BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 02/06 pp12–24
Volume 2, Number 1, March 2019
DOI: 10.26599/BDMA.2018.9020028

Disseminating Authorized Content via Data Analysis in Opportunistic Social Networks

Chenguang Kong*, Guangchun Luo, Ling Tian, and Xiaojun Cao

Abstract: Authorized content is a type of content that can be generated only by a certain Content Provider (CP). The content copies delivered to a user may bring rewards to the CP if the content is adopted by the user. The overall reward obtained by the CP depends on the user's degree of interest in the content and the user's role in disseminating the content copies. Thus, to maximize the reward, the content provider is motivated to disseminate the authorized content to the most interested users. In this paper, we study how to effectively disseminate the authorized content in Interest-centric Opportunistic Social Networks (IOSNs) such that the reward is maximized. We first derive Social Connection Pattern (SCP) data to handle the challenging opportunistic connections in IOSNs and statistically analyze the interest distribution of the users contacted or connected. The SCP is used to predict the interests of possible contactors and connectors. Then, we propose our SCP-based Dissemination (SCPD) algorithm to calculate the optimum number of content copies to disseminate when two users meet. Our dataset based simulation shows that our SCPD algorithm is effective and efficient to disseminate the authorized content in IOSNs.

Key words: opportunistic social networks; content dissemination; interest centric; data analytics

1 Introduction

An Opportunistic Social Network (OSN) enables users to leverage short-range communication technologies such as Wi-Fi[1], Near Field Communication (NFC)[2] and 5G[3], to form an on-the-fly social network. Through opportunistic contacts among mobile nodes in OSNs, individual users can share free or self-generated content such as news, pictures, and videos. Merchants or organizations can also disseminate their commercial or advertisement content to interested customers through opportunistic communication of social users. For example, coupon brochures may be distributed by the merchant to potential customers in a shopping mall, a conference organizer may send conference invitations to potential interested people, and a bookseller at a book exhibition may deliver free books to interested readers. Such commercial or advertisement content is called authorized content[4, 5], which has interesting features. First, the content is not free of cost. The Content Provider (CP) has to consume a certain amount of resources or services to generate content copies (e.g., goods listed in coupon brochures and book samples). Therefore, only a limited number of content copies are at the CP's disposal. Second, the user who possesses a copy of the content cannot generate copies of the authorized content by himself/herself. Third, each content copy can bring the CP a certain benefit or reward when the content is adopted by a user. For example, a user (or customer) redeems the coupon and purchases goods that can increase the merchant's

• Chenguang Kong and Xiaojun Cao are with the Department of Computer Science, Georgia State University, Atlanta, GA 30309, USA. E-mail: cmkchg@gmail.com; cao@cs.gsu.edu.
• Guangchun Luo and Ling Tian are with the School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China. E-mail: gcluo@uestc.edu.cn; lingtian@uestc.edu.cn.
∗ To whom correspondence should be addressed.
Manuscript received: 2018-04-18; accepted: 2018-05-03

income. Different users can bring different rewards to the content provider which is highly related to how much the users are interested in the content. Thus, the content provider aims to disseminate the limited number of authorized content to the interested users who can maximize the reward received by the CP.

In the OSN, the intermittent network connectivity and contact uncertainty make the content dissemination process unpredictable and difficult. Some studies have investigated the data dissemination in OSNs. The users' interests and preferences are used for content dissemination in OSNs, as shown in Refs. [6–9]. In Refs. [10–14], users with frequent communication or common interests form communities, which are used to find the effective routing for data dissemination. Geographic information is used by several studies to help detect the receivers in OSNs[15]. The work in Ref. [16] develops the geographic location based geo-community and geo-centrality technology to model the regularity of users' mobility in opportunistic social networks. In Refs. [17] and [18], the researchers try to improve the dissemination efficiency and reduce the communication traffic, by leveraging the connection of popular users. The authors in Ref. [17] propose to select a subset of users as the dissemination seeds, according to their online social spreading impacts and offline mobility impacts. Some studies have been conducted on information diffusion services or content delivery by utilizing graph theory[19] and game theory[20] in mobile social networks. However, no previous study has explored the pattern and regulation of users' communication to develop efficient dissemination schemes for authorized content in opportunistic networks.

In this paper, the authorized content dissemination problem in the Interest-centric Opportunistic Social Network (IOSN) is defined for the first time. In the IOSN, users move around motivated by their interests. Therefore, the connection and communication activities of IOSN users are influenced by the distribution of users' interests, and present certain interest related patterns. We propose the Social Connection Pattern (SCP) to analyze the communication and interest distribution data from users in IOSNs. The Social Connection Pattern based Dissemination (SCPD) algorithm is accordingly developed to identify the optimized content dissemination strategies when users meet. The proposed SCPD calculates the maximum expected reward if a certain number of content copies are delivered to the user it has just met. Then the SCPD calculates the number of content copies to deliver such that the overall expected reward is maximized. Our dataset based simulation shows that our SCPD algorithm is effective and efficient to disseminate the authorized content in interest-centric OSNs.

The major contributions of our approach are as follows. First, our model is built to estimate and predict the overall interest property of the opportunistic connections instead of individual user, which can efficiently detect reliable and stable communication patterns. Although an individual user may have arbitrary connections, the interest pattern of the connected users can be effectively captured in our data analytical model. Second, we propose an online data analysis process that efficiently reduce the computation and storage cost of data analysis. Third, by predicting the possible interests of future connectors, our social connection pattern based dissemination algorithm innovatively and efficiently helps users to calculate the number of content copies to be disseminated so that the total reward is maximized.

The rest of this paper is organized as follows. The process of the authorized content dissemination in OSNs is introduced in Section 2. Section 3 describes the definition and formulation of SCP in IOSNs. Based on the data analysis of the social connection pattern, we develop our dissemination algorithm in Section 4. The SCPD algorithm is analyzed in Section 5. We evaluate our algorithm with our dataset based simulation in Section 6 and conclude this work in Section 7.

2 Authorized Content Dissemination in OSNs

In this section, we introduce the authorized content dissemination problem in IOSNs. In an IOSN, a CP generates a limited number of authorized content copies, and disseminates those copies to users during the opportunistic contacts. We assume that the user retains 1 copy for possible use after receiving serveral content copies, no matter whether the user is interested in the content or not. Furthermore, the users are motivated to help disseminate the rest of received content copies to others in future opportunistic contacts because of the incentive mechanism provided by the CP. Many possible incentive mechanisms can motivate

users to help disseminate the content, such as virtual check[8] and Tit-For-Tat (TFT)[21]*. The dissemination process terminates if no further content copies can be disseminated in the network. For example, in Fig. 1, the CP has 4 content copies to disseminate. When D and A meet CP at time t_1, the CP delivers 1 and 3 copies to D and A, respectively. Subsequently (i.e., at time t_2), user A meets with B, user A retains 1 content copy and providers the remaining 2 content copies to user B. After retaining 1 content copy, user B sends the other copy to user C who user B meets at time t_3.

