Space-Crossing: Community-Based Data Forwarding in Mobile Social Networks Under the Hybrid Communication Architecture

Zhong Li, Cheng Wang, Siqian Yang, Changjun Jiang, and Ivan Stojmenovic, *Fellow, IEEE*

***Abstract*—In this paper, we study two tightly coupled issues, space-crossing community detection and its influence on data forwarding in mobile social networks (MSNs). We propose a communication framework containing the hybrid underlying network with access point (AP) support for data forwarding and the base stations for managing most of control traffic. The concept of physical proximity community can be extended to be one across the geographical space, because APs can facilitate the communication among long-distance nodes. Space-crossing communities are obtained by merging some pairs of physical proximity communities. Based on the space-crossing community, we define two cases of node local activity and use them as the input of inner product similarity measurement. We design a novel data forwarding algorithm Social Attraction and Infrastructure Support (SAIS), which applies similarity attraction to route to neighbor more similar to destination, and infrastructure support phase to route the message to other APs within common connected components. We evaluate our SAIS algorithm on real-life datasets from MIT Reality Mining and University of Illinois Movement (UIM). Results show that space-crossing community plays a positive role in data forwarding in MSNs. Based on this new type of community, SAIS achieves a better performance than existing popular social community-based data forwarding algorithms in practice, including Simbet, Bubble Rap and Nguyen's Routing algorithms.**

***Index Terms*—Mobile social networks, infrastructure support, space-crossing community, data forwarding.**

I. Introduction

THE mobile social network (MSN) is a mobile communication system involving social relationships of the users [1]. The majority of mobile social applications (e.g., Facebook, WeChat, Myspace) are currently implemented over the centralized cellular networks due to the high efficiency of service implementation and management under such a centralized architecture. However, if data transmitting totally relies on the base stations in cellular networks, there will be some disadvantages, such as the issues related to overload and privacy. To be specific, first, according to Cisco's global mobile data traffic forecast [2], global mobile data traffic will increase nearly 11-fold between 2013 and 2018. By the end of 2014, the number of mobile-connected devices will exceed the number of people on earth, and by 2018 there will be nearly 1.4 mobile devices per capita. So, we can see that receiving and forwarding large amount of information via base stations will make the stations overloaded and easily become the bottlenecks; second, if much privacy information is laid on stations, it is usually prone to be stolen by the third part, meanwhile, not every user likes to give their privacy social information to the public centralized base stations [1]. For example, there are some studies [3]–[6] show that the data in the server would be stolen by some adversary adversaries. The adversary will use the methods of machining learning, trajectory reconstruction and so on to obtain some sensitive information, like the patient gene, daily trip trajectory, and the sensitive personal information in mobile phones.

An intuitive solution to these problems is to introduce the ad hoc framework and realize some peer-to-peer mobile social applications. Moreover, MSNs under such an ad hoc framework have been extensively studied [7]–[12]. An observation to all these studies is that they limit the underlying communication networks for mobile social applications to the case at the other extreme, i.e., the purely decentralized ad hoc networks, allowing no base station to get involved. Furthermore, they totally neglected the low performance in terms of service stability and global control resulted from the pure ad hoc networks, regardless of the real-world communication architecture for MSN. In a common real-world scenario for MSNs, the centralized cellular (base station based) networks and decentralized ad hoc networks usually coexist. We propose a more practical framework where ad hoc networks undertake the heavier task, i.e., data transmission, while cellular networks assist doing a much lighter task, i.e., management and coordination of some relevant global information.

Cellular networks can be assumed to be 3G or LTE networks. We adopt the WIFI networks for the ad hoc networks. Specially, in the practical framework stated above, we take a full consideration for the role of access points (APs) in WIFI networks for data transmission, and model the WIFI network to a hybrid communication paradigm consisting of both direct ad hoc

Manuscript received June 8, 2014; revised December 5, 2014 and February 14, 2015; accepted April 8, 2015. Date of publication April 21, 2015; date of current version September 7, 2015. This work is partially supported by the Integrated Project for Major Research Plan of the National Natural Science Foundation of China under Grant No. 91218301, the Key Program for International S&T Cooperation Projects of China under Grant No. 2012DFG11580, Shanghai ShuGuang Project under Grant No. 14SG20, Shanghai Rising-Star Program under Grant No. 14QA1403700, the National Natural Science Foundation of China (NSFC) under Grant No. 61202383, the Program for New Century Excellent Talents in University (NCET) under Grant No. NCET-12-0414, the Natural Science Foundation of Shanghai under Grant No. 12ZR1451200, the Research Fund for the Doctoral Program of Higher Education of China (RFDP) under Grant No. 20120072120075. The associate editor coordinating the review of this paper and approving it for publication was M. C. Vuran. *(Corresponding author: Changjun Jiang.)*

Z. Li is with the College of Information Science and Technology, Donghua University, Shanghai 201620, China.

C. Wang, S. Yang, and C. Jiang are with the Department of Computer Science, Tongji University, Shanghai 201804, China (e-mail: cjjiang@tongji.edu.cn).

I. Stojmenovic is with the School of Electrical Engineering and Computer Science (SEECS), University of Ottawa, Ottawa, ON K1N 6N5, Canada.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2015.2424965

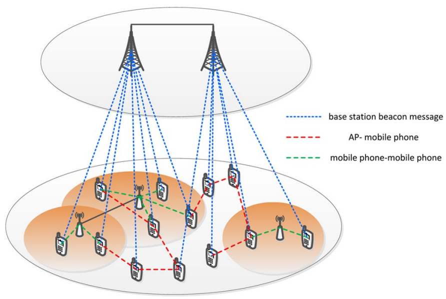


Fig. 1. Proposed architecture.

communications among users and AP-user communications. We also can adopt other short-distance communication standards for the ad hoc networks, such as Bluetooth, with the hybrid communication network consisting of direct ad hoc communications among users using Bluetooth and AP-user communications using WIFI. Some assumptions about APs are given in Section II-B. In this paper, we mainly address the routing schemes for data transmission in MSNs under such a communication framework.[1] The overall architecture is illustrated in Fig. 1.

