http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (15): 3103-3109, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3103.3109

A User-interest-based Routing Optimization Algorithm for Data Dissemination in Mobile Social Networks

Minggang Zhai[1], Ming Zhao[1,*], Liping Liu[1], Na Wei[2]
[1]School of Software, Central South University, 410083, Changsha, China
[2]School of Information Science and Engineering, Hunan University, 410082, Changsha, China

Abstract: Data dissemination is an important research subject in Mobile Social Networks (MSNs). This study mainly studies user-interest-based accurate data dissemination based on geographic communities in mobile social networks. Our goal is to establish interestingness measure model of the geo-community, plan superuser's moving route, so that the superuser can disseminate data to the other regular users of the network who are interested in the data carried by the superuser.We have established a geo-community interestingness measure model based on users distribution over geo-communities and establish routing algorithm for the superuser according to the user's interestingness measurement of the geo-community, in order to achieve accurate data dissemination purposes. We give the superuser's routing algorithm. We get good results in the aspect of the number of packets, data transmission delay and energy consumption.

Key words: Mobile social network, accurate data dissemination, user interestingness measure, geography community

INTRODUCTION

In recent years, with the growing popularity of smart mobile devices, users with the same interests can establish a wireless network connection and communicate with each other in the absence of network infrastructure by cooperating with each other. This network is called mobile social network, it can be applied to a variety of absence of end to end connectivity scenes (Ioannidis *et al*., 2009; Han *et al*., 2012). For example, large-scale post-disaster reconstruction, the battlefield and a wide range of sensor networks. In these circumstances network connection is intermittent, network infrastructure is absence and there is connections. Thus the mobile social networks are also viewed as delay tolerant networks (Delay/Disruption Tolerant Networks, DTN) (Han *et al*., 2012; Rao *et al*., 2012). Based on above several features, mobile social networks' intermittent, uncertainty connections make accurate data dissemination facing tremendous challenges (Gao and Cao, 2011).

There are already some studies about the social characteristics of user mobility in mobile social networks. The user nodes usually belong to certain geo-communities and usually connect with the other nodes that have commonalities (such as hobbies, social relations and occupational) with them. Du *et al*. (2011) relying on theoretical model studies a single node's geography moving regularity. They point out in the mobile network user nodes are often active in a few fixed geo-community, rather than moving randomly in the network. For the first time, Fan *et al*. (2010) combine the node's social moving characteristics with the geography features of the node's moving. They find that nodes' sojourn time in the geo-communities follow a power law distribution. They employ a semi-Markov process to model user mobility based on the geo-community structure of the network. And through the modeling of real data sets they get the steady-state distribution the users over geo-community. This laid the foundation for our user-interest-based accurate data dissemination. In addition, We can easily infer that there is a relevance between the user's interest and a certain geo-community. In this study, the data dissemination mechanism is based on limited broadcast, that is, the data is sent from a specific source node (superuser node) to the regular users in the various geo-communities of the mobile social networks who are interested in the data carried by the superuser. As regular user's mobility characteristic known, the superuser's data dissemination routing path depends on the interestingness measure of the geo-communities. Therefore, Establishment of the interestingness measure of the geo-communities over regular user's interest model and proper planning the superuser's moving trajectory is of great significance to improve overall network performance.

Corresponding Author: Ming Zhao, School of Software, Central South University, 410083, Changsha ,China

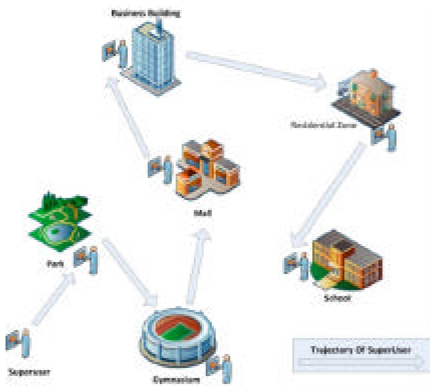

Fig. 1: Geo-Community-based and user interest-based data dissemination illustration

The purpose of the introduction of super-user node is to facilitate network connectivity and faster data transfer speeds. For example, in Message Ferry or Data MULES (Zhao *et al*., 2004, 2005; Shah *et al*., 2003), they use super node and assuming that the super node move along a fixed path to promote connectivity with other nodes in the network. It is designed to reduce energy consumption rather than accelerate data transmission of the network. Wu and Wang (2012) and Wu *et al*. (2013) study a adaptive network loads and the location of the target node ferry moving path planning mechanism, but their research orients the static nodes in the network. Tariq *et al*. (2006) try to study the ferry path planning problem in mobile networks, but they imposed too many artificial restrictions to mobile nodes in the network, which is not suitable for reality scenes. Fan *et al*. (2012) study Geo-Community-Based broadcasting for data dissemination in Mobile Social Networks. In their study, the moving node is modeled as a Semi-Markov model, they plan a super-user node moving path based on trace-driven simulations. It is the first to study data broadcasting in a realistic MSNet setting. However, the study did not consider the wishes of the user's interest in the geo-community and cannot achieve accurate data dissemination.