Time-related connection relationships between users are of two types: **contactor** and **connector**. User u's contactors are the users who have directly contacted with user u (e.g., A and D are contactors of CP since time t_1 as shown in Fig. 1). User u's connectors are the users who have not directly contacted with user u while a connection path exists from those users to u (e.g., B and C are connectors of CP at time t_3). The connectors can be further identified according to the length of the communication path between users (i.e., the number of **hops**). For example, in Fig. 1, user C is a 3-hop connector of CP since C is connected CP through the communication between CP-A, A-B, and B-C. Similarly, we can also consider the contactors of a user as the 1-hop connectors of the user.

Each user has different interests in the content, which has a significant influence on the possible reward obtained by the CP. Generally, the more likely a user adopts or redeems a content copy, the more reward the content provider will obtain. This reward is proportional to the user's interest in the content. Hence, the reward

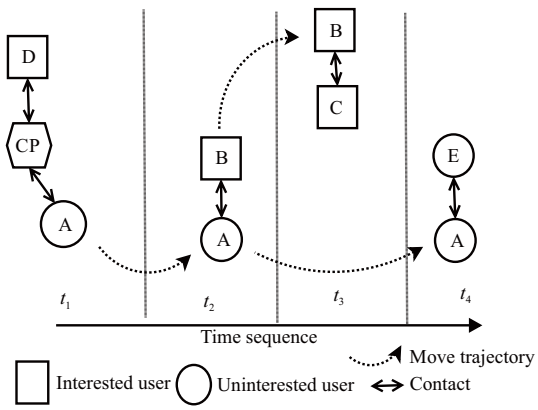


Fig. 1 An example of authorized content dissemination in OSNs.

*The work on incentive mechanism is beyond the scope of this paper.

amount can be measured as the overall interest of the content receivers. To maximize the overall reward from disseminating the content copies, the CP should consider the interests of both possible contactors and connectors. In Fig. 1, D is a contactor of CP who is interested in the content, so CP prefers to deliver a copy to D. On the other hand, user A is not highly interested in the content but is able to help disseminate the content to other interested connectors in the future (e.g., B and C). Thus, a good strategy for the CP is to deliver 1 copy to D, and 3 content copies to A during the contact. User A will retain 1 copy for itself and deliver the remaining 2 to B and C. User E will not receive any content because it is neither interested in the content nor able to contact with other interested users (that have extra content copies).

When a content holder who has received a certain number of content from the content provider or other users meets a new contactor, it needs to make its dissemination decision to help maximize the reward generated by the content copies it holds. Hence, each user who holds a number of content copies has to figure out the following two questions when it meets a new contactor.

(1) What type of users can the new contactor probably encounter directly or connect remotely in the future?

(2) How many content copies should be delivered to the new contactor so that the overall reward is maximized?

To help the user answer these two questions and make the dissemination decision, we propose our SCP and SCPD algorithm in the following sections.

3 Interest-Centric Opportunistic Social Networks

In this section, based on the user mobile behaviors, SCP is proposed to efficiently predict the possible connection pattern of mobile users in IOSNs.

We select two typical IOSN datasets as the study cases for user behavior investigation: Infocom 2006[22] and Sigcomm 2009[23]. The information on these two datasets is listed in Table 1.

Table 1 Information on the OSN datasets.

Name	Nodes	Duration	Contacts	Topics
Infocom 2006	98	4 days	227 657	27
Sigcomm 2009	76	4 days	285 879	154

3.1 User's contact in IOSN

Figure 2 shows an example of a user's contact history in an active period according to the Infocom 2006 dataset. X-axis is the time stamp while y-axis is the ID of the user encountered. The preceding figures show an unbalanced distribution of the contacts. The user in this IOSN frequently encounters a certain number of users (such as users with ID 14 and 37) while rarely contact with some others. The frequent encountering lasts in the whole active period instead of a short time duration. This indicates that users have a high and stable probability to encounter certain type of users. The similar observation of Sigcomm 2009 is shown in Fig. 3.

For these two datasets, we assign a **weight** in the range of $[0, 50]$ to each user to indicate how the user is interested in a specific content. The values of the weight are obtained by matching the topics of content and users' interested topics.

Figures 4 and 5 show the weight distribution of our datasets. In these figures, the y-axis is the weight of contactors at different times. Compared with the distribution of the contact ID, the distribution of the contacts' weights are concentrated in a smaller scale (such as weights 15 and 44). The weight distribution is stable through the entire active period. The weight is distributed in multiple discrete scales, instead of single scale, because each content/location holds different properties that attract various types of users. For instance, in a conference section "online social networks", the attendees may be interested in either "networks" or "big data".

3.2 Social connection pattern in IOSN

The SCP of a user u_i consists of two elements: a social connection pattern matrix $\boldsymbol{P}_{it}$ and a counting vector $\boldsymbol{C}_{it}$. $\boldsymbol{P}_{it}$ records the interest distribution of the

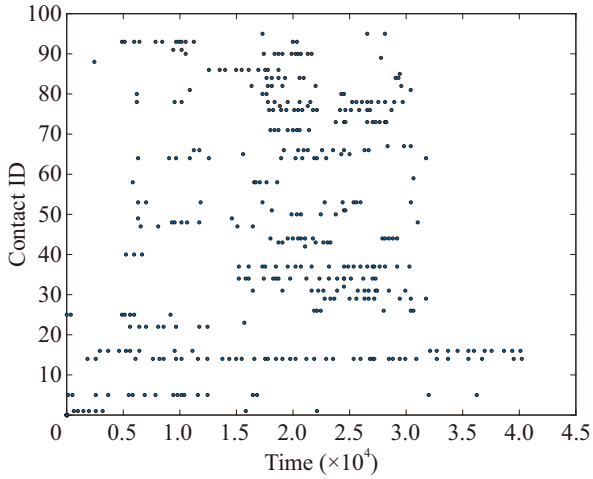


Fig. 2 An example of a user's contact history in Infocom 2006 dataset.

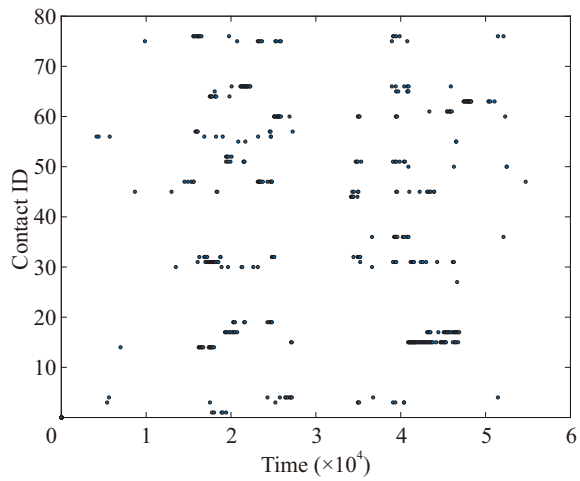


Fig. 3 An example of a user's contact history in Sigcomm 2009 dataset.

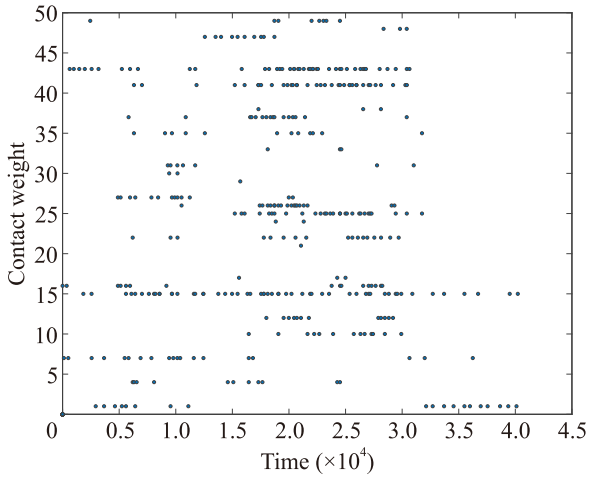


Fig. 4 An example of the weight of a user's contact in Infocom 2006 dataset.