This communication framework has certain advantages. First, as a popular wireless broadband access technology, WIFI is recognized as one of the primary offloading technologies [15]–[17]. The hybrid underlying network can be seen as an offloading network for base stations (delivering data originally targeted for cellular networks by WIFI). Next, as a self-organized mode, the ad hoc communication can still be used as an extension of centralized base stations and prevent users from exposing much privacy information to public. Finally, the base station is only used to send some periodic beacon information to do some simple and auxiliary global work. The global and high efficient control capability is kept in the network.

Under the hybrid communication architecture, when studying the data forwarding in mobile social networks, we need to pay attention to two importance aspects. One is the infrastructure support; the other is the social attributes of nodes. So, we propose a space-crossing community method to solve the high efficiency data forwarding problem in such MSNs.

Traditionally, a community is defined as a group of tight-knit nodes with more internal than external links [18]–[20]. The definition of community is quite subjective and depends on the specific application. There is not a uniform definition about community [1]. In mobile social networks (MSNs), the links usually represent the social friendship. The friendship can be formed according to various factors. One of the representative friendships/links results from device physical contact frequency. It has demonstrated by some studies [12], [21] that the physical contact frequency and the friendship impact with each other. In our paper, we just focus on this kind of mobile social network. It can be called the geographic proximity mobile social network that also has been studies by some researchers in literature [7], [8], [22]. If the underlying network in MSNs uses the ad hoc way, according to the traditional definition of the community, the community will represent a group of nodes with high physical contact frequency. It also can be termed as the physical proximity community. The links among devices are the direct physical contact/communication, for example, the Bluetooth and/or WIFI connection of the mobile phones or portable tablets. The community structure shows stronger communication capability. Now, in our hybrid communication architecture, the underlying network has AP supports. Some long-distance nodes without direct links could also communicate frequently through APs although there are no direct contacts (links) among them. These long-distance nodes with strong communication capability should be partitioned into the same group to show the friendships among them. So, the traditional definition of the community "a group of tight-knit nodes with more internal links than external links" will not fit for the hybrid communication architecture since there are no direct links between some long-distance nodes. Thus, the definition of community changes in nature. Then, we give a new definition of community in this hybrid communication architecture, called space-crossing community. The space-crossing community can be described/defined as the following two steps: 1) According to the direct contacts among some APs and mobile users, we can first obtain some physical proximity communities by using the traditional community detection method. 2) But, it is not the end since some long-distance nodes without direct contacts have not been dealt with. Then, spanning physical space, we combine some physical proximity communities via several APs to obtain the space-crossing communities. So, the space-crossing community reflects the aspects of the infrastructure support and the social community attribute simultaneously.

Based on space-crossing community structure, we define *local activity* of a node (including the mobile user and the stationary access point) with respect to its belonging community. According to the contact records between *user pairs* and *user-AP pairs*, the local activity of a mobile user u in its belonging space-crossing community $ComSC$ is defined as the ratio of u's encounter probability with other nodes in $ComSC$ to the sum of encounter probability between any two nodes in $ComSC$. Note that call logs among users (explicit social links) are not considered in our local activity definition. Similarly, according to the contact records of *user-AP pairs*, each stationary AP r also has a statistical ratio value of r's encounter probability with other nodes in $ComSC$ to the sum of encounter probability between any two nodes in $ComSC$. Specially, the local activity of an AP r is defined as the sum of the statistical ratio values of APs in common connected components within its common space-crossing communities. Then, based on the node local activity, we use the inner product similarity and design an efficient data forwarding scheme to fully demonstrate the role of the space-crossing community structure.

The main contributions of this paper can be summarized as follows:

- To the best of our knowledge, the proposed architecture, consisting of mobile phones, APs and base stations, was not considered for any MSN scenario.
- We give the detailed descriptions of the space-crossing community detection method in MSNs by taking the hybrid underlying network with APs support into consideration.
- A SAIS (Similarity Attraction and Infrastructure Support) data forwarding scheme is proposed, which is based on the detection results of space-crossing communities.

[1]Some studies about the data forwarding protocol, like literature [13], [14], are also based on the underlying network with AP support. However, 1) they do not consider the role of base stations; 2) the APs have no "brains" to decide whether to do data forwarding, i.e., they are passive. So, our network architecture is different from them.

The rest of the paper is organized as follows. Section II presents the network model. Sections III and IV study the space-crossing community detection and its impact on data forwarding in MSNs respectively. We conduct extensive experiments and analyze our results in Section V. We simply review the related work in Section VI. Finally, we conclude the paper in Section VII.

II. Network Model

A. Dynamic Graph

We model this hybrid underlying network with APs support as a *dynamic graph* which can be defined as a time sequence of network graph, denoted by $\mathcal{G} = \{G_0, G_1, \ldots, G_t, \ldots\}$, where $G_t = (V_t, E_t)$ represents a time dependent network snapshot recorded at time $t; V_t$ denotes the set of nodes, including the set of mobile users and the set of stationary APs; $E_t = \{(u, v)|u, v \in V_t\}$ denotes the edge set. Both node and edge sets change over time.

B. Assumption of AP

In our mobile social network, the ratio of the number of APs to the number of mobile users is small. The coverage area of all APs does not occupy the entire mobile social network. Some of the APs are interconnected via wired or wireless links. Many studies usually utilize an unrealistic assumption, in which that large amount of APs are all connected with a backbone network in city area. This assumption aims at simplifying the network analysis. In this kind of network environment, the data flooding along the backbone results in heavy workloads on all APs. It is not a recommendable data forwarding scheme. In fact, we can partition some AP connected components according to the geographical area. In our paper, for simplicity, we partition some AP connected components based on the following steps: 1) system dispatches the sequence numbers (natural number) to each AP; 2) according to ascending sequence number (from small to large sequence number), random numbers (2–5) of APs are grouped into different connected components in which APs are connected one by one (a chain). So, our underlying network with AP support could not only guarantee a certain amount of copies in the network, but also avoid introducing a large scale data flooding.

C. Contact Aggregation for Edges

Contact records of Bluetooth and WIFI access points are usually contained in social datasets. To show a cumulative effect on mobile social network, a median-based sliding window mechanism will be implemented.