The main objective of this study is , by counting the user's interest value of the regular node in the geo-community, planning routing trajectory for supernode to achieve accurate and efficient data dissemination. Our study focuses on mobile social networks. Accurate data dissemination can improve network resource utilization efficiency and reduce the waste of the data. How to establish interestingness measure model is the key to this problem, then select geo-communities with a high interest measure value for superuser to distribute data, finally, achieve the target to improve the efficiency and accuracy of the data dissemination. Fig. 1 shows a schematic diagram of our data dissemination.

The novelty and contributions of this study are as follows:

- We propose the concept of accurate data dissemination in mobile social networks
- We establish interestingness measure model of the geo-community. It lays the foundation for our accurate data dissemination
- A simulation based on simulated datasets is made. We propose optimal path algorithm for the superuser and realize the data dissemination process foe the superuser based on the interestingness measure of the geo-communities

The remainder of this study is organized as follows: Section 2 provides an overview of related researches about user's interestingness measure model. Then, we explore both social and geographic regularities of user mobility in simulated traces and further establish interestingness measure model for the geo-communities in Section 3. We propose superuser route schemes in Section 4. Section 5 evaluates the performance of our approach, via simulated trace-driven simulation. Section 6 is the outlook for user-interest-based data dissemination in mobile social networks. And the last section is our conclusions.

OVERVIEW ABOUT THE USER'S INTEREST

Relevant studies on user's interest model:

Access methods of user's interest: There is two main access methods of user's interest: implicit obtain and explicit obtain. Implicit obtain is usually the more common way and it almost does not affect the user's normal work. We can receive the datum of the user's behavior within a period of time by not disturbing the user. After analyze these data, we can know which items the users are interested in and we can also know the interest degree values of the items. Such as server-side mining the user's access log, or applying intelligent Agent to track user's behavior information. Explicit-obtain requires the user to interrupt work for providing feedback. Feedback information includes project's relevance , interestingness for the project and other items. Explicit-obtain can usually obtain information with a higher degree of accuracy. Because the information is not obtained through indirect inference, instead it is provided by the users themselves. For example, the system provides users with a customized form that asks the user to fill demand for personalized information.

In this study, an explicit way to obtain users' interests is applied to ensure the accuracy of the user interest information.

Representation of the user's interest model: Representation of user's interest model mainly includes representation based on vector space model, representation based on neural networks, representation based on the 'user-item' evaluation matrix, representation based on the coarse-grained and fine-grained interestingness, representation based on Ontology and representation based on semantic networks. Among them, representation based on vector space model refers using keywords vectors in the vector space to represent the user model. Vector space model is the most common method to document representation. The basic idea is to use vectors to represent the document, each document is represented as $\{(\pi_1, w_1), (\pi_2, w_2), \ldots (\pi_n, w_n)\}$, where, π_n is then-th keyword, w_n is the weight of the corresponding keyword π_n. Vector space model is currently the most popular user's interest model representation. Keywords in the vector space model may be all of the words which the user is interested in appearing in the document, it can also be a part of feature words selected from the document.

We use feature words to represent the user's interest in this study. Such as 'sports', 'entertainment', 'finance', 'music', etc. To facilitate experiments, we also incorporates the 'user- item' evaluation matrix approach. A matrix $R_{m \times n}$ is used to denote user model, where m is total number of users in the system, n is the number of items (Items in this study is used to represent user interest categories). Each element of the matrix R^k_j denotes user k's interest value to the item j. The range of interest value is [0, 10],where '0' denotes 'No interest in the item', '10' denotes 'Very interested in the item'.

Related research about the geographic regularity of human mobility: Fan *et al*. (2012) proposed geographic community concept in mobile social networks. They pointed out that users are concentrated in different geographic communities and the sojourn time in various geographic communities follows a power law distribution (long-tailed distribution). In their studies, node mobility model is modeled as semi-Markov model. Then they also verified the stability of the model on the time and space prediction.