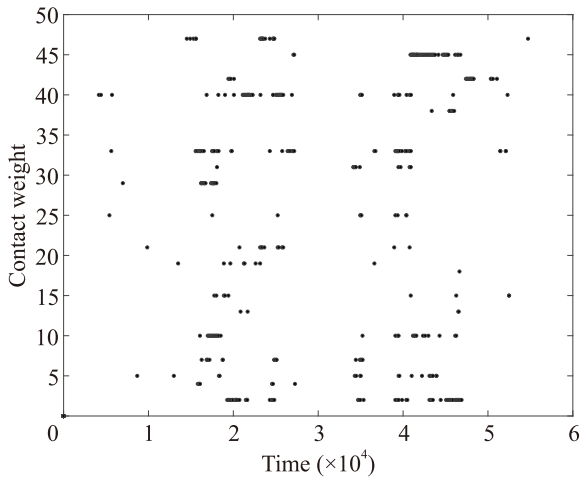


Fig. 5 An example of the weight of a user's contact in Sigcomm 2009 dataset.

users who have directly contacted or connected through intermediate users with user i until time t. The counting vector $\boldsymbol{C}_{it}$ counts the number of connectors of user i until time t. The j-th row in $\boldsymbol{P}_{it}$ is used to record the interest distribution of the j-hop connectors of user i at time t as shown in Eq. (1), where N is the largest hop number[†].

In $\boldsymbol{P}_{it}$, a vector $\boldsymbol{x}_{ijt} = [p_{ij1t}, ..., p_{ijwt}, ..., p_{ijWt}]$ is used to describe the interest distribution of j-hop connectors of user i at time t. p_{ijwt} is the probability that a j-hop connector of user i at time t has a weight of w. W denotes the largest possible weight.

$$\boldsymbol{P}_{it} = [\boldsymbol{P}_{i1t}, \boldsymbol{P}_{i2t}, ..., \boldsymbol{P}_{iNt}]^{\mathrm{T}} = \begin{bmatrix} p_{i11t} & p_{i12t} & \cdots & p_{i1Wt} \\ p_{i21t} & p_{i22t} & \cdots & p_{i2Wt} \\ \vdots & \vdots & \vdots & \vdots \\ p_{iN1t} & p_{iN2t} & \cdots & p_{iNWt} \end{bmatrix} \tag{1}$$

According to the "small world" property of opportunistic social networks discussed in Refs. [24–26], the opportunistic social networks have a high average clustering coefficient and a low average path length. Thus, in most scenarios, N in Eq. (1) can be set as a rationally small number to reduce the storage cost and computation cost while ensuring the social connection pattern matrix covering most connectable users.

The counting vector $\boldsymbol{C}_{it}$ counts the number of connectors of user i at time t, which describes the potential that user i will meet with others as shown in Eq. (2).

$$\boldsymbol{C}_{it} = [c_{i1t}, c_{i2t}, ..., c_{ijt}, ..., c_{iNt}] \tag{2}$$

where c_{i1t} is obtained by counting the direct contactors of user i, and c_{ijt} is the count of j hop connectors.

3.3 SCP probability learning

To leverage SCP for content dissemination, we need to learn the SCP probabilities first. The SCP probabilities are learned by capturing and analyzing the weight information of possible connectors. When a new user u with weight w_u joins the OSN (e.g., a new customer enters a shopping mall), the social connection pattern matrix and counting vector are initialized by setting the weight distribution of connectors at each hop uniformly and setting the counting vector to zero. When this user u encounters another user v with weight w_v at time t, they will exchange their social connection pattern matrix and counting vector ($\boldsymbol{P}_{ut}$, $\boldsymbol{C}_{ut}$, $\boldsymbol{P}_{vt}$, $\boldsymbol{C}_{vt}$). Then the connection pattern matrix $\boldsymbol{P}_{ut}$ and counting vector $\boldsymbol{C}_{ut}$ are learned and updated.

$\boldsymbol{P}_{u1t}$ is learned by tracing the weights of the 1-hop connectors when encountering them.

$$p'_{u1wt} = \begin{cases} \frac{p_{u1wt} \cdot c_{u1t} + 1}{c_{u1t} + 1}, & \text{if } w_v = w; \\ \frac{p_{u1wt} \cdot c_{u1t}}{c_{u1t} + 1}, & \text{otherwise} \end{cases} \tag{3}$$

$$c'_{u1t} = c_{u1t} + 1 \tag{4}$$

In Eqs. (3) and (4), p'_{u1wt} is the updated interest distribution of user u's contactors and c'_{u1t} is the updated number of contactors of user u at time t.

When user u meets user v, the information of $\boldsymbol{P}_{v(j-1)t}$ can be used to update the value of $\boldsymbol{P}_{ujt}$. Assume n_{ujt} is a possible j-hop connector of user u, the SCP probability distribution of more than 1 hop connector P_{ujt} ($1 < j \leqslant N$) can be formulated as

$$\boldsymbol{P}_{ujt}(w) = \sum_{w'=1}^{W} p(n_{ujt} = w | n_{u1t} = w') \cdot p(n_{u1t} = w') = \sum_{w'=1}^{W} p_{n_{u1t}(j-1)t}(w) \cdot \boldsymbol{P}(n_{u1t} = w') \tag{5}$$

At time t, suppose a j-hop connector n_{ujt} of u is recorded connecting to u through a 1-hop connector n_{u1t} of u. In Eq. (5), $p(n_{ujt} = w | n_{u1t} = w')$ is the conditional probability that n_{ujt} has weight w if n_{u1t} has weight w'. Hence, we have $p(n_{ujt} = w | n_{u1t} = w') = p_{n_{u1t}(j-1)t}(w)$ if user n_{u1t} has weight w'. Therefore, we can train $\boldsymbol{P}_{ujt}(w)$ from $\boldsymbol{P}_{n_{u1t}(j-1)t}(w)$ of user u's 1-hop connector n_{u1t}.

With the difference of connectivity of users (reflected by counting vector $\boldsymbol{C}_{vt}$), the updating approach can be described as follows:

$$p'_{ujwt} = \frac{1}{c'_{u1t}} p_{v(j-1)wt} \cdot c_{v(j-1)t} + \left(1 - \frac{1}{c'_{u1t}}\right) p_{ujwt} \cdot c_{ujt} = \frac{c_{v(j-1)t} p_{v(j-1)wt} + (c'_{u1t} - 1) c_{ujt} p_{ujwt}}{c_{v(j-1)t} + (c'_{u1t} - 1) c_{ujt}} \tag{6}$$

After the social connection pattern matrix is updated, the counting vector is updated as follows:

$$c'_{ujt} = \left(1 - \frac{1}{c'_{u1t}}\right) c_{ujt} + \frac{1}{c'_{u1t}} c_{v(j-1)t} \tag{7}$$

In Eqs. (6) and (7), p'_{ujwt} is the updated interest distribution of user u's j-hop connector and c'_{ujt} is the updated number of j-hop connectors of user u at time

[†] N and the matrix can be simplified when removing the isolated users from the network.

t. $\frac{1}{c'_{u1t}}$ represents the ratio of user v across all samples. Similarly, user v updates its $\boldsymbol{P}_{vt}$ and $\boldsymbol{C}_{vt}$.