Let $l(u, v, t) = 1$ denote that there *starts* a contact between node u and v at time t $(0 \le t < \infty)$. Then, we have $\sum_{t=t_{now}-\Delta}^{t_{now}} l(u, v, t)$ denote the overall numbers of contacts between node u and v from time $t_{now} - \Delta$ to t_{now}; and have $\sum_{t=t_{now}-\Delta}^{t_{now}} l_*(t)$ denote the overall numbers of contacts for all nodes from time $t_{now} - \Delta$ to $t_{now}, 0 < \Delta < t_{now}$.[2] We define the encounter ratio between node u and v at current time t_{now}[3] as

$$w(u, v, t_{now}) = \frac{\sum_{t=t_{now}-\Delta}^{t_{now}} l(u, v, t)}{\sum_{t=t_{now}-\Delta}^{t_{now}} l_*(t)}.$$

[2]In the paper, we take the length of sliding window Δ as a constant, not a variable. The window length of sliding window mechanism is usually empirically determined [7], [23]. We have Δ equal 6∗3600 s, since the social based comparison algorithm, like literature [7] used in the experiment, is also set this value.

[3]In the following sections, for brevity, denote the current time t_{now} by t, without confusion.

An edge between nodes u and v will be created if $w(u, v, t_{now})$ is larger than the median of $\{w(u', v', t_{now})|u', v' \in V_t \text{ and } w(u', v', t_{now}) \neq 0\}$.

Remark 1: To cope with the asymmetric detection of Bluetooth devices in datasets, we have $w(u, v, t_{now}) = w(v, u, t_{now})$, by assigning the larger value to the other.

D. Space-Crossing Community

In the hybrid underlying network, every mobile user and stationary access point can be viewed as an independent participant (node). First, due to frequent interactions (physical proximity) between the mobile user pairs and user-AP pairs, users and APs will form some dense groups, called physical proximity communities. Then, with the help of connectivity among some APs, a part of long-distance nodes that are in different physical proximity communities containing APs could have the capability to communicate with each other. We combine those physical proximity communities through APs to form some groups across the geographical space, called *space-crossing communities*. Both the physical proximity and space-crossing communities reflect structures with relatively strong communication capability. We allow that space-crossing communities can overlap with each other. The detailed description of *Space-Crossing Community Detection* algorithm is provided in Section III.

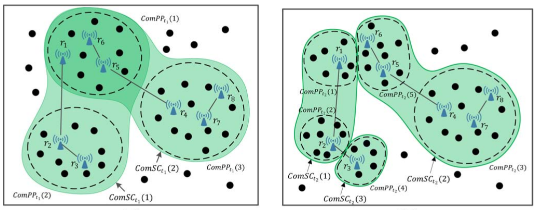
III. Space-Crossing Community Detection

In the dynamic and hybrid underlying networking environment, we detect the space-crossing communities in two basic ways. First, according to our dynamic graph, we use AFOCS [11] to obtain physical proximity communities at each time slot. Second, using combination criterion CA (at time slot t_1) or combination criterion CB (in all subsequent time slots), we get final space-crossing communities at each time slot. Finally, we have two sequences, one is a dynamic time sequence of physical proximity community structure, denoted by $\{\mathcal{PP}_0, \mathcal{PP}_1, \ldots, \mathcal{PP}_t, \ldots\}$. Let $ComPP_t(i)$ represent the $i-th$ physical proximity community at time t in $\mathcal{PP}_t$. The other is a dynamic time sequence of space-crossing community structure, denoted by $\{\mathcal{SC}_0, \mathcal{SC}_1, \ldots, \mathcal{SC}_t, \ldots\}$. Let $ComSC_t(i)$ represent the $i-th$ space-crossing community at time t in $\mathcal{SC}_t$.

At initial network snapshot, based on network graph G_1 (*the nodes in* G_1 *contain APs and mobile users*) defined in Section II-A, the *initial locating community phase of algorithm AFOCS* [11] is applied to obtain initial set of physical proximity communities $\mathcal{PP}_1$, as illustrated in technical report [59, Appendix B, Algorithm 1].

We give a simple but practical combination criterion CA to obtain the initial set of space-crossing communities $\mathcal{SC}_1$. Combination criterion CA is described as following steps.

(1) System dispatches the sequence numbers (natural number) to each AP. Each AP maintains a mark (*undone*) that means the belonging community did not apply yet the CA combination operation at current time slot.
(2) Suppose that, for simplicity, according to ascending sequence number (from small to large sequence number), random numbers of APs are grouped into different connected components in which APs are connected one by one (a chain).



(a) (b)

Fig. 2. There are three connected components of APs ($r_1 - r_2 - r_3, r_4 - r_5 - r_6, r_7 - r_8$). Subfigure (a) shows the result of the space-crossing community detection at time slot t_1. As network evolves, subfigure (b) shows the dynamic changes/updates of space-crossing communities at time slot t_2. The black solid line means the wire links between APs. The black dash line circles the physical proximity communities. The green area covers the space-crossing communities.

(3) A certain AP r will first check 1) if the AP r' in its left order (i.e., the sequence number of r' is smaller than r) has a link with it; 2) the mark of combination criterion CA of AP r' is *undone*; 3) they are not in the same physical proximity communities.

(4) If the above three conditions are satisfied, combine the physical proximity communities containing r and r' into a new space-crossing community; and set the mark of combination criterion CA of AP r and r' as *done*.

(5) Otherwise, the same operations (3) and (4) will be done for the AP r' in its right order (the sequence number of r' is larger than r).

Assuming there are R APs in a certain connected component of APs, the number of space-crossing communities will be at most $\frac{R}{2}$. Pseudo-codes of combination criterion CA in simulation are shown in technical report [59, Appendix C, Algorithm 6]. Our combination criterion eases the workload on APs and avoids forming a large space-crossing community around the backbone network of APs.

Fig. 2(a) gives an intuitive illustration of the combination criterion CA at time slot t_1. Suppose that, after applying the initial locating community phase of algorithm AFOCS, we have three physical proximity communities $ComPP_t(1) \sim ComPP_t(3)$, which contain mobile users and APs. There are three connected components of APs ($r_1 - r_2 - r_3, r_4 - r_5 - r_6, r_7 - r_8$). According to combination criterion CA, in the first connected component, space-crossing communities $ComSC_t(1)$ can be formed by $r_1 - r_2$; in the second connected component, space-crossing communities $ComSC_t(2)$ can be formed by $r_4 - r_5$. Since r_7 and r_8 are already in the same physical proximity community, there is no need to apply combination criterion and r_7 and r_8 are also in the same space-crossing community.

