Computer Science and Information Systems 18(2):441–460 https://doi.org/10.2298/CSIS200128004C

Message Propagation in DTN Based on Virtual Contact of Behavior Model

Ho-Hsiang Chan[1] and Tzu-Chieh Tsai[2]

[1]Department of Computer Science
National Chengchi University
Taipei, Taiwan
100753503@nccu.edu.tw
[2]Department of Computer Science
National Chengchi University
Taipei, Taiwan
ttsai@cs.nccu.edu.tw

Abstract. Delay Tolerant Network (DTN) is a kind of network structured to deliver message intermittently. Network connections are not persistent between nodes, instead they must rely on nodes making geographic location movements to incur contact with other nodes and establish intermittent communication sessions to allow messages delivery. We will refer to encounters via geographic location movements as "physical contact." Many DTN researches mainly focus on message delivery via physical contact. However, this paper believes that in a realistic environment, encounters between nodes not only happen geographically in nature, but also occur virtually in cyberspace. When both nodes go online on the same social media platform, it is an encounter we refer as *virtual contact*. How messages deliver for virtual contact is store-post-and-forward, just like what happens in a DTN, but it is no longer restrained by geographical locations. This paper considers a scenario in which nodes make virtual contact in cyberspace and incur message delivery based on their own behavior patterns. The verifying experiment is conducted using both survey and simulation. First of all, we handed out questionnaires for students to fill out. The questionnaire inquired them to rank their most frequent activities performed on social media platforms. According to the responses, we conclude the top 3 frequent activities when the students use social media platforms and classify them into 3 groups according to a weighted behavior pattern scheme. The classification includes *Social Group*, *Read-Only* Group and *Interest Group*. It does not matter which group a student is assigned to. In the simulation, he or she will get to decide whether to deliver/receive messages or not based on a randomized selection on 3 behavior pattern. Finally, we analyze the simulation result to determine how messages propagated in different behavior pattern groups. It is derived from the simulation that to quicken message propagation, directing messages to one of the behavior groups yields the maximize benefits. This provides the basis for further researches on collecting data of desired scenarios to establish respective propagation models.

Keywords: Delay Tolerant Network, physical contact, virtual contact, behavior pattern

1. Introduction

1.1 Background

Nowadays the Internet is flooded with information, and it comes in different flavors. As different kinds of information spread across the world each day, people are busy receiving in and sending out information on various social platforms. For example, economic news, political comments, and sport coverage are among the most covered ones and get forwarded repetitively. To understand how quickly information is dispersed, it is necessary to evaluate the number of forwards/shares made and the