4 SCP-Based Content Dissemination

In this section, we introduce the SCP-based content Dissemination (SCPD) algorithm to disseminate the authorized content in IOSNs. We assume an existence of M content copies generated by a CP. When the dissemination process ends, all users in the network have at most 1 content copy. Otherwise, they will continue to disseminate the additional copies to others. The objective is to disseminate those M copies to the users in IOSNs so that the total reward of the users who receive the content copies is maximized.

Suppose user u holds $s+1$ ($M > s \geqslant 1$) copies of the content and needs to disseminate s content copies in the dissemination process starting from u. The s content copies are expected to generate some reward through the dissemination process, denoted as $g_u(s)$. When user u encounters another user v at time t, user u needs to calculate how many content copies should be delivered to v. Assume user u delivers m content copies to user v, we define $f_v(m)$ as the expected reward to be generated by the m content copies through user v. As shown in Fig. 6, the total expected reward of s content copies is $g_u(s)$ before user u meets user v. After user u disseminates m content copies to user v, the total expected reward of the s content copies now comes from the sum of the $s-m$ copies held by u and m copies held by v, which can be denoted by $g_u(s-m)+f_v(m)$. Clearly, user u prefers to disseminate m content copies to user v if the total expected reward of the s content copies increases after the dissemination (i.e., $g_u(s-m)+f_v(m) > g_u(s)$). The key idea of our SCPD algorithm is to find out the optimal m, denoted by m^* such that the increase of total expected reward is maximized as follows:

$$m^* = \text{argmax}_m\{(g_u(s-m)+f_v(m)) - g_u(s)\}, \quad \text{s.t. } 0 \leqslant m \leqslant s \tag{8}$$

where m^* is the optimum number of content copies to be delivered to v by u.

In Eq. (8), the expected reward of the m content copies disseminated to user v (i.e., $f_v(m)$) is calculated based on the SCP of user v, which describes the possible interest of v's potential connectors. The value of $f_v(m)$ can be calculated as

$$f_v(m) = \max_{m_j} w_v + \sum_{j=1}^{N}\sum_{w=1}^{W} p_{vjwt} \times m_j \tag{9}$$

$$\text{s.t. } \sum_{j=1}^{N} m_j \leqslant m-1 \tag{10}$$

$$\frac{m_j}{m_{j-1}} \leqslant \frac{c_{vjt}}{c_{v(j-1)t}}, \text{ for all } 1 < j \leqslant N \tag{11}$$

$$\forall m_j \in \mathbb{Z} \tag{12}$$

where w_v is the interest weight of user v; p_{vjwt} is the element in the social connection pattern matrix P_{vt} and m_j is the number of content copies retained by the j-hop connectors of user v in the dissemination process starting from v. The reward generated by the j-hop connector is $\sum_{w=1}^{W} p_{vjwt} \times m_j$. Thus, Eq. (9) shows that the total reward created by the m content copies is $w_v + \sum_{j=1}^{N}\sum_{w=1}^{W} p_{vjwt} \times m_j$. Recall that v retains 1 copy for self-usage if it receives any content as shown in Eq. (10). The constraint in Eq. (11) ensures that a sufficient number of intermediate users can connect to the j-hop connectors. According to the counting vector $\boldsymbol{C}_{vt}$, the total number of c_{vjt} users in j-hop are connected to user v through $c_{v(j-1)t}$ users in $(j-1)$-hop. Hence, to deliver m_j content to the users in the j-hop, we need the collaboration of at least $\frac{c_{v(j-1)t} m_j}{c_{vjt}}$ users in the $(j-1)$-hop as shown in Eq. (11).

Similarly, user u has its expectation on the reward from the content copies it holds. If user u holds s content copies, the expected reward $g_u(s)$ is calculated as in Eq. (13).

$$\begin{aligned} g_u(s) &= \max_{s_j} w_u \sum_{j=1}^{N}\sum_{w=1}^{W} p_{ujwt} \times s_j, \\ &\text{s.t. } \sum_{j=1}^{N} s_j \leqslant s, \\ &\frac{s_j}{s_{j-1}} \leqslant \frac{c_{ujt}}{c_{u(j-1)t}}, \text{for all } 1 < j \leqslant N, \\ &\forall s_j \in \mathbb{Z} \end{aligned} \tag{13}$$

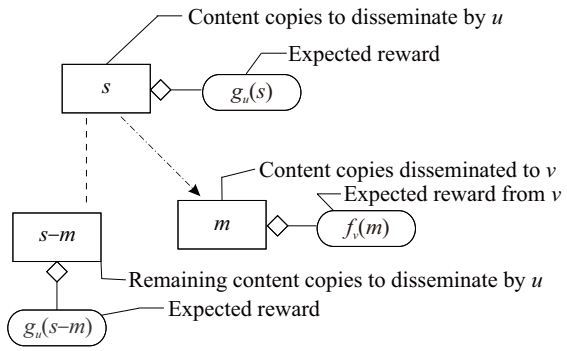


Fig. 6 Expected reward change during the content dissemination process.

To efficiently resolve this maximization problem in Eqs. (9) and (13), we propose the Reward Maximization Algorithm (RMA), a Branch and Bound algorithm[27] as shown in Algorithm 1. The steps of our algorithm are described as follows.

(1) **Initialization**. We set $i = 1$ and the value of m_i as 1 to $m - 1$ to generate *m possible solutions s* as shown in Lines 5–7 in Algorithm 1. The $f_v(m)$ of each possible solution is calculated by ignoring the constraint in Eq. (11). The size of the possible solutions $|s|$ is the value of m_i.

(2) **Bound**. As shown in Lines 8–11, we check if any possible solution satisfies the constraint in Eq. (11) as well. If yes, select the possible solution with the largest $f_v(m)$ that satisfies the constraint in Eq. (11) as the *available solution*, and remove any other possible solutions with smaller $f_v(m)$. If the *available solution* is the only possible solution left, the algorithm terminates; otherwise, proceed to step (3).

(3) **Branch**. We set the value of m_{i+1} as 1 to $m - |s| - 1$ to generate new possible solutions, as shown in Lines 14–26. The $f_v(m)$ of each new possible solution is calculated by ignoring the constraint in Eq. (11). The size of the new possible solutions $|s'|$ is $|s| + m_{i+1}$. Update $i = i + 1$ and go to step (2).

With the RMA algorithm calculating $f_v(m)$ and $g_u(s)$, users u can identify the optimal m^* in Eq. (8) by traversing all m in $0 \leqslant m \leqslant s$. Based on the calculation of m^*, we have our SCPD scheme described in Algorithm 2.