For subsequent time slot t ($2 \leq t < \infty$), based on $\mathcal{PP}_{t-1}$, the *adaptive finding community phase of algorithm AFOCS* [11] is applied to obtain the set of physical proximity communities $\mathcal{PP}_t$. In the adaptive finding community phase of algorithm AFOCS, dynamic network changes are classified into four simple actions: adding new nodes, adding edges, removing nodes and removing edges. For each kind of changes, Algorithms 2-5 in technical report [59, Appendix B] gives corresponding methods to adaptively find the updated physical proximity communities.

Based on above $\mathcal{PP}_t$ ($2 \leq t < \infty$), we apply combination criterion CB to obtain the set of space-crossing communities $\mathcal{SC}_t$ ($2 \leq t < \infty$). Combination criterion CB is described as follows.

(1) At time slot t ($2 \leq t < \infty$), each AP maintains an initial mark (*undone*), meaning that the community did not yet apply the CB combination operation at the current time slot.

(2) When a certain AP r whose mark of combination criterion CA is *done* finds that the size of its belonging physical proximity communities changes.

(3) The following steps are similar to the combination criterion CA (3)–(5). The difference lies in changing the mark of combination criterion CA into the mark of combination criterion CB.

The combination criterion CB adaptively and locally combines (updates) the changed physical proximity communities into new space-crossing communities as network evolves. Pseudo-codes of combination criterion CB in simulation are shown in technical report [59, Appendix C, Algorithm 7].

Fig. 2(b) illustrates the combination criterion CB at time slot t_2. Based on the left picture, suppose that four kinds of network changes (adding new nodes, adding edges, removing nodes and removing edges) have taken place. After applying the adaptive finding community phase of algorithm AFOCS, we have a new physical proximity community structure containing $ComPP_t(1) \sim ComPP_t(5)$. We find that the physical proximity communities where APs r_1, r_2, r_3, r_5, r_6 belong to have changed. Therefore, using combination criterion CB, in Fig. 2(b), due to r_1 and r_2, $ComPP_t(1)$ and $ComPP_t(2)$ are locally combined to form a new space-crossing community $ComSC_t(1)$; due to r_5, $ComPP_t(3)$ and $ComPP_t(5)$ are also locally combined to form a new space-crossing community $ComSC_t(2)$.

We provide the detailed call order of different algorithms and combination criteria at different time phases in space-crossing community detection in technical report [59, Appendix D].

Remark 2: The analysis on the effectiveness and complexity of algorithm CA and CB is provided in technical report [59, Appendix E].

Remark 3: The simple combination criteria CA and CB can ensure the operability of the detection method. The idea of coupled combination mainly guarantees that the APs will not bear much overhead for the combination workload. In real application, based on AP communication load, a more complex and alternative combination criterion can be applied, e.g., as provided in technical report [59, Appendix H].

Remark 4: AFOCS is a viable choice (some advantages are given in the related work) for our basic physical proximity community detection. From technical prospective, other detection methods mentioned in technical report [59, Appendix A] also can be used in finding physical proximity community.

IV. SAIS Data Forwarding Scheme

In this section, based on the results of Space-Crossing Community Detection, we design a SAIS (Similarity Attraction and Infrastructure Support) data forwarding scheme to validate the positive role of the space-crossing community.

A. Two Cases of Local Activity and Inner Product Social Similarity

Definition 1 Local Activity:

Case I For Mobile Users: Let $\mathrm{LAV}_t(i,u)$ denote the local activity of a mobile user u in a space-crossing community $ComSC_t(i)$ at time t. Then,

$$\text{LAV}_t(i,u) = \frac{\sum_{(u,v)\in ComSC_t(i)} w(u,v,t)}{\sum_{(v',v'')\in ComSC_t(i)} w(v',v'',t)}, 1 \le i \le K,$$

where v' and v'' are any two nodes in $ComSC_t(i)$; $w(u,v,t)$ has been defined in Section II-C; K represents the number of space-crossing communities. The numerator represents the sum of the encounter ratio between node u and other nodes in community $ComSC_t(i)$ and the denominator represents the sum of the encounter ratio between any two nodes in community $ComSC_t(i)$.

Case II For Stationary APs: Let $CC_t(i,j)$ denote the $j-th$ connected component of APs in a certain space-crossing community $ComSC_t(i)$. We define the local activity of every AP $r_k \in CC_t(i,j)$ as

$$\text{LAP}_t(i,r_k) = \sum_{r_m\in CC_t(i,j)} \text{LAV}_t(i,r_m),$$

where for every AP $r_m \in CC_t(i,j), \text{LAV}_t(i,r_m)$ is obtained according to the method provided in *Case I*, treating APs as ordinary mobile users.

Here, a calculation example of the local activity of an AP is provided. As illustrated in Fig. 2(a), there are two connected components of APs $(r_1-r_2-r_3, r_5-r_6)$ in space-crossing community $ComSC_t(1)$ and two connected components of APs $(r_4-r_5-r_6, r_7-r_8)$ in space-crossing community $ComSC_t(2)$, respectively.Then, we have

$$\begin{aligned}
\text{LAP}_t(1,r_1) &= \text{LAP}_t(1,r_2) = \text{LAP}_t(1,r_3)\\
&= \text{LAV}_t(1,r_1) + \text{LAV}_t(1,r_2) + \text{LAV}_t(1,r_3);\\
\text{LAP}_t(1,r_5) &= \text{LAP}_t(1,r_6)\\
&= \text{LAV}_t(1,r_5) + \text{LAV}_t(1,r_6);\\
\text{LAP}_t(1,r_4) &= \text{LAP}_t(1,r_7) = \text{LAP}_t(1,r_8) = 0;\\
\text{LAP}_t(2,r_4) &= \text{LAP}_t(2,r_5) = \text{LAP}_t(2,r_6)\\
&= \text{LAV}_t(2,r_4) + \text{LAV}_t(2,r_5) + \text{LAV}_t(2,r_6);\\
\text{LAP}_t(2,r_7) &= \text{LAP}_t(2,r_8)\\
&= \text{LAV}_t(2,r_7) + \text{LAV}_t(2,r_8);\\
\text{LAP}_t(2,r_1) &= \text{LAV}_t(2,r_1);\\
\text{LAP}_t(2,r_2) &= \text{LAP}_t(2,r_3) = 0.
\end{aligned}$$

The local activity of a node (including the mobile user and the stationary AP) can represent the importance of the node in a certain community. A larger local activity means that the node has more interactions with other members in the community. In data forwarding, local activity is important because if the message is given to a node having low local activity, it will cause a low efficiency in terms of delivery ratio.

