Zadeh, A. Goel, K. Munagala and A. Sharma, "On the Precision of Social and Information Networks," in *Proc. ACM COSN*, pp. 63-74, 2013.
[8] J. Leskovec and A. Krevl, SNAP Datasets: Stanford Large Network Dataset Collection, http://snap.stanford.edu/data, 2014.

[9] Y. Wu, N. Pitipornvivat, J. Zhao, S. Yang, G. Huang and H. Qu, "egoSlider: Visual Analysis of Egocentric Network Evolution," in *IEEE Trans. on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 260-269, 2015.

[10] S. Wu, A. D. Sarma, A. Fabrikant, S. Lattanzi and A. Tomkins, "Arrival and Departure Dynamics in Social Networks," in *Proc. ACM WSDM*, pp. 233-242, 2013.

[11] T. Zhang, P. Cui, C. Faloutsos, W. Zhu and S. Yang, "Come-and-Go Patterns of Group Evolution: A Dynamic Model," in *Proc. ACM KDD*, 2016.

[12] B. Yin, Y. Yang and W. Liu, "Exploring Social Activeness and Dynamic Interest in Community-based Recommender System," in *Proc. ACM WWW*, pp. 771-776, 2014.

[13] F. Chung and L. Lu, "Complex Graphs and Networks," in *Providence: American Mathematical Society*, vol. 107, 2006.

[14] G. Ghoshal, L. Chi and A. Barabàsi, "Uncovering the Role of Elementary Processes in Network Evolution," in *Scientific Reports*, 2013.

[15] O. Yagan, D. Qian, J. Zhang and D. Cochran, "Conjoining Speeds up Information Diffusion in Overlaying Social-Physical Networks," in *IEEE J. on Selected Areas in Communications*, vol. 31, no. 6, pp. 1038-1048, 2013.

[16] L. Fu, W. Huang, X. Gan, F. Yang and X. Wang, "Capacity of Wireless Networks with Social Characteristics," in *IEEE Trans. on Wireless Communications*, vol. 15, no. 2, pp. 1505-1516, 2016.

[17] X. Wang, W. Huang, S. Wang, J. Zhang and C. Hu, "Delay and Capacity Tradeoff Analysis for MotionCast," in *IEEE/ACM Trans. on Networking*, vol. 19, no. 5, pp. 1354-1367, 2011.

[18] X. Wang, L. Fu and C. Hu, "Multicast Performance with Hierarchical Cooperation," in *IEEE/ACM Trans. on Networking*, vol. 20, no. 3, pp. 917-930, 2012.

[19] M. Gomez-Rodriguez, J. Leskovec and A. Krause, "Inferring Networks of Diffusion and Influence," in *ACM Trans. on Knowledge Discovery from Data*, vol. 5, no. 4, 2012.

[20] Y. Jin, J. Ok, Y. Yi and J. Shin, "On the Impact of Global Information on Diffusion of Innovations over Social Networks," in *Proc. IEEE INFOCOM*, pp. 3267-3272, 2013.

[21] Z. Lu, Y. Wen and G. Cao, "Information Diffusion in Mobile Social Networks: The Speed Perspective," in *Proc. IEEE INFOCOM*, pp. 1932-1940, 2014.

[22] J. Liu, L. Fu, J. Zhang, X. Wang and J. Xu, "Modeling Multicast Group in Wireless Social Networks: A Combination of Geographic and Non-geographic Perspective," in *IEEE Trans. on Wireless Communications*, vol. 16, no. 6, pp. 4023-4037, 2017.

[23] C. Jiang, Y. Chen and K. J. Ray Liu, "Evolutionary Dynamics of Information Diffusion Over Social Networks," in *IEEE Trans. on Signal Processing*, vol. 62, no. 17, pp. 4573-4586, 2014.

[24] L. Gao, Y. Xu and X. Wang, "MAP: Multiauctioneer Progressive Auction for Dynamic Spectrum Access," in *IEEE Trans. on Mobile Computing*, vol. 10, no. 8, pp. 1144-1161, 2011.

[25] J. Fan, J. Chen, Y. Du, W. Gao, J. Wu and Y. Sun, "Geocommunity-Based Broadcasting for Data Dissemination in Mobile Social Networks," in *IEEE Trans. on Parallel and Distributed Systems*, vol. 24, no. 4, pp. 734-743, 2013.