Algorithm 1 Reward Maximization Algorithm

1: **Input**: Number of content copies m
Social Connection Pattern matrix of user v $\boldsymbol{P}_v$
Counting vector of user v $\boldsymbol{C}_v$
2: **Output**: Maximum reward obtained if user v disseminates the m content copies, denoted by $f_v(m)$
3: Initialize the possible solution set $L = null$ and *optimal reward solution* $F = 0$
4: **for** each $i \in [1, m]$ **do**
5: Set a new *possible solution* s as $s = \{m_1, ..., m_N\}$, where $m_1 = i$ and $m_j = 0\ \forall j \neq 1$
6: Calculate the *maximum reward* of s by ignoring the constraint in Eq. (11)
7: Add s to L
8: **if** {s is an *available solution* satisfying all constraints} AND {the maximum reward of s > *optimal reward solution* F} **then**
9: Set F =maximum reward of s
10: Remove any possible solution in L with maximum reward smaller than F
11: **end if**
12: **end for**
13: **while** L contains other *possible solutions* besides the *available solution* **do**
14: Select the *possible solution* s' with the largest maximum reward from L
15: Find the smallest j with $m_j = 0$ in s'
16: Calculate the total number of content copies assigned n in s'
17: **for** $i \in [1, m - n]$ **do**
18: Set a new *possible solution* s'' with $s'' = s'$.
19: Set the j-hop value m_j in s'' as $m_j = i$.
20: Calculate the maximum reward of s''
21: **if** {s is an *available solution* } AND {the maximum reward of $s > F$} **then**
22: Set F =maximum reward of s''
23: **end if**
24: Remove any possible solution in L with maximum reward smaller than F
25: **end for**
26: Remove s'
27: **end while**
28: **return** F

5 SCPD Analysis

Our dissemination model is developed on the basis of the observation of IOSNs, in which most mobile social network users move around the network according to personal interests. In IOSNs, mobile social network users with similar personal interests have similar

Algorithm 2 SCP based dissemination scheme

1: User u inputs its interests for content-interest matching when joining the network
2: User u initializes its SCP
3: **while** TRUE **do**
4: **if** User u meets with user v **then**
5: u and v exchange personal interests, caching content information, and SCP
6: **for** content c in {content set that v holds > 1 copies AND u does not hold} **do**
7: v calculates the optimum number of content copies m^* to deliver to u by using RMA algorithm
8: v transmits m^* copies of content c to u
9: **end for**
10: **for** content c' in { content set that u holds > 1 copies AND v does not hold} **do**
11: u calculates the optimum number of content copies m'^* to deliver to v by using RMA algorithm
12: u transmits m'^* copies of content c' to v
13: **end for**
14: u and v update their SCP values
15: **end if**
16: **end while**

movement trajectories and pattern. Therefore, IOSN users possess a stable distribution of the encounters.

5.1 Stability

The accuracy of the SCP based dissemination scheme is influenced by the stability of the SCP, i.e., how likely that mobile social network users with similar interests have similar movement trajectories. In a typical IOSN, users move around the locations that match their personal interests, therefore, users with similar interests would have similar movement patterns. As a result, IOSN users would have a stable SCP, which is related to personal interests. This condition is demonstrated by Infocom 2009 and Sigcomm 2006 datasets in previous sections.

The SCP may shift during the dissemination lifetime. This shift may be unnoticeable and may have a limited impact on the dissemination process because of the short lifetime of the process. Time is an important factor that causes the shifting of the SCP. Many locations provide varying information at different times, thereby attracting users with various interests. For instance, the conference rooms in a conference may host discussions of different topics. The attenders to those locations would have different interests according to the topics at a particular time. To handle the shifting caused by time, we can set a time decoy when we study the SCP. Over time, the older encounters would have less impact on SCP, whereas the impact of new encounters is increased. Hence, the SCP evolution equation, Eqs. (3) and (6), can be replaced by

$$p'_{ujwt} = \begin{cases} \frac{\alpha \cdot p_{ujwt} \cdot c_{ujt} + 1}{c_{ujt} + 1}, & \text{if } w_v = w; \\ \frac{\alpha \cdot p_{ujwt} \cdot c_{ujt}}{c_{ujt} + 1}, & \text{otherwise} \end{cases} \tag{14}$$

$$p'_{ujwt} = \frac{c_{v(j-1)t}\, p_{v(j-1)wt} + \alpha(c'_{u1t} - 1) c_{ujt}\, p_{ujwt}}{c_{v(j-1)t} + \alpha(c'_{u1t} - 1) c_{ujt}} \tag{15}$$

in which α ($0 < \alpha \leqslant 1$) is a time decoying coefficient that describes the influence of old records on current social connection pattern.

5.2 Population

The population of the IOSN users in the network actually plays an important role in the dissemination accuracy. When few IOSN users exist, the information is insufficient to derive a stable SCP. In that case, the individual ISON user will have a higher influence on the SCP of their encounters, which may degrade the performance of the SCP based dissemination. To address the problem caused by a small population in the network, instead of initializing the SCP distribution as 0, we can initialize the SCP distribution in Eq. (1) as a Gaussian distribution centralized on the interests of the user (e.g., $p_{ijwt} \sim N(w_0, 1)$), where w_0 is the weight of user u_i. The reason for such initialization is that IOSN users have a high probability to meet others with similar interests in IOSNs.

5.3 Scalability

In the SCP based dissemination scheme, estimating the interest distribution of all accessible users is important (i.e., the users can be communicate and disseminate content through direct contactor multiple hop forwarding). To ensure an accurate estimation of the expected dissemination reward, the SCP should cover most users who may be involved in the dissemination process. However, capturing the information of the users with a very large hop distance may be difficult. First, the increase in hop numbers will increase the computation cost for every user when it calculates the optimum number of copies to deliver. Second, the stability and accuracy of the SCP is reduced on the users with large hop distance because of the influence of the intermediate users. Third, the network can normally be traversed by a small hop number. A large hop number in SCP will repeatedly count only a segment of the users. Hence, we should consider the scalability of the network to choose the appropriate hop number N in SCP. If the network is large or spare, a larger N should be set to avoid missing any important users in the dissemination. On the other hand, a smaller N is more suitable for a network with small scale or dense connections.

6 Simulation and Evaluation

In this section, we present the numerical results and performance analysis from our dataset based simulation.

Two typical opportunistic social network datasets are used in our simulation: Infocom 2006[22] and Sigcomm 2009[23]. In our simulation, the reward received by the content provider is represented by the weight sum of the content receivers who retain 1 content copy after the termination of the dissemination process. We use the topics gathered from the participants to calculate the interest weight of a social user. The weights of users are normalized within the range of $[0, 100]$. If topic information is missed in the datasets, we randomly

generate the topic when needed.

When deploying our SCPD algorithm, we use a number of contact records as the training pool to formulate the social connection pattern for each user. Two different lengths of the training process are adopted: 10 000 and 30 000 contacts records. To evaluate the efficiency of our algorithm, we compare it with the Flooding algorithm[28]. In the Flooding algorithm, when a user C meets a new contactor B and B has not obtained the content previously, C delivers half of C's content copies to B regardless of the interest of B.