Definition 2 Activity Vector: For each mobile user u, we define the activity vector at time t as

$$\begin{aligned}
&A_t(u)\\
&= (\text{LAV}_t(1,u), \text{LAV}_t(2,u), \ldots, \text{LAV}_t(i,u), \ldots, \text{LAV}_t(K,u));
\end{aligned}$$

for each AP r, we define the activity vector at time t as

$$\begin{aligned}
&A_t(r)\\
&= (\text{LAP}_t(1,r), \text{LAP}_t(2,r), \ldots, \text{LAP}_t(i,r), \ldots, \text{LAP}_t(K,r)),
\end{aligned}$$

where $\text{LAV}_t(i,u)$ and $\text{LAP}_t(i,r)$ denote the local activity of mobile user u and AP r respectively in space-crossing community $ComSC_t(i)$ at time t. K represents the number of communities after applying the space-crossing community detection method.

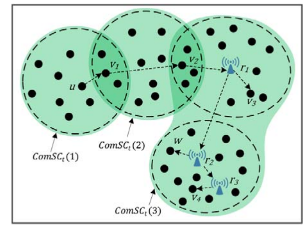


Fig. 3. A session from node u to w. The black dotted arrow means a data forwarding path. Similarity attraction phase occurs at transmissions from u to v_1, v_1 to v_2, v_2 to r_1, r_1 to v_3, r_3 to v_4 and r_2 to w. Infrastructure support phase occurs at transmissions from r_1 to r_2 and r_2 to r_3.

There are some similarity measurements, such as Euclidean distance, Hamming distance, inner product measurement, Pearson correlation coefficient and Jaccard coefficient [24], which are often used in the research of social network, recommendation system and clustering analysis in web search [11], [12], [25]–[28]. When being applied in different environments, each measurement has its advantage and disadvantage [12], [26], [29].

In the paper, we want to concern on two key aspects, one is local activity, the other is the distribution of the node belonging communities. We require the larger node local activity and more similar distribution of the node belonging communities to the destination, simultaneously. However, some of above measurements can not meet our requirement. For example, the cosine angular distance and Pearson correlation coefficient only focus on the distribution of the node belonging communities with the destination; the Euclidean distance puts emphasis on the component value (local activity) of two vectors; the Jaccard coefficient only considers the number of common communities with the destination. So, according to the characteristic of our input (activity vector), inner product measurement is properly chosen as our social similarity, defined in Definition 3.

Definition 3 [12] Inner Product Social Similarity Social Similarity: Given two activity vectors $A_t(u)$ of node u and $A_t(w)$ of node w, we define the social similarity between u and w at time t as $SS_t(u,w)$, having $SS_t(u,w) = A_t(u) \cdot A_t(w)$, where the symbol $\cdot$ denotes the inner product of vectors.

Space-crossing community can reflect the node's belonging property from the view of physical communication among nodes. Local activity can show the importance of a node in a certain space-crossing community. Based on destination-oriented aim, a larger inner product social similarity can not only guarantee to find a node that has similar distribution of the node belonging communities to the destination, but also can gain larger local activity entries in the vector. Thus, a larger inner product social similarity to the destination indicates that the relay node has higher chance to approach to the destination.

B. SAIS Data Forwarding Scheme

In this section, we provide the details of SAIS (Similarity Attraction and Infrastructure Support) algorithm. It includes two phases: *Similarity Attraction Phase* and *Infrastructure Support Phase*, as illustrated in Fig. 3.

1) Similarity Attraction Phase: In this phase, each node has its activity vector. When the message holder (including the mobile user and the stationary AP) meets another node, they will calculate their inner product social similarities to the destination respectively. The calculation is enabled by beacon messages sent by base stations periodically, so that the source node can learn the activity vector of the destination. The message holder tries to send the message to a node which has larger social similarity than itself and let the node send the message to the destination consecutively.

2) Infrastructure Support Phase: In this phase, a message holder AP delivers the message to other APs in its common connected components.

3) SAIS Algorithm: The pseudo-codes of the SAIS algorithm are described in technical report [59, Appendix F, Algorithm 8]. We do not distinguish the two phases in sequence. This is because the message exchange in SAIS is compatible with each phase.

- Mobile user holding the message chooses its next relay from the APs and mobile users according to the requirement of *Similarity Attraction Phase*.
- An AP holding the message first delivers the message to other APs according to the requirement of *Infrastructure Support Phase*; and then all these AP message holders choose their next relays from mobile users according to the requirement of *Similarity Attraction Phase*.

Note that, the implementation and future application of the space-crossing community detection and SAIS algorithm are provided in technical report [59, Appendix G].

V. Performance Evaluation

In this section, we evaluate the performance of SAIS based on two tracing datasets. One is MIT Reality Mining [30], the other is UIM (University of Illinois Movement) [31]. The detailed dataset selection and simulation setup are provided in technical report [59, Appendix I].

A. Comparisons With Other Forwarding Schemes

We compare SAIS algorithm with Simbet [32], BUBBLE RAP [7] and Nguyen's Routing [11] (i.e., three popular social community-based routing algorithms).

B. Comparison Fairness

For the sake of fairness, first, we select settings or parameters which bring about the best performances for these comparison algorithms. Additionally, since the comparison algorithms are not based on the hybrid underlying network, we treat APs as the ordinary mobile users to fit the operability requirement and use the fair AP strategy (spreading the messages in AP connected component) for the comparison algorithms. That is to say, Simbet, BUBBLE RAP, and Nguyen's Routing also can be used in the environment with AP supports. *They all have the same multiple copies strategy on APs with our SAIS forwarding scheme*.

C. Experiment Results and Analysis

Figs. 4 and 5 show the delivery ratio, overhead ratio and average latency of SAIS, Simbet, BUBBLE RAP and Nguyen's Routing algorithms in MIT and UIM datasets, respectively. We can see, for the two datasets, the delivery ratio of SAIS achieves best among those algorithms while the overhead ratio and the average latency are lowest.