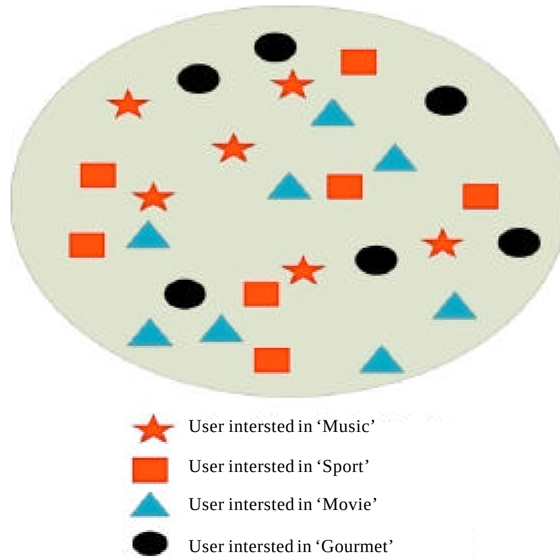

Fig. 2: Users with different interest in a geo-community

INTERESTINGNESS MODEL OF THE GEO-COMMUNITY

Relevance between Geo-community and user's Interest: Obviously, geo-communities and user's interest have a strong correlation in mobile social network. Such as in the gym, gymnasium, swimming pool and other similar geo-communities, users love sports will very intensive. Correspondingly, the entire community about the degree of interest in sport is relatively high. Thus, there is a interest value ε_i for each geo-community on certain data. The range of interestingness values is also [0, 10]. Greater the interest value, relevance between geo-community and the data, the greater contribution to the total interestingness of the geo-community. Suppose there are m geo-communities, then there is an interest value matrix $R_{m \times n}$, where, n denotes the number of items.

Interestingness of All Users in the Geo-comm-unity: In Fig. 2, there are many users with various interest types. It has different interestingness with the data carried by the superuser.

We use σ_i denotes value of all the user's interest in the current geo-community. Namely:

$$\sigma_i = \sum_{k=1}^{N} (R^k_j \times U^k_i) \quad (1)$$

where, U^k_i denotes the steady-state contact probability per unit time between geo-Community i and $User_k$, R^k_j denotes the interestingness that $User_k$ to the item j.

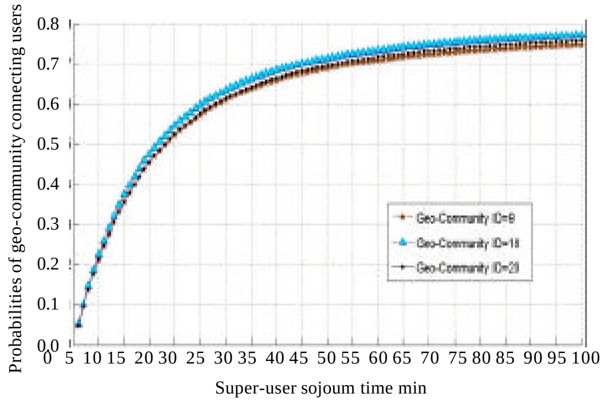

Fig. 3: Curve of Super-user sojourn time and connectivity of geo-community

Interestingness of the Geo-community:

Steady-state distribution U_j^k of all users: Fan *et al.* (2010) provide us with the calculation formula of the steady-state distribution U_j^k. Namely:

$$U_i^k = \frac{\bar{d}_i^k \bullet \pi_i^k}{\sum_{i=1}^{J} (\bar{d}_i^k \bullet \pi_i^k)} \tag{2}$$

The steady-state distribution $[U_i^k]$ is the probability distribution of $User_k$ over geo-communities at any instant and ishence the long-term average distribution of $User_k$ over geo-communities. In Section 3.3.2, we use U_i^k to represent the contact probability per unit time between geo-Community j and $User_k$.

In Fig. 3, with the increase of the sojourn time that superuser stay in the geo-community, the contact probability per unit time between geo-Community i and $User_k$ is increasing.

Geo-community interestingness calculation: Based on the above analysis, the geo-community interestingness model is as follows:

$$X_i(t_i) = I_i \times (1 - \frac{1}{N} \sum_{k=1}^{N} (1 - U_i^k)^{t_i}) \tag{3}$$

where, $I_i = \varepsilon_i + \sigma_i$, U_i^k denotes the contact probability per unit time between geo-Community i and $User_k$. T_i denotes the sojourn time that superuser stays in the geo-community i. Through the establishment of geo-community interestingness model, we can effectively distinguish each geo-community out, to facilitate the superuser choose the appropriate geo-community and the appropriate sojourn time for data dissemination.