Figure 7 shows the dissemination results of the Infocom 2006 dataset. In the simulation, we randomly select a user as the content provider, who intends to disseminate M ($10 \leqslant M \leqslant 90$ as shown in the x-axis) content copies to users in the network. The y-axis denotes the total reward obtained from the content receivers who hold 1 content copy when the dissemination process terminates. The results show that the total reward obtained by the SCPD algorithm is significantly higher than that obtained by the Flooding algorithm. In fact, the proposed SCPD can outperform the Flooding scheme by as much as 40%. This indicates that SCPD can efficiently disseminate the content to users with higher interest in the content. It also demonstrates that the usage of social connection pattern is able to help predict users' possible connections in the future, and help effectively determine the number of content to deliver when two users meet.

In Fig. 7, the SCPD algorithm achieves a similar total reward as the flooding scheme when $M = 10$ or $M = 90$ while outperforming the flooding scheme the most when $M = 50$. This can be explained as follows. When $M = 50$, half of the users in the network will receive the content after the dissemination process. Hence how those users are selected to receive the content is extremely important. The SCPD can take advantage of the SCP and user's interest information to maximize the reward achieved. However, when M is small, most users in the network cannot receive the content. Even though the proposed SCPD can identify the users with high interest in the content, the interested users may be too far from the CP (in terms of connection hops). Thus, with an extremely small number of content copies to disseminate, both algorithms likely deliver them to users who are closer to the CP (i.e., users who are fewer hops away from the CP). In this case, the social connection pattern and user interests may not matter significantly. Similarly, when the number of content copies is large (e.g., $M = 90$), most of the social users receive the content whereas the SCP and user interests have a minimal impact on the performance of the dissemination process. Thus the performance difference between SCPD and the Flooding algorithm is small when M is very small or very large. Similar conclusions are oberved from the results of the Sigcomm 2009 dataset, as shown in Fig. 8.

We also evaluate the performance of our SCPD algorithm with different training lengths. By testing the results with different training lengths, we can evaluate whether each user actually formulates a stable SCP. The total reward results on the different training lengths of the Infocom 2006 and Sigcomm 2009 datasets are shown in Figs. 9 and 10, respectively. In the simulation, we set two different training lengths: 10 000 contact records and 30 000 contact records. The results show that the total reward of the content receivers with different training lengths is quite close to each other for both the Infocom 2006 and Sigcomm 2009

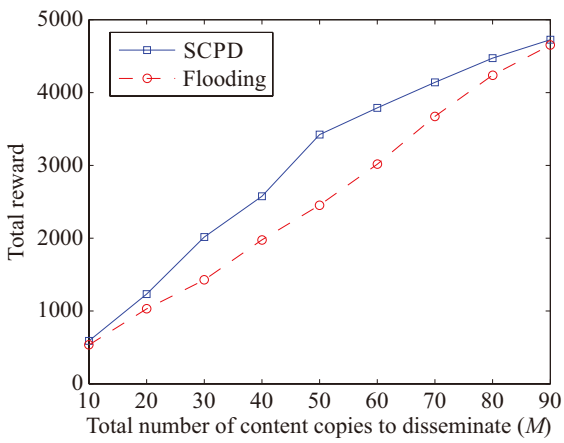


Fig. 7 Dissemination results of Infocom 2006 dataset.

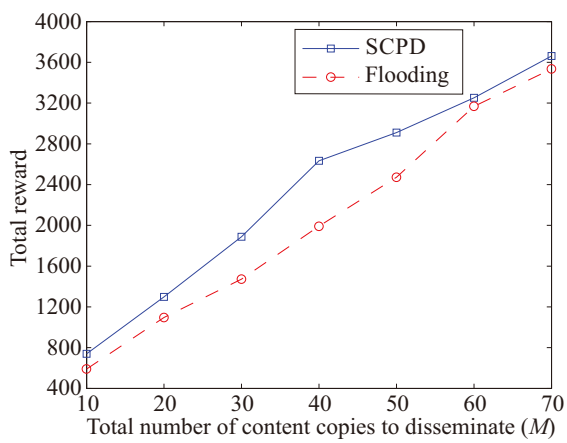


Fig. 8 Dissemination results of Sigcomm 2009 dataset.

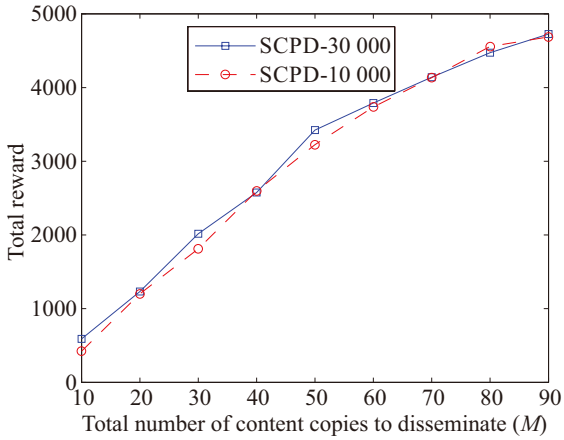


Fig. 9 Dissemination results of Infocom 2006 dataset with different training lengths.

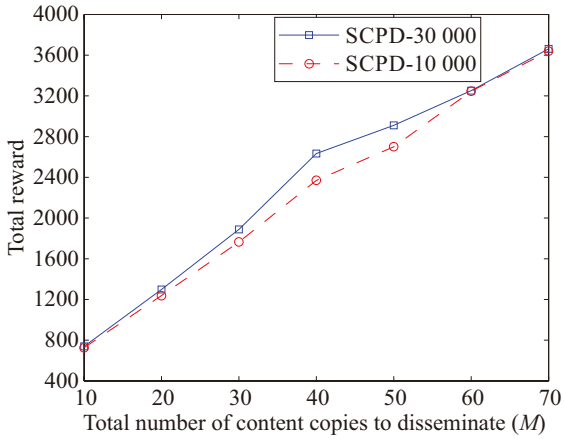


Fig. 10 Dissemination results of Sigcomm 2009 dataset with different training lengths.

datasets. This indicates that the SCP of each user after 20 000 additional record training is similar to the pattern formulated in the previous 10 000 contact records. Thus, the proposed SCPD can reliably predict the future connections according to the SCP.

Figure 11 shows the distribution of the content copies disseminated to each user. The x-axile is the user ID, which is sorted by the number of content copies received by the user. The y-axile is the number of content copies received by the user. We compare the dissemination results with different total content copies (i.e., $M = 30, 50, 70$), from the Flooding algorithm and our SCPD algorithm. In Fig. 11, the total number of users receiving 1 or more content copies is M as only M users in the network participate in the dissemination process. Figure 11 shows that several critical forwarders have received a large number of content copies (typically $> \frac{1}{2}M$), and forward them to other users.