In terms of delivery ratio, Fig. 4(a) and Fig. 5(a) show that SAIS performs best among those algorithms. In MIT Reality Mining dataset, its delivery ratio is higher than Nguyen's Routing with 56.21 percent, BUBBLE RAP with 70.32 percent and Simbet with 127.6 on average. In UIM dataset, its delivery ratio is higher than Nguyen's Routing with 70.32 percent, BUBBLE RAP with 79.3 percent and Simbet with 197.12 on average. The peak value of SAIS arises later that the other three comparison algorithms. After that, due to the increasing TTL and the copies of messages, it is normal that the delivery ratios of all algorithms decrease slightly. However, the degree of the decline in SAIS is smaller than the comparison algorithms.

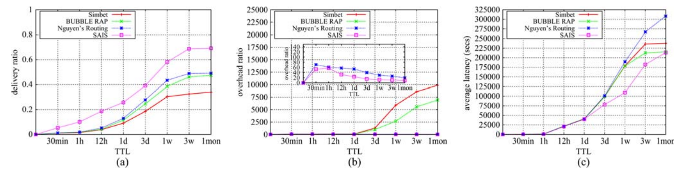


Fig. 4. Simulation Results on MIT Reality Mining Dataset. Subfigure (a), Subfigure (b) and Subfigure (c) show the delivery ratio, overhead ratio and average latency respectively.

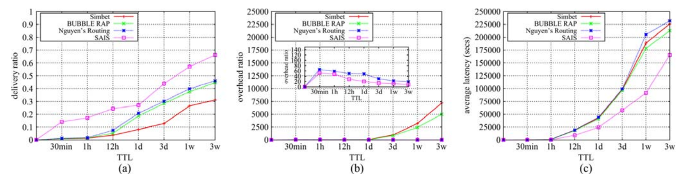


Fig. 5. Simulation Results on UIM Dataset. Subfigure (a), Subfigure (b) and Subfigure (c) show the delivery ratio, overhead ratio and average latency respectively.

This advantage is mainly brought about by the space-crossing community structure. In the current hybrid underlying network, if we use the traditional community detection method, since the remote geographical area, two nodes with frequent communication through APs will be partitioned into two different physical proximity communities. Then, the social similarity between the two nodes will be smaller than the cast that they are in the same community. So, due to small social similarity, the data forwarding may not occur between the two nodes even they encounter with each other. However, in reality, due to AP facilitation, a message transmission is proper to occur between the two nodes. That is to say, the transmission capability between them is strong. Now, our space-crossing detection is fit for the hybrid underlying network. On the premise of not increasing the work load of APs, the above two nodes will be partitioned into the same space-crossing community. The social similarity between the two nodes becomes large. A data forwarding will be more likely to occur than before. So, based on space-crossing community structure, using the inner product social similarity and local AP transmission, SAIS has better performance than three comparison algorithms in terms of delivery ratio. Especially, we independently do some simulations to test the role of space-crossing community in technical report [59, Appendix J].

For Simbet and BUBBLE RAP, besides the problem of without considering the space-crossing community, another problem is the employment of global betweenness in entire or partial phase of data forwarding. In Simbet and BUBBLE RAP, the concept of community is considered. That is to say, each node has its belonging community(ies), expect solitary nodes. If we deliver the message to a node having high global betweenness, although it indeed has high contact frequency with other nodes with respect to the entire network, it may be in a community which is irrelevant to or does not overlap with the destination community. As a relay, this node is not proper and increases the time of reaching the destination. For Nguyen's Routing, it tends to send messages to nodes having many interests with the destination. However, it may deliver them to the nodes which have low local activity in their communities (or interests groups). This is the main reason of the low delivery ratio of Nguyen's Routing.

In Fig. 4(b) and Fig. 5(b), on average, SAIS keeps a low overhead ratio of 25.28 in MIT Reality Mining and 24.84 in UIM. It is better than Nguyen's Routing which is 44.72 in MIT Reality Mining and 42.13 in UIM. The overhead ratios of Simbet and BUBBLE RAP are much higher than SAIS and Nguyen's Routing. The reason is SAIS and Nguyen's Routing prefer to choose the similar interests nodes as relays, which can control the number of copies in sessions. In the enlarged legend, the overhead ration of SAIS and Nguyen's Routing are descending with TTL increasing. This is because both of them use social similarity strategy (SAIS uses the inner product as similarity, Nguyen's Routing uses the number of common interests as similarity) to delivery message. As time goes by, the social phenomena becomes increasingly clear, which makes the algorithms more and more suitable for the social network, i.e., just fewer copies can handle the data forwarding. Fig. 4(c) and Fig. 5(c) show that the delays of those algorithms all increase with TTL increasing. SAIS performs better than the three comparison algorithms. Due to the help of space-crossing communities, some long-distance nodes can communicate through short paths across the geographical space, which lead to the low delay of SAIS.

The partial results of this work have been presented in the conference paper [33]. We provide the summary of differences and the comparison experiments in technical report [59, Appendices K and L].

VI. Related Work

There exist many classical community detection algorithms [11], [34]–[48] and interesting data forwarding algorithms [7], [8], [10], [11], [13], [14], [32], [49]–[53] in mobile social networks or under the hybrid communication networks. Due to space limits, we refer them in technical report [59, Appendix A].

VII. Conclusion

In this paper, we study a more realistic communication framework for MSNs, i.e., the hybrid underlying network with access point (AP) support undertakes the work of data forwarding and the base stations help doing auxiliary global work. Due to the help of APs, the AP connectivity motivates a new concept of space-crossing community. We propose a space-crossing community detection method and describe a high efficient data forwarding scheme SAIS based on it.

In the future, we will put emphasis on studying/measuring the communication overhead in real scenario under the circumstance of frequent community update and the large scale networks. At the same time, the forwarding willingness (including the trust and incentive mechanism) will be considered in the future work.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive comments.

References

[1] N. Kayastha, D. Niyato, P. Wang, and E. Hossain, "Applications, architectures, and protocol design issues for mobile social networks: A survey," in *Proc. IEEE*, vol. 99, no. 12, pp. 2130–2158, Dec. 2011.