In Fig. 4,we show the different interest values of each geo-community over different sojourn time periods.

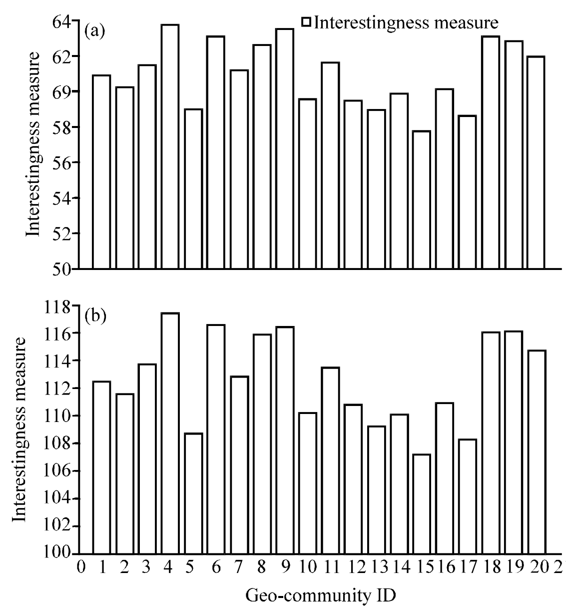

Fig. 4: Various geo-community interest values among different super-user sojourn time

According to the image, the model of geo-community has a comparative distinguishable effect.

DESIGNING ALGORITHMS FOR THE SUPERUSER ROUTE

Static routing algorithm: Superuser's total routing time consists of two parts:

(1) Waiting time T_w: Namely total waiting time in all geo-communities, (2) Travel time T_t: Namely total time moved between all the geo-communities. The total routing time is $T = T_w + T_t$. Assuming that only T_w contributes to the

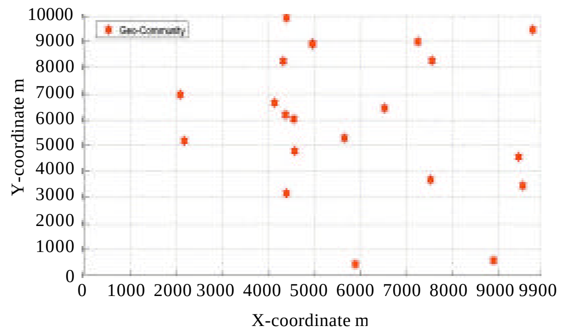

Fig. 5: Geo-community location map

superuser's data dissemination, Superuser's path planning can be divided into two sub-problems: Select the appropriate geo-community set and assign the appropriate time for each geo-community which is in the set, then order these geo-communities together to form a tour.

Superuser's goal is to maximize the dissemination rate p before the deadline T. Then the problem can be modeled as the following optimization problem:

$$\max p = \sum_{i=1}^{J} X_i(t_i)$$

$$\text{s.t.} \sum_{i=1}^{J} t_i + T_t(t_0, t_1, \ldots, t_J) \le T \tag{4}$$

$$t_i \ge 0, 1 \le i \le J,$$

The problem can be simplified as knapsack problem. We proposes the following geo-community efficiency concept:

$$e_i = \frac{X_i'(t)|_{t=0}}{t_{cur,i}} \tag{5}$$

where, $X_i'(0)$ indicates the gradient of X_i (t_i) at $t_i = 0$, $t_{cur,\,1}$ represents the traveling time from the current community to Community i. The algorithm of the problem is that we choose the geo-community which is highest efficiency and put it to the knapsack each time. The superuser can wait at each chosen community until the increment of geo-centrality descends below a certain threshold σ. Obviously, the set of communities generate the highest interestingness while consuming the shortest traveling time in total.

Greedy adaptive route algorithm: In our greedy adaptive route algorithm we still choose the geo-community interestingness as basis for data dissemination, but instead, it computes geo-community interestingness of non-contacted users for each community repeatedly.

Given a collection of geo-communities $S = \{1, 2, ..., J\}$, a collection of users $U = \{1, 2, ..., N\}$. Let $\mathfrak{R}$ be a collection of contacted users. Assume X_i (t_i) denotes geo-community interestingness function of geo-community i during waiting time t_i. $\overline{X}_I$ (t_i) denotes such interestingness of non-contactedusers covered by geo-Community i.