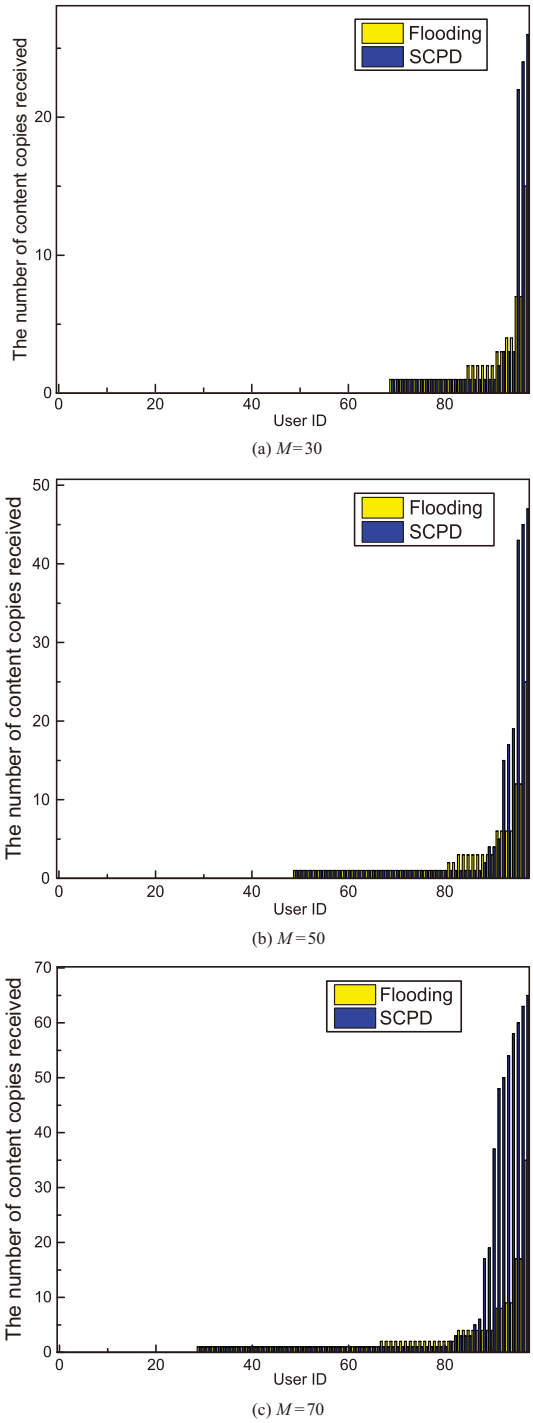


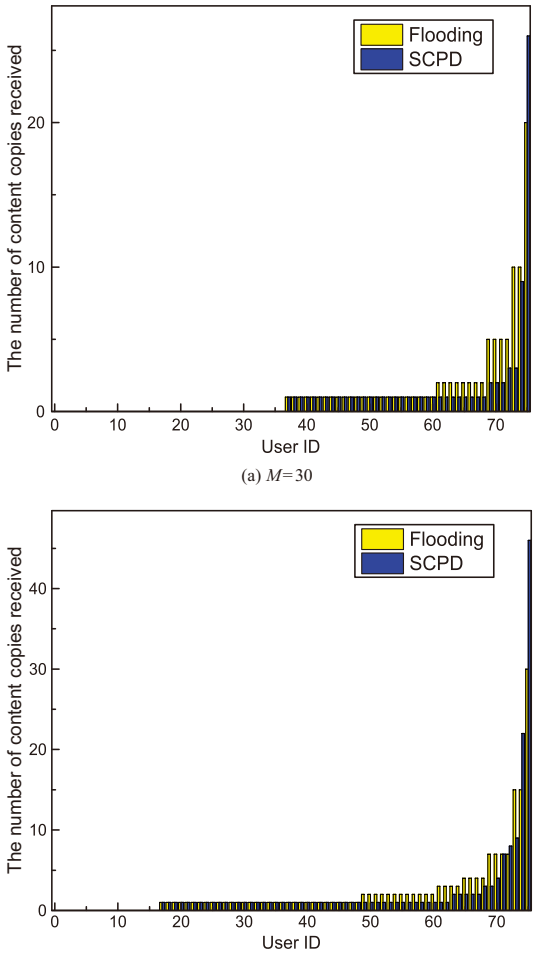
Fig. 11 Distributions of received content copies for Infocom 2006 dataset.

Figure 11 also indicates that more users receive a large number of content copies in the SCPD dissemination algorithm than the Flooding dissemination algorithm. The largest number of content copies delivered to users is much larger in SCPD (65 as in Fig. 11c) than that in the Flooding algorithm (35 as in Fig. 11c). This indicates that the SCPD algorithm is

able to identify the critical forwarders (e.g., the users who receive a large number of content copies) who can efficiently help disseminate the content to the interested users.

In the SCPD algorithm, besides the users with a large number of content copies (the critical forwarders) and the users with 0 content copy received, most of the other users receive 1 content copy. It implies that most of the users receiving 1 content copy obtain the copy directly from the critical forwarders that have a large number of content copies. However, more users in Flooding algorithm receive more than 1 copy. Those users act as intermediate forwarders to disseminate the content. Hence, the average dissemination path length between the receivers and the content provider in SCPD is smaller than that in Flooding algorithm. The reason is that the SCPD algorithm can efficiently detect the users who likely contact users offering high reward. Since the users with high reward have similar interests, they are able to contact some common users with high probability. Hence, the content provider can disseminate numerous content copies to the common contacted users who can then forward the content to the interested users with short dissemination paths. When more content copies are disseminated, as shown in Figs. 11b and 11c, more users act as intermediate forwarders who have received fewer content copies than the critical forwarders. When M is large, we need to disseminate the content to users who may not be contacted by the critical forwarders. Hence, we need to rely on other users as the intermediate forwarders to disseminate the content. As a result, the number of users receiving more than 1 content copies increases.

The evaluation results based on Sigcomm 2009 dataset are shown in Fig. 12. Given the total number of content copies as $M = 40$ and $M = 60$, we calculate the distribution of the content copy number received by users, as shown in Figs. 12a and 12b, respectively. Figure 12 shows a similar distribution as that in Fig. 11, which further demonstrates the effectiveness and efficiency of our SCPD algorithm.



(a) M=30

(b) M=60

Fig. 12 Distributions of received content copies for Sigcomm 2009 dataset.

7 Conclusion

In IOSNs, users move around the interested locations and have a stable probability to meet users with certain interests, thereby allowing the prediction of the potential social connections. In this paper, we have formulated the problem of authorized content dissemination in IOSNs. We have derived the SCP to statistically analyze the interest distribution of both direct contactors and indirectly connectors. We have also proposed the SCPD algorithm to effectively disseminate the authorized content via effective data analysis in the IOSN while maximizing the reward generated by all content receivers. Our dataset based simulation has shown that our SCPD algorithm outperforms the existing Flooding algorithm by as much as 40%.

References

[1] Wi-Fi Alliance, Official industry association, http://www.wi-fi.org/, 2011.

[2] NFC, NFC forum, http://nfc-forum.org/, 2016.

[3] Z. Su and Q. C. Xu, Content distribution over content centric mobile social networks in 5G, *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 66–72, 2015.

[4] C. G. Kong and X. J. Cao, Semi-controlled authorized information dissemination in content-based social networks, in *Proc. 23rd Int. Conf. Computer Communication*

and Networks, Shanghai, China, 2014.

[5] C. G. Kong and X. J. Cao, Disseminating authorized content in interest-centric opportunistic social networks, in *Proc. 24th Int. Conf. Computer Communication and Networks*, Las Vegas, NV, USA, 2015.

[6] K. Lin, C. Chen, and C. Chou, Preference-aware content dissemination in opportunistic mobile social networks, in *Proc. 2012 IEEE INFOCOM*, Orlando, FL, USA, 2012.

[7] R. I. Ciobanu, R. C. Marin, C. Dobre, V. Cristea, and C. X. Mavromoustakis, ONSIDE: Sociallyaware and interestbased dissemination in opportunistic networks, in *Proc. 2014 IEEE Network Operations and Management Symp.* (*NOMS*), Krakow, Poland, 2014.

[8] T. Ning, Z. P. Yang, H. Y. Wu, and Z. Han, Self-interest-driven incentives for ad dissemination in autonomous mobile social networks, in *Proc. 2013 IEEE INFOCOM*, Turin, Italy, 2013.

[9] Y. M. Zhao, W. Song, and Z. Han, Social-aware data dissemination via device-to-device communications: Fusing social and mobile networks with incentive constraints, *IEEE Trans. Serv. Comput.*, doi:10.1109/TSC.2016.2599160.

[10] B. Jedari and F. Xia, A survey on routing and data dissemination in opportunistic mobile social networks, arXiv preprint arXiv: 1311.0347, 2013.

[11] M. J. Xiao, J. Wu, and L. S. Huang, Community-aware opportunistic routing in mobile social networks, *IEEE Trans. Comput.*, vol. 63, no. 7, pp. 1682–1695, 2014.

[12] A. K. Pietilänen and C. Diot, Dissemination in opportunistic social networks: The role of temporal communities, in *Proc. 13th ACM Int. Symp. Mobile Ad Hoc Networking and Computing*, Hilton Head, SC, USA, 2012.

[13] K. Thilakarathna, A. C. Viana, A. Seneviratne, and H. Petander, Mobile social networking through friendto-friend opportunistic content dissemination, in *Proc. 14th ACM Int. Symp. Mobile Ad Hoc Networking and Computing*, Bangalore, India, 2013.

[14] C. Boldrini, M. Conti, and A. Passarella, Contentplace: Social-aware data dissemination in opportunistic networks, in *Proc. 11th Int. Symp. Modeling Analysis and Simulation of Wireless and Mobile Systems*, Vancouver, Canada, 2008.

[15] J. L. Fan, Y. Du, W. Gao, J. M. Chen, and Y. X. Sun, Geographyaware active data dissemination in mobile social networks, in *Proc. 7th IEEE Int. Conf. on Mobile Adhoc and Sensor Systems*, San Francisco, CA, USA, 2010.

[16] J. L. Fan, J. M. Chen, Y. Du, W. Gao, J. Wu, and Y. X. Sun, Geocommunity-based broadcasting for data dissemination in mobile social networks, *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 734–743, 2013.

[17] X. F. Wang, Z. G. Sheng, S. S. Yang, and V. C. M. Leung, Tagassisted social-aware opportunistic device-to-device sharing for traffic offloading in mobile social networks, *IEEE Wirel. Commun.*, vol. 23, no. 4, pp. 60–67, 2016.

[18] X. Zhang, Z. H. Zhang, J. Xing, R. Yu, P. W. Zhang, and W. B. Wang, Exact outage analysis in cognitive two-way relay networks with opportunistic relay selection under primary user's interference, *IEEE Trans. Veh. Technol.*, vol. 64, no. 6, pp. 2502–2511, 2015

[19] J. T. Li, Q. F. Qi, Q. C. Xu, and Z. Su, Graph based content delivery in mobile social networks, in *Proc. 12th Int. Conf. Mobile Ad-Hoc and Sensor Networks* (*MSN*), Hefei, China, 2016, pp. 348–352.

[20] Q. C. Xu, Z. Su, and S. Guo, A game theoretical incentive scheme for relay selection services in mobile social networks, *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6692–6702, 2016.

[21] U. Shevade, H. H. Song, L. L. Qiu, and Y. Zhang, Incentiveaware routing in DTNs, in *Proc. 2008 IEEE Int. Conf. Network Protocols*, Orlando, FL, USA, 2008.

[22] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau, Crawdad trace cambridge/haggle/imote/infocom2006, http://crawdad.cs.dartmouth.edu/cambridge/haggle/imote/infocom, 2006.

[23] A. K. Pietilainen and C Diot, The thlab/sigcomm2009 (v20120715), http://crawdadorg/thlab/sigcomm2009/, 2012.

[24] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, Measurement and analysis of online social networks, in *Proc. 7th ACM SIGCOMM Conf. Internet Measurement*, San Diego, CA, USA, 2007.

[25] C. O. Rolim, V. R. Q. Leithardt, A. G. Rossetto, T. F. M. dos Santos, A. M. Souza, and C. F. R. Geyer, Six degrees of separation to improve routing in opportunistic networks, *Int. J. UbiComp.*, vol. 4, no. 3, pp. 11–22, 2013.

[26] T. Hossmann, T. Spyropoulos, and F. Legendre, A complex network analysis of human mobility, in *Proc. 2011 IEEE Conf. Computer Communications Workshops*, Shanghai, China, 2011.

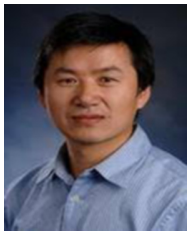
[27] A. H. Land and A. G. Doig, An automatic method of solving discrete programming problems, *Econometrica*, vol. 28, no. 3, pp. 497–520, 1960.

[28] Y. S. Wang and J. Wu, Social-tie-based information dissemination in mobile opportunistic social networks, in *Proc. 14th IEEE Int. Symp.* "*A World of Wireless Mobile and Multimedia Networks*" (*WoWMoM*), Madrid, Spain, 2013.

Chenguang Kong received the BS degree from University of Science and Technology of China and MS degree from University of Hong Kong in 2009 and 2011, respectively. In 2016, he received the PhD degree from Georgia State University. He currently is a data scientist in Airsage Inc. He has published numbers of conference and journal papers. His research interests include social networks analysis and big data.

Guangchun Luo received the BS, MS, and PhD degrees from University of Electronic Science and Technology of China (UESTC) in 1995, 1999, and 2004, respectively. He is currently a professor in the School of Information and Software at UESTC. He has published over sixty journal and conference papers in his fields. His research interests include computer networks and big data.

Ling Tian received the BS, MS, and PhD degrees from University of Electronic Science and Technology of China (UESTC) in 2003, 2006, and 2010, respectively. She is currently a professor in UESTC. She had been a visiting scholar in Georgia State University (GSU) in 2013. She has edited 2 books and holds over 10 Chinese patents. Her research interests include image/video coding and big data.



Xiaojun Cao received the BS degree from Tsinghua University and MS degree from Chinese Academy of Sciences in 1996 and 1999, respectively. In 2004, he received the PhD degree in Computer Science and Engineering from the State University of New York at Buffalo. Currently, he is a professor at Computer Science Department of Georgia State University. He and his group are working on modeling, analysis, protocols/algorithms design as well as data processing for networks, and cyber physical systems. His research has been sponsored by U.S. National Science Foundation (NSF), Centers for Disease Control and Prevention, IBM, and Cisco University Research Program. He is a recipient of NSF CAREER Award between 2006 and 2011. He actively serves the research communities in various roles. He currently chairs IEEE ComSoc Optical Networking Technical Committee (ONTC), and serves as an associate editor for *IEEE Communications Magazine* and *Photonic Network Communications*. He has served over many international conferences as a member of the organizing committee such as the program co-chair for IEEE International Conference on Big Data and Cloud Computing 2016, and the symposium co-chair for IEEE GLOBECOM 2016 and IEEE ICC 2019.