[2] "Cisco visual networking index: Global mobile data traffic forecast update, 2013–2018," Cisco, San Jose, CA, USA, Feb. 5, 2014. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/serviceprovider/visual-networking-index-vni/white paper c11-520862.html

[3] M. Fredrikson *et al.*, "Privacy in pharmacogenetics: An end-to-end case study of personalized warfarin dosing," in *Proc. USENIX Assoc. USENIX Security*, 2014, pp. 17–32.

[4] M. Göotz, S. Nath, and J. Gehrke, "Maskit: Privately releasing user context streams for personalized mobile applications," in *Proc. ACM SIGMOD*, 2012, pp. 289–300.

[5] X. Chen, X. Wu, X.-Y. Li, Y. He, and Y. Liu, "Privacy-preserving high-quality map generation with participatory sensing," in *Proc. IEEE INFOCOM*, 2014, pp. 2310–2318.

[6] Y. Agarwal and M. Hall, "Protectmyprivacy: Detecting and mitigating privacy leaks on ios devices using crowd sourcing," in *Proc. ACM MobiSys*, 2013, pp. 97–110.

[7] P. Hui, J. Crowcroft, and E. Yoneki, "Bubble rap: Social-based forwarding in delay tolerant networks," in *Proc. ACM MobiHoc*, 2008, pp. 241–250.

[8] W. Gao, Q. Li, B. Zhao, and G. Cao, "Multicasting in delay tolerant networks: A social network perspective," in *Proc. ACM MobiHoc*, 2009, pp. 299–308.

[9] Y. Ren, J. Yang, M. C. Chuah, and Y. Chen, "Mobile phone enabled social community extraction for controlling of disease propagation in healthcare," in *Proc. IEEE MASS*, 2011, pp. 646–651.

[10] J. Fan *et al.*, "Geo-community-based broadcasting for data dissemination in mobile social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 734–743, Apr. 2012.

[11] N. P. Nguyen, T. N. Dinh, S. Tokala, and M. T. Thai, "Overlapping communities in dynamic networks: Their detection and moibile applications," in *Proc. ACM MobiCom*, 2011, pp. 85–96.

[12] Z. Li, C. Wang, S. Yang, C. Jiang, and X. Li, "Lass: Local-activity and social-similarity based data forwarding in mobile social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 1, pp. 174–184, Jan. 2015.

[13] J. Jeong, S. Guo, Y. Gu, T. He, and D. Du, "TBD: Trajectory-based data forwarding for light-traffic vehicular networks," in *Proc. IEEE ICDCS*, 2009, pp. 231–238.

[14] J. Jeong, S. Guo, Y. Gu, T. He, and D. H. Du, "TSF: Trajectory-based statistical forwarding for infrastructure-to-vehicle data delivery in vehicular networks," in *Proc. IEEE ICDCS*, 2010, pp. 557–566.

[15] A. Aijaz, H. Aghvami, and M. Amani, "A survey on mobile data offloading: Technical and business perspectives," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 104–112, Apr. 2013.

[16] M. Fidan and S. Thrasyvoulos, "Is it worth to be patient? Analysis and optimization of delayed mobile data offloading," in *Proc. IEEE INFOCOM*, 2014, pp. 2364–2372.

[17] N. Cheng, N. Lu, N. Zhang, X. S. Shen, and J. W. Mark, "Vehicular wifi offloading: Challenges and solutions," *Veh. Commun.*, vol. 1, no. 1, pp. 13–21, Jan. 2014.

[18] M. Girvan and M. E. Newman, "Community structure in social and biological networks," *Proc. Nat. Academy Sci.*, vol. 99, no. 12, pp. 7821–7826, 2002.

[19] M. Porter, J. Onnela, and P. Mucha, "Communities in networks," *Notices AMS*, vol. 56, no. 9, pp. 1082–1097, 2009.

[20] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, no. 3–5, pp. 75–174, 2010.

[21] C. Wang *et al.*, "Modeling data dissemination in online social networks: A geographical perspective on bounding network traffic load," in *Proc. ACM MobiHoc*, 2014, pp. 53–62.

[22] J. Fan *et al.*, "Geocommunity-based broadcasting for data dissemination in mobile social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 4, pp. 734–743, Apr. 2013.

[23] T. Hossmann, T. Spyropoulos, and F. Legendre, "Know thy neighbor: Towards optimal mapping of contacts to social graphs for dtn routing," in *Proc. IEEE INFOCOM*, 2010, pp. 866–874.

[24] P. Jaccard, *Distribution de la Flore Alpine: Dans le Bassin des dranses et dans quelques régions voisines*. Surrey, U.K.: Eau Rouge, 1901.

[25] F. Papadopoulos, M. Kitsak, M. Á. Serrano, M. Boguñá, and D. Krioukov, "Popularity versus similarity in growing networks," *Nature*, vol. 489, pp. 537–540, Sep. 2012.