[26] Y. Matsubara, Y. Sakurai, B. A. Prakash, L. Li and C. Faloutsos, "Rise and Fall Patterns of Information Diffusion: Model and Implications," in *Proc. ACM KDD*, pp. 33-41, 2012.

[27] D. Wang, H. Park, G. Xie and S. Moon, "A Genealogy of Information Spreading on Microblogs: a Galton-Watson-based Explicative Model," in *Proc. IEEE INFOCOM*, pp. 2391-2399, 2013.

[28] W. Gao and G. Cao, "User-Centric Data Dissemination in Disruption Tolerant Networks," in *Proc. IEEE INFOCOM*, pp. 3119-3127, 2011.

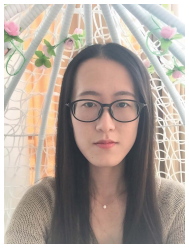
[29] L. Fu, J. Zhang and X. Wang, "Evolution-cast: Temporal Evolution in Wireless Social Networks and its impact on Capacity," in *IEEE Trans. on Parallel and Distributed Systems*, vol. 25, no. 10, pp. 2583-2594, 2014.

[30] W. Hoeffding, "Probability inequalities for sums of bounded random variables," in *J. American statistical association*, pp. 13-30, 1963.

[31] P. Erdős and A. Rényi, "On Random Graphs," in *Publ. Math. Debrecen*, pp. 290-297, 1959.

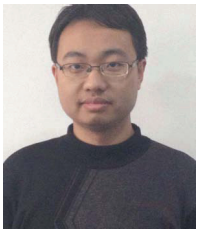
[32] P. Gupta, A. Goel, J. Lin, A. Sharma, D. Wang and R. Zadeh, "WTF: The Who to Follow Service at Twitter," in *Proc. ACM WWW*, pp. 505-514, Rio de Janeiro, Brazil, 2013.

[33] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. T. Sukhavasi, C. Patel and S. Geirhofer "Network densification: the dominant theme for wireless evolution into 5G," in *IEEE Communications Magazine*, vol, 52, no. 2, 2014.



Jiaqi Liu received her B. E. degree in Electronic Engineering from Shanghai Jiao Tong University, China, in 2014. She is currently pursuing the PHD degree in Electronic Engineering in Shanghai Jiao Tong University. Her research of interests are in the area of wireless networks, social networks and evolving networks.

Luoyi Fu received her B. E. degree in Electronic Engineering from Shanghai Jiao Tong University, China, in 2009 and Ph.D. degree in Computer Science and Engineering in the same university in 2015. She is currently an Assistant Professor in Department of Computer Science and Engineering in Shanghai Jiao Tong University. Her research of interests are in the area of social networking and big data, scaling laws analysis in wireless networks, connectivity analysis and random graphs.



Zhe Liu received the B.Eng. degree in telecommunication engineering from Hebei University, Baoding, China, in 2012, and the Ph.D. degree in telecommunication engineering, at Xidian University, Xi'an, China, in 2017. His current research interest is in the area of operating system.

Xiao-Yang Liu received his B.Eng. degree in computer science from Huazhong University of Science and Technology, China, in 2010. He is currently a joint PhD in the Department of Electrical Engineering, Columbia University, and in the Department of Computer Science and Engineer, Shanghai Jiao Tong University. His research interests include tensor theory and deep learning, nonconvex optimization, big data analysis, and applications to Internet of Things.

Xinbing Wang received the B.S. degree (with hons.) in automation from Shanghai Jiao Tong University, Shanghai, China, in 1998, the M.S. degree in computer science and technology from Tsinghua University, Beijing, China, in 2001, and the Ph.D. degree with a major in electrical and computer engineering and minor in mathematics from North Carolina State University, Raleigh, in 2006. Currently, he is a Professor in the Department of Electronic Engineering, and Department of Computer Science, Shanghai Jiao Tong University, Shanghai, China. Dr. Wang has been an Associate Editor for IEEE/ACM TRANSACTIONS ON NETWORKING,IEEE TRANSACTIONS ON MOBILE COMPUTING, and ACM Transactions on Sensor Networks. He has also been the Technical Program Committees of several conferences including ACM MobiCom 2012,2014, ACM MobiHoc 2012-2017, IEEE INFOCOM 2009-2017.