Algorithm 1 gives the greedy adaptive route algorithm of superuser where, T denotes the total time limit of this data dissemination turn. Subscript cur denotes the current community where the superuser stays. t_{stay} denotes the waiting time at the current community, $t_{stay,\,I}$ is the traveling time from the current community to community j. Assuming the moving speed of the superuser is known, then $t_{cur,\,j}$ is a constant. $\overline{X}_i'(0)$ stands for the gradient of $\overline{X}_I$ (t_i) at $t_i = 0$. This algorithm can be dynamically updated geo-community interestingness to overcome the overlap among geo-communities in the network by facing non-contacted users each step, but produced more computational overhead.

Algorithm 1: Greedy Adaptive Route Algorithm (GARA)

1: $\mathfrak{R}$; ←ø; S, T
2:Compute X'_i (0) for every i∈S
3: Stop at the Geo-community with maximal $X_i'(0)$ as the first station
4:While t≤T
5:if $User_k \in Community_{cur}$ then
6: $\mathfrak{R} \leftarrow \mathfrak{R} \cup User_k$
7: end if
8: a [I] = $\overline{X}'_i$ (0), 1<i<J, I≠cur
9: w = a [j] = max a
10:if:

$$(\overline{X}'_{cur}(t_{stay}) \le w) \wedge (\overline{X}_j(T - t - t_{cur,j})) \ge (\overline{X}_{cur}(T) - \overline{X}_{cur}(t_{stay}))$$

then
11: Move to Community j, $t = t_{stay} + t_{sur,\,j}$
12: else
13: Stay at the current community; t_{stay}++
14:end if
15:end While

SIMULATION

Parameter settings: We use Matlab for simulation analysis. There are N = 1000 regular users in our simulation and J = 20 geo-communities. Infocom06 is currently the only public datasets which provide all the geographic information of the internet infrastructure APs and in the project infocom06, there are only 78 regular users. But our interest-based data dissemination superuser path planning would be more meaningful if it used on a larger scale in mobile social networks. Therefore, we use simulated datasets in our experiment. In

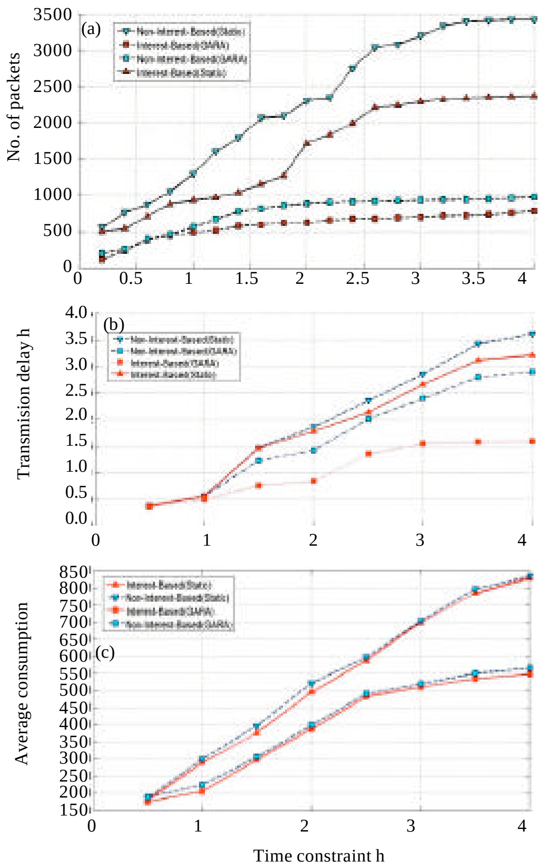

Fig. 6 (a-c): Comparison figure, (a): No. of packets comparison figure, (b): Transmission delay and (c): Average consumption

this study we consider the 1000 users' situation and assuming that steady-state distribution of all users is already known. Meanwhile, coordinates of all geo-communities is randomly generated. The coordinates of each community is shown in Fig. 5.

In the experiment, the three key parameters in the mobile social network data dissemination are as follows:

Number of the packet: The total number of packets that superuser send to the network

Transmission delay: The average time of the data is send successfully by superuser

Average consumption: Sum of the total distance traveled by superuser and the total number of packets that superuser send to the network

Performance comparison: We assumes that superuser immediately disseminate data to users who are interested in the data when he reaches a geo-community. Interest matching process is instantaneous. Superuser's moving speed between geo-communities is v = 15 m sce^{-1}. We have compared the three key parameters within different time limit under different scenarios, respectively. It is shown in Fig. 6.