[26] H. J. Ahn, "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem," *Inf. Sci.*, vol. 178, no. 1, pp. 37–51, Jan. 2008.
[27] J. Wu and Y. Wang, "Social feature-based multi-path routing in delay tolerant networks," in *Proc. IEEE INFOCOM*, 2012, pp. 1368–1376.
[28] G. Guo, J. Zhang, and N. Yorke-Smith, "A novel bayesian similarity measure for recommender systems," in *Proc. AAAI Press IJCAI*, 2013, pp. 2619–2625.
[29] H. Ma, I. King, and M. R. Lyu, "Effective missing data prediction for collaborative filtering," in *Proc. ACM SIGIR*, 2007, pp. 39–46.
[30] N. Eagle, A. Pentland, and D. Lazer, "Inferring friendship network structure by using mobile phone data," *Proc. Nat. Academy Sci.*, vol. 106, no. 36, pp. 15 274–15 278, 2009.
[31] K. Nahrstedt and L. Vu, "CRAWDAD data set uiuc/uim (v. 2012-01-24)," Jan. 2012. [Online]. Available: http://crawdad.cs.dartmouth.edu/uiuc/uim
[32] E. Daly and M. Haahr, "Social network analysis for routing in disconnected delay-tolerant manets," in *Proc. ACM MobiHoc*, 2007, pp. 32–40.
[33] Z. Li, C. Wang, S. Yang, C. Jiang, and I. Stojmenovic, "Improving data forwarding in mobile social networks with infrastructure support: A space-crossing community approach," in *Proc. IEEE INFOCOM*, 2014, pp. 1941–1949.
[34] M. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, 2004, Art. ID. 026113.
[35] M. Newman, "Analysis of weighted networks," *Phys. Rev. E*, vol. 70, no. 5, 2004, Art. ID. 056131.
[36] E. Leicht and M. Newman, "Community structure in directed networks," *Phys. Rev. Lett.*, vol. 100, no. 11, 2008, Art. ID. 118703.
[37] A. Clauset, M. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, no. 6, 2004, Art. ID. 066111.
[38] R. Guimera and L. Amaral, "Functional cartography of complex metabolic networks," *Nature*, vol. 433, no. 7028, pp. 895–900, 2005.
[39] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Statist. Mech.: Theory Exp.*, vol. 2008, 2008, Art. ID. P10008.
[40] M. Rosvall and C. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Academy Sci.*, vol. 105, no. 4, 2008, Art. ID. 1118.
[41] G. Palla, I. Derenyi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, Jun. 2005.
[42] P. Hui, E. Yoneki, S. Y. Chan, and J. Crowcroft, "Distributed community detection in delay tolerant networks," in *Proc. ACM MobiArch*, 2007, pp. 1–10.
[43] Y.-R. Lin, Y. Chi, S. Zhu, H. Sundaram, and B. L. Tseng, "Analyzing communities and their evolutions in dynamic social networks," *ACM Trans. Knowl. Discovery Data (TKDD)*, vol. 3, no. 2, p. 8, 2009.
[44] M.-S. Kim and J. Han, "A particle-and-density based evolutionary clustering method for dynamic networks," *Proc. VLDB Endowment*, vol. 2, no. 1, pp. 622–633, 2009.
[45] N. P. Nguyen, T. N. Dinh, Y. Xuan, and M. T. Thai, "Adaptive algorithms for detecting community structure in dynamic social networks," in *Proc. IEEE INFOCOM*, 2011, pp. 2282–2290.
[46] D. Duan, Y. Li, Y. Jin, and Z. Lu, "Community mining on dynamic weighted directed graphs," in *Proc. ACM CNIKM*, 2009, pp. 11–18.
[47] S. Bansal, S. Bhowmick, and P. Paymal, "Fast community detection for dynamic complex networks," in *Complex Netw.*, vol. 116, pp. 196–207, 2011.
[48] R. Cazabet, F. Amblard, and C. Hanachi, "Detection of overlapping communities in dynamical social networks," in *Proc. IEEE SocialCom*, 2010, pp. 309–314.
[49] J. Wu, M. Xiao, and L. Huang, "Homing spread: Community home-based multi-copy routing in mobile social networks," in *Proc. IEEE INFOCOM*, 2013, pp. 2319–2327.
[50] H. Zhu *et al.*, "Zoom: Scaling the mobility for fast opportunistic forwarding in vehicular networks," in *Proc. IEEE INFOCOM*, 2013, pp. 2832–2840.
[51] Y. Wu, Y. Zhu, and B. Li, "Infrastructure-assisted routing in vehicular networks," in *Proc. IEEE INFOCOM*, 2011, pp. 1485–1493.
[52] Z. Ying, C. Zhang, and Y. Wang, "Social based throwbox placement in large-scale throwbox-assisted delay tolerant networks," in *Proc. IEEE ICC*, 2014, pp. 2472–2477.
[53] N. Banerjee, M. D. Corner, and B. N. Levine, "An energy-efficient architecture for dtn throwboxes," in *Proc. IEEE INFOCOM*, 2007, pp. 776–784.
[54] A. Lancichinetti and S. Fortunato, "Community detection algorithms: A comparative analysis," *Phys. Rev. E*, vol. 80, no. 5, 2009, Art. ID. 056117.
[55] A. Khadivi, A. Rad, and M. Hasler, "Network community-detection enhancement by proper weighting," *Phys. Rev. E*, vol. 83, no. 4, Apr. 2011, Art. ID. 046104.
[56] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," *New J. Phys.*, vol. 11, 2009, Art. ID. 033015.
[57] M. T. Thai and P. M. Pardalos, *Handbook of Optimization in Complex Networks*. Berlin, Germany: Springer-Verlag, 2012.
[58] A. Keränen, J. Ott, and T. Kärkkäinen, "The one simulator for DTN protocol evaluation," in *Proc. ICST SIMUTools*, 2009, pp. 1–10.
[59] Z. Li, C. Wang, S. Yang, C. Jiang, and I. Stojmenovic, "Space-crossing: community-based data forwarding in mobile social networks under the hybrid communication architecture," Tech. Rep. [Online]. Available: https://www.dropbox.com/s/qte54kunahb107q/space-crossing-20150419.pdf?dl=0

Zhong Li received the Ph.D. degree from the Department of Computer Science, Tongji University, in 2015. Her research interests include wireless communication, social network analysis and distributed computing.

Cheng Wang received the Ph.D. degree from the Department of Computer Science, Tongji University, in 2011. Currently, he is a Research Professor at Tongji University. His research interests include wireless networking, mobile social networks, and mobile cloud computing.

Siqian Yang received the B.S. degree from the Department of Computer Science and Technology, Tongji University, in 2012. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Tongji University, Shanghai, China. His research interests include social network analysis, delay tolerant networking and distributed computing.

Changjun Jiang received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1995. Currently, he is a Professor with the Department of Computer Science and Engineering, Tongji University, Shanghai. His current areas of research are concurrent theory, Petri net, and formal verification of software, wireless networks, concurrency processing and intelligent transportation systems.

Ivan Stojmenovic (AM'88–M'04–SM'05–F'08) received the Ph.D. degree in mathematics. He is a Full Professor at the University of Ottawa, Canada. He held or holds regular and visiting positions in Saudi Arabia (Distinguished Adjunct Professor at the King Abdulaziz University, Jeddah), China (Tsinghua University, DUT, Beihang), Serbia, Japan, the USA Canada, France, Mexico, Spain, United Kingdom (as Chair in applied computing at the University of Birmingham), Hong Kong, Brazil, Taiwan, and Australia. He published over 300 different papers, and edited seven books on wireless, ad hoc, sensor and actuator networks and applied algorithms with Wiley.