

# Community Detection in the Social Internet of Things Based on Movement, Preference and Social Similarity

A. Meena KOWSHALYA , M. L. VALARMATHI

Department of CSE, Government College of Technology, Coimbatore, India.  
meenakowshalya.gct@gmail.com

**Abstract:** Internet of Things (IoT) is one paradigm many visions technology. One of the many visions of *Internet of Things* is to make *Things* sociable. This is achieved by integrating *IoT* and Social networking which may lead to a new paradigm called Social Internet of Things (SIoT). *SIoT* is defined as collection of intelligent objects that can autonomously interact with its peers via owners. In a *SIoT* scenario, detecting and characterizing a network structure is very important. In this paper, we propose a new community detection algorithm that detects communities in *SIoT* using three metrics namely social similarity, preference similarity and movement similarity. To the best of our knowledge this is the first work that detects communities in large scale Social Internet of Things using social, preference and movement similarity. The experimental results show that the proposed community detection scheme achieves higher quality results in terms of detection rate and execution time when compared to existing methods.

**Keywords:** Social Internet of Things (SIoT), Internet of Things (IoT), Community Detection, Preference Similarity, Social Similarity, Movement Similarity

## 1. Introduction

The world around us is composed of electrical and electronic gadgets such as sensors, actuators, RFIDs, etc., collaborating with humans and things. These objects have become a part of our fabric. Social Internet of Things is turning to be a pioneer successful paradigm for collaboration among peer communities. SIoT is composed of objects that are not only smarter but also socially conscious [12]. Relationship between smart objects can be classified into four types [2]. Parental object relationship, Co-work / Co-location object relationship, Ownership object relationship and Social object relationship. Parental object relationship is established when objects of the same manufacturer tend to collaborate with each other. Co-work or Co-location object relationship are established when objects meet each other at work place of owners or if an owner moves to a different location, their objects interact with other(s) object in that location. Ownership object relationship is defined as a relationship established by objects belonging to the same owner. Social object relationship is established when owners engage in social networking activities and objects tend to interact socially via owners. Many assumptions can be made from the concept of smart objects and their relationships. There must exist community of objects belonging to owners having common interest. Such objects tend to meet each other frequently leading to social similarity. Such objects exhibit similarity

in behavioral pattern called preference similarity. Also these objects tend to move to similar places of interests exhibiting movement similarity. This paper takes into account movement, preference and social similarities to detect communities across SIoT environments. The rest of the paper is organized as follows. Section 2 presents the state of art in community detection of large scale social networks. Section 3 describes how movement, preference and social similarity is constructed. Section 4 presents the experimental results followed by Conclusion and future work.

## 2. Related Work

### 2.1 Social similarity based community detection

The authors in [10] proposed an agglomerative hierarchical clustering algorithm based on similarity sequence. A sequence was constructed by sorting the similarity in descending order. Using agglomerative clustering optimal community structure was obtained. The authors in [8] uses long term social history and relationship to detect communities distributively. [13] discusses a protocol called ICast which chases users interests based on social communities formed. It uses weak ties between nodes that rarely happen and has proposed a new idea of weak ties connecting different communities. The authors in [6] proposes an opportunistic IoT where movement location information and

contact of humans are used for information sharing and dissemination. A bidirectional nature between humans and opportunistic IoT were identified to exist. [11] proposes a Community Detection in an Integrated IoT and SN (CDIISN) in which nodes in social networks are converted into basic nodes and IoT nodes. Community detection is done for these two nodes separately. The authors in [7] has detected communities in Social Internet of Things using Ant Colony Optimization technique which considered many social relationship and properties between objects.