Overall, the proposed interest-based data dissemination mechanism show fewer packet number, lower transmission delay and energy consumption.

CONCLUSION

This study mainly studies user-interest-based accurate data dissemination based on geographic communities in mobile social networks. The main idea is planning data dissemination paths for superuser based on geo-community interestingness. Specifically, we firstly propose the concept of accurate data dissemination. The data is disseminated accurately to the users who are interested in the data while users who are not interested in the data cannot obtain the data. Then, we establish interestingness measure model of the geo-community according to existing researches. The model of geo-community interestingness has a comparative distinguishable effect, it has laid a good foundation for superuser to select the appropriate geo-community to disseminate data. We develop a static and dynamic path planning algorithm for superuser respectively and we compare our proposed algorithm to the current distribution mechanism. Experimental results show that our proposed mechanism show fewer packet number, lower transmission delay and energy consumption. It achieves better results.

REFERENCES

Du, Y., J. Fan and J. Chen, 2011. Experimental analysis of user mobility pattern in mobile social networks. Proceedings of the 2011 IEEE Wireless Communications and Networking Conference (WCNC), March 28-31, 2011, Cancun, Quintana Roo, pp: 1086-1090.

Fan, J., J. Chen, Y. Du, W. Gao, J. Wu and Y. Sun, 2012. Geocommunity-based broadcasting for data dissemination in mobile social networks. IEEE Trans. Parallel Distrib. Syst., 24: 734-743.

Fan, J., Y. Du, W. Gao, J. Chen and Y. Sun, 2010. Geography-aware active data dissemination in mobile social networks. Proceeding of the IEEE 7th International Conference on Mobile Adhoc and Sensor Systems (MASS), November 8-12, 2010, San Francisco, CA., pp: 109-118.

Gao, W. and G. Cao, 2011. User-centric data dissemination in disruption tolerant networks. Proceedings of the IEEE INFOCOM, April 10-15, 2011, Shanghai, pp: 3119-3127.

Han, B., P. Hui, V.S.A. Kumar, M.V. Marathe, J. Shao and A. Srinivasan, 2012. Mobile data offloading through opportunistic communications and social participation. IEEE Trans. Mobile Comput., 11: 821-834.

Ioannidis, S., A. Chaintreau and L. Massoulie, 2009. Optimal and scalable distribution of content updates over a mobile social network. In Proceedings of the IEEE INFOCOM, April 19-25, 2009, Rio de Janeiro, pp: 1422-1430.

Rao, W., K. Zhao, Y. Zhang, P. Hui and S. Tarkoma, 2012. Maximizing timely content advertising in DTNs. Proceedings of the 2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, June 18-21, 2012, Seoul, pp: 254-262.

Shah, R., S. Roy, S. Jain and W. Brunette, 2003. Data MULEs: Modeling a three-tier architecture for sparse sensor networks. Proceedings of the 1st IEEE 2003 International Workshop on Sensor Network Protocols and Applications, May 11, 2003, Anchorage, AK., pp: 30-41.

Tariq, M.M.B., M. Ammar and E. Zegura, 2006. Message ferry route design for sparse ad hoc networks with mobile nodes. Proceedings of the International Symposium on Mobile Ad Hoc Networking and Computing, May 22-25, 2006, Florence, Italy, pp: 37-48.

Wu, J. and Y. Wang, 2012. Social feature-based multi-path routing in delay tolerant networks. Proceedings of the 3lst IEEE International Conference on Computer Communications, March 25-30, 2012, Philadelphia, PA.,.

Wu, J., M. Xiao and L. Huang, 2013. Homing spread: Community home-based multi-copy routing in mobile social networks. Proceedings of the 32nd IEEE International Conference on Computer Communications, April 14-19, 2013, Turin, Italy, pp: 2319-2327.

Zhao, W., M. Ammar and E. Zegura, 2004. A message ferrying approach for data delivery in sparse mobile ad hoc networks. Proceedings of 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing, Tokyo, Japan, May 24-26, ACM Press, pp: 187-198.

Zhao, W., M. Ammar and E. Zegura, 2005. Controlling the mobility of multiple data transport ferries in a delay-tolerant network. Proceedings of INFOCOM2005, Miami, Florida, USA., March 13-17, ACM, New York, pp: 1407-1418.