## 2.2 Preference similarity based community detection

[14] proposes a peer to peer content driven communication in mobile social networks where end to end connectivity is not possible. The proposed protocol takes into account the information about user's interests, content based data storing and forwarding, host mobility in a disconnected, delay tolerant MANET for preference extraction. [4] proposes a community detection algorithm based on extracting the nodes interests from a file. The algorithm groups common interest's nodes into one community which often meets each other. [9] has proposed a dynamic resource discovery based on preference and movement similarity through which efficient resource discovery was possible from constructed sub communities. [3] proposes a location recommender system consisting of offline modelling and online recommendation. The offline modelling models each individual's personal preferences with a weighted category hierarchy and infers the expertise of each user. The online recommendation part selects candidate local experts in a geospatial range that matches the user's preferences using a preference-aware candidate selection algorithm. The score of the candidate locations are calculated based on the opinions of the selected local experts

## 2.3 Movement based similarity

Very little work exists to detect communities using movement similarity. The authors in [1] uses distance and interconnection factors of mobile nodes to quantize the social relationships. Subgroups are build based on social awareness and movements of nodes. This works well even for sparse networks. [5] uses Tabu search scheme and local search scheme to

implement a movement pattern aware delay tolerant optimal routing. The scheme outputs optimal sets of nodes through efficient movement based search techniques. [15] uses the movement pattern of nodes that are extracted from the movement records of nodes. Commonly visited locations for the relay nodes and the destination node are thus obtained.

## 3. Proposed Model

The IoT layers were modified to suit SIoT components. The proposed model has 4 layers namely base layer, virtual network layer, global network layer and application layer as shown in Figure 1. In the base layer every region is divided into individual units for node identification. The nodes with similar preferences show similar trajectory movements. These patterns together with social interests form the virtual network layer. All these patterns are tied together and merged to form a global network layer. The global network layer is used for resource discovery and service search. The application layer is not modified and is used as such.

### 3.1 Trajectories of objects

The GPS (Global Positioning System) provides current location information of objects. The position information is represented as  $P(x_i, y_i, z_i, t_i)$ .  $P(x_i, y_i, z_i)$  are three dimensional coordinates of node  $i$  and  $t_i$  represent timestamp of node  $i$  at position  $P$ . At any time the user will reside in a particular position over a period of time. The average threshold used in the implementation is 16 minutes. Let  $P(x_i, y_i, z_i, t_i)$  be the position of node  $i$  at time  $t_i$  and  $P(x_j, y_j, z_j, t_j)$  be the position of node  $j$  at time  $t_j$ . Then the visit of the user id

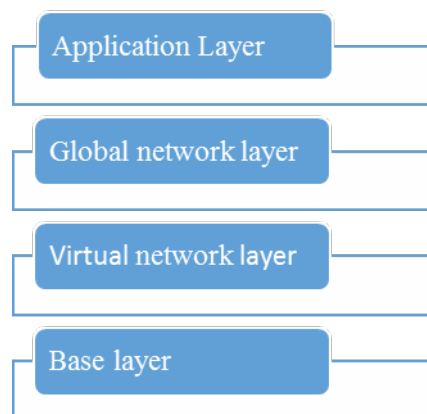


Figure 1. Layers in the proposed model

$T_{ij} = (t_j - t_i)$ , when  $t_j - t_i$  greater than  $t$  the mobile users are in the same region. The motion trajectories of nodes along with their timestamp is also required.

Let  $T_i$  be the vector used to represent the motion trajectories of a node  $i$ . The vector is given as  $T_i = (A_1(x_1, y_1, z_1, \text{timestamp}, \alpha_1), (A_2(x_2, y_2, z_2, \text{timestamp}, \alpha_2), \dots, (A_n(x_n, y_n, z_n, \text{timestamp}, \alpha_n)$ . Timestamp has the information about location, start time and end time.

Timestamp =  $(t_{\text{start}}, t_{\text{end}}, P_i)$

where  $t_{\text{start}}$  is the start time when node moves into the region  $A_i$  and  $t_{\text{end}}$  is when it comes out of  $A_i$ . The weight  $\alpha$  is the average difference of start time and end time of the entire list of records  $\alpha$  is calculated as

$$\alpha_i = \frac{\sum_{i=0}^{m-1} \text{list}(i).t_{\text{end}} - \text{list}(i).t_{\text{start}}}{t_{\text{total}}} \quad (1)$$

where  $m$  is the length of the list and  $t_{\text{total}}$  is the total time of processing the entire list.

### 3.2 Preference similarity

The parameters used to extract preferences mainly are resources and services used by objects. For example, an object would have used services of its neighbors, shared a nearby printer, etc. Utilizing common keywords, common resources have been extracted and are denoted as  $R_r$  and  $R_s$ . Similarity between  $R_r$  and  $R_s$  is calculated as

$$\text{sim}(r, s) = \frac{\sum_{i=1}^m w_{ri} * w_{si}}{\sqrt{\sum_{i=1}^m w_{ri}^2} * \sqrt{\sum_{i=1}^m w_{si}^2}} \quad (2)$$

Where  $m$  is the total number of common keywords,  $w_{ri}$  and  $w_{si}$  are the weights of the  $i^{\text{th}}$  common keyword in the two resources respectively. Considering similarities of all resources a global matrix is formed. The elements of the similarity matrix can be sorted to generate a list of similar resources. The top  $k$  similar resources can be found using equation (3):

$$k = \min \frac{\sum_{i=1}^k \text{sim}_{(r,s)}^i}{\sum_{i=1}^k \text{sim}_{(r,s)}^i} - \delta, \quad 0 < \delta < 1, 0 < k \leq 1 \quad (3)$$

where  $1$  is the length of the list and  $\delta$  is a user specific parameter to determine top  $k$  resources. For the top  $k$  resources the preference vector becomes,

$$w_{l_j} = \sum_{i=1}^l w_{R_i}^{l_j} / k \quad (4)$$

where  $R_i = \{(e_1, w_1); (e_2, w_2) \dots (e_n, w_n)\}$ ,  $e_i$  refers to the  $i^{\text{th}}$  keyword and  $w_i$  refers to the weight of the  $i^{\text{th}}$  keyword,  $w_{R_i}^{l_j}$  is the weight of the keyword  $l_j$  in resource  $R_i$ . Then the preference vector of the node  $R_n^p$  is

$$R_n^p = (l_j, w_{l_j}), \quad j > 0$$

### 3.3 Social connectivity

#### 3.3.1 Cooperativeness

Cooperativeness represents whether or not the trustee node is socially cooperative with the trustor. It's assumed that nodes with common friends are cooperative and behaves differently with others. In a SIoT environment, nodes cooperativeness can be predicted by its social ties. Socially cooperative nodes improve the application performance. Each device/object possesses a list of friends likely to be cooperative. This list will be updated by owners periodically. According to [12] the  $CO_{ij}^{\text{cooperativeness}}(t)$  at time  $t$  is calculated as

$$CO_{ij}^{\text{cooperativeness}} = \frac{\text{friends}(i) \cap \text{friends}(j)}{\text{friends}(i) \cup \text{friends}(j)} \quad (5)$$

#### 3.3.2 Community interest

Community interest as proposed by [12] is another factor that enables communication between objects of communal interest. The objects are classified according to their parental relationships, Co-work or Co-location relationships, social object relationships and owner object relationships. Objects with the same community interest are supposed to interact with each other very often leading to increased application performance. The  $CI_{ij}^{\text{community interest}}(t)$  at time  $t$  is calculated according to equation (6).

$$CI_{ij}^{\text{community interest}} = \frac{\text{community}(i) \cap \text{community}(j)}{\text{community}(i) \cup \text{community}(j)} \quad (6)$$

### 3.4 Construction of communities

The preference similarity, movement similarity and social similarity are used to construct sub

communities. The preference and movement similarity is constructed based on discussions from [9]. The authors in [9] derived preference and movement similarity for efficient resource discovery in Social Internet of Things.

### 3.4.1 Preference Similarity

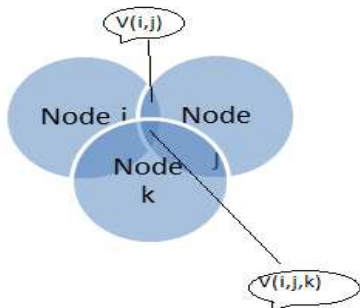
The preference similarity  $Sim_{ij}^p$  is calculated as

$$Sim_{ij}^p = \frac{\sum_{e=1}^m w_{ie} * w_{je}}{\sqrt{\sum_{e=1}^m w_{ie}^2} * \sqrt{\sum_{e=1}^m w_{je}^2}} \quad (7)$$

Where  $m$  is the total number of common keywords of the preference vector  $Sim_{ij}^p$ ,  $w_{ie}$  and  $w_{je}$  are the weights of the  $k^{th}$  common keyword of the preference vector  $Sim_{ij}^p$  and  $Sim_{ij}^q$ .

### 3.4.2 Movement based similarity

Two or more nodes can be in the same region at the same time. If a region is considered as a small square unit, then node  $i$  and node  $j$  at time  $t_i$  and  $t_j$  may reside in the same region. Figure 2 shows overlap between nodes  $i$ , node  $j$  and node  $k$ .



**Figure 2.** Overlap of nodes (stay region) at the same time

The movement similarity between two nodes is calculated as

$$Sim_{ij}^m = \frac{3 \sum_{i=1}^n \sum_{j=1}^n V(i, j) \alpha_i * \alpha_j}{4 n \pi r^3 \sqrt{\sum_{i=1}^n \alpha_i^2} * \sqrt{\sum_{j=1}^n \alpha_j^2}} \quad (8)$$

where  $V(i, j)$  is the overlap (stay) region time of nodes  $i$  and  $j$ .  $\alpha_i$  and  $\alpha_j$  are the weights of the stay regions.

### 3.4.3 Social similarity

The social similarity between two nodes  $i$  and  $j$  is calculated as

$$Sim_{ij}^s = \frac{\sum_{e=1}^m CO_{ij} * CI_{ij}}{\sqrt{\sum_{e=1}^m CO_{ij}^2} * \sqrt{\sum_{e=1}^m CI_{ij}^2}} \quad (9)$$

where  $CO_{ij}$  represents the cooperativeness of two nodes  $i$  and  $j$ ,  $CI_{ij}$  represents the community interest of  $i$  and  $j$ ,  $m$  is the total number of common keywords. Thus the total similarity is the sum of movement, social and preference similarity and is as shown

$$Sim_{tot} = \alpha Sim_{ij}^p + \beta Sim_{ij}^m + \delta Sim_{ij}^s \quad (10)$$

$$\alpha + \beta + \delta = 1$$

where  $Sim_{ij}^p$  is the preference similarity vector,  $Sim_{ij}^m$  is the movement similarity vector,  $Sim_{ij}^s$  is the social similarity vector,  $\alpha, \beta, \gamma$  are the weights of the vectors. The values of the weights set in the implementation are  $\alpha = 0.4, \beta = 0.3$  and  $\gamma = 0.3$ .  $T_{Sim}$  is a predefined threshold set as 0.3 in the implementation. At start, a node broadcasts hello packets to all direct neighbors. The neighboring nodes calculate their  $Sim_{tot}$  of node  $i$ . If  $Sim_{tot}$  is greater than the threshold value, node  $i$  is added to the sub community. Every node saves its top  $k$  neighbors. It is evident that if  $T_{Sim} = 0$ , then there exists only one community, ie the Social Internet of Things network itself (size is the same as global community).

#### Algorithm: Community Detection

```

Input: Preference, social and movement
similarity vectors of node  $n_i$ 
Output: Subcommunity lists
For all neighbors nb of  $n_i$ 
    Send broadcast message fro nb to  $n_i$ 
    The total similarity  $Sim_{tot}$  is computed by all
    nb nodes
    If  $Sim_{tot} > T_{Sim}$ 
        Then
            add  $n_i$  into the community
            Send join message to  $n_i$ 
    If message accepted
        Then
            add  $n_i$  to the community
    return community list
end

```

## 4. Experimental Results

Experiments were conducted using ns3 and SocNetV1.9, a social network tool. For experimental purpose, datasets were collected from Crowdad, a large online repository of data store. The dataset was used by 76 persons during SIGCOMM 2009 conference in Barcelona, Spain. A simple application was developed for the same purpose. The data sets include traces of Bluetooth proximity, opportunistic message creation and dissemination and the social profiles of the participants. 100 smart phones were distributed to all participants during the first two days of the conference. Each device consists of social profile of the participant that included some basic information such as home, city, country and affiliation. Each participant was asked to log on to their Facebook profile in order to include the list of Facebook friends and interests in the social profile. The participants were allowed to edit the social profile before it was uploaded on the device and recorded in our traces. The final trace contains data from all 76 devices that shows activity's of all participants during the experiment.

The Community detection algorithm was run on the trace producing four sub communities according to their movement pattern, social similarity and preferences. The social internet of things networks is as shown in Figure 3. The detected community 1 comprises of 12 nodes, sub community 2 with 16 nodes, sub community 3 with 29 nodes and sub community 4 with 19 nodes. Figure 4 shows the result of the proposed community detection algorithm. The proposed algorithm is better in terms of detection accuracy and execution time compared to SPOON[4], LOC [15] and RDPM [9]. The algorithm was run initially for the graph shown in Figure 3. At 10.14 seconds the algorithm extracted top  $k$  similarity preference vectors consisting of keyword "devices", "print", "new", Facebook", "activity", "interests". At 13 seconds approximately the algorithm listed the user ID and group number of nodes based on the interaction. The movement of nodes was sparse and at 25.34 seconds the algorithm listed movements of nodes. Few interaction with outside nodes were also seen. For keeping things simple, outside interaction were omitted. Only interaction between 76 users was taken into account.

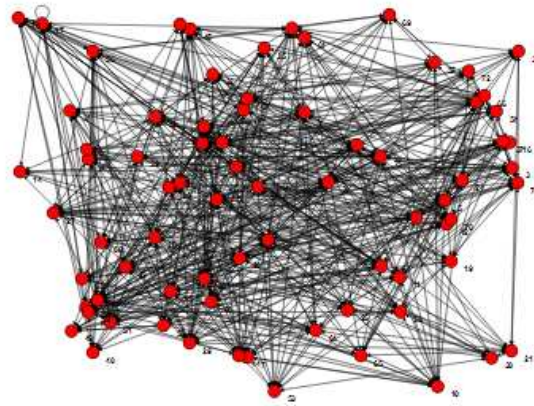
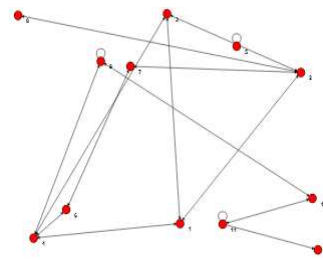
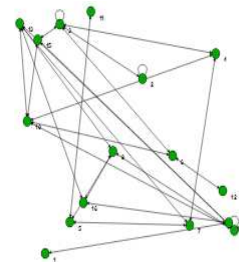


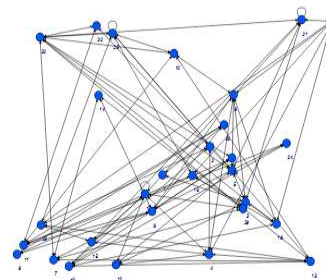
Figure 3. Layout of the network



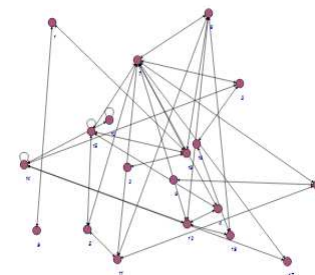
a)



b)

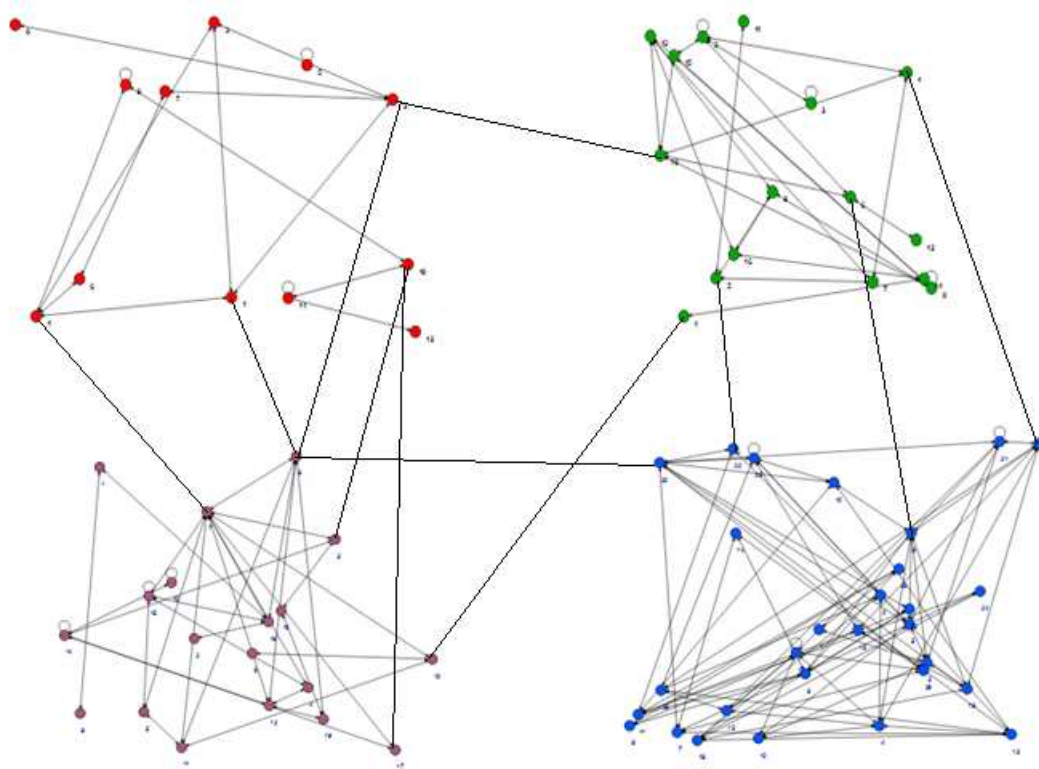


c)



d)

Figure 4. Detected sub-communities



**Figure 5.** Interaction between intra and inter sub - communities

Finally the algorithm showed 4 sub communities based on social ties, preferences and movement similarity. Figure 4 shows four sub-communities formed based on movement, preference and social relationship between nodes. To the extent possible, the interactions between intra and inter communities were also studied as shown in Figure 5.

The proposed methods was compred with RDPM [9], LOC [15] and SPOON [4].

Simulation area of 1000 \* 1000 m was used. The total time of simulation was 9 hours. The proposed method has the fastest execution time compared to the existing methods and detected 4 communities. RDPM [9], LOC [15] and SPOON [4] detected only 3 sub communities. This proves that the proposed work is much better in terms of detection accuracy and execution time. Table 2 compares the proposed method with existing schemes. This is the 1<sup>st</sup> work that utilizes social similarity, preference

**Table 1.** Execution time and number of Communities detected

S.No	Scheme	Execution time in hours	Communities detected
1	RDPM [9]	12	3
2	LOC [15]	10 (only for community formation)	3
3	SPOON [4]	28 (only for community formation)	3
4	Proposed method	09	4

**Table 2.** Comparison of the proposed method with existing schemes

S.No	Scheme	Social Relationships	Movement Patterns	Preference Similarity
1	RDPM [9]		✓	✓
2	LOC [15]		✓	
3	SPOON [4]			✓
4	Proposed method	✓	✓	✓

similarity and movement similarity for community detection.

## 5. Conclusion

The Social Interest of Things has not only paved way for smarter environments but also to a socially conscious environment. Out of the many challenges in SIoT, detecting communities is of prime importance since things belonging to the same community tend to cooperate with each other. This paper has proposed a new community detection methodology which takes into account three parameters namely movement similarity, preference similarity and social similarity for detecting communities across large scale SIoT environments. The movement based similarity was determined from the trajectory locations of nodes and the overlap stay regions of nodes. The preferences were extracted from the preference vectors that collect the similarity patterns of resources and services used by nodes. The social similarity was found from community interest and cooperativeness of the nodes. The proposed method detected 4 communities from the real time dataset. The proposed method outperforms the existing community detection schemes in terms of execution time and detection accuracy.

## REFERENCES

1. AN, J., X. GUI, W. ZHANG, J. JIANG, **Nodes Social Relations Cognition for Mobility-aware in the Internet of Things**, In Internet of Things (iThings/CPSCom), 2011 International Conference on and 4th International Conference on Cyber, Physical and Social Computing 2011 Oct 19, pp. 687-691.
2. ATZORI, L., A. IERA, G. MORABITO, **Siot: Giving a Social Structure to the Internet of Things**, IEEE Communications Letters. 2011 Nov; vol. 15(11), pp. 1193-5.
3. BAO, J., Y. ZHENG, M. F. MOKBEL, **Location-based and Preference-aware Recommendation using Sparse Geo-social Networking Data**, In Proceedings of the 20th International Conference on Advances in Geographic Information Systems 2012 Nov. 6, pp. 199-208.
4. CHEN, K., H. SHEN, H. ZHANG, **Leveraging Social Networks for P2P Content-based File Sharing in Disconnected MANETs**, IEEE Trans. on Mobile Computing. 2014 Feb; vol. 13(2), pp. 235-49.
5. GONZALEZ, M. C., C. A. HIDALGO, A. L. BARABASI, **Understanding Individual Human Mobility Patterns**, Nature, vol. 453(7196), 2008, pp. 779-82.
6. GUO, B., Z. YU, X. ZHOU, D. ZHANG, **Opportunistic IoT: Exploring the Social Side of the Internet of Things**, In Computer Supported Cooperative Work in Design (CSCWD), 2012 IEEE 16th International Conference on 2012 May 23 pp. 925-929.
7. KOWSHALYA, A. M., M. L. VALARMATHI, **Detection of Sybil's across Communities over Social Internet of Things**, Journal of Applied Engineering Science (Istrazivanja i projektovanja za privredu) 14.1 (2016).
8. LI, F., J. WU, **MOPS: Providing Content-based Service in Disruption-tolerant Networks**, In Distributed Computing Systems, 2009. ICDCS'09. 29th IEEE International Conference on 2009 Jun 22 (pp. 526-533). IEEE.
9. LI, Z., R. CHEN, L. LIU, G. MIN, **Dynamic Resource Discovery based on Preference and Movement Pattern Similarity for Large-Scale Social Internet-of-Things**, accepted for publication in IEEE IoT 2015.
10. LU, H., Q. ZHAO, Z. GAN, **A Community Detection Algorithm Based on the Similarity Sequence**, International Conference on Web Information Systems Engineering 2014 Oct 12 (pp. 63-78), Springer International Publishing.
11. MISRA, S., R. BARTHWAL, M. S. OBAIDAT, **Community Detection in an Integrated Internet of Things and Social Network Architecture**, In Global Communications Conference (GLOBECOM), 2012 IEEE 2012 Dec 3 pp. 1647-1652, IEEE.
12. NITTI, M., R. GIRAU, L. ATZORI, **Trustworthiness Management in the Social Internet of Things**, IEEE Transactions on Knowledge and Data Engineering. vol. 26(5), 2014, pp. 1253-66.

13. PAGANI, E., L. VALERIO, G. P. ROSSI, **Weak Social Ties Improve Content Delivery in Behavior-aware Opportunistic Networks**, Ad Hoc Networks, vol. 25, 2015, pp. 314-29.
14. QURESHI, B., G. MIN, D. KOUVATSOS, M. ILYAS, **An Adaptive Content Sharing Protocol for P2P Mobile Social Networks**, In Advanced Information Networking and Applications Workshops (WAINA), 2010 IEEE 24th International Conference on 2010 Apr 20, pp. 413-418.
15. YOU, L., J. LI, C. WEI, L. HU, **MPAR: A Movement Pattern-aware Optimal Routing for Social Delay Tolerant Networks**, Ad Hoc Networks. 2015 Jan 31, vol. 24, pp. 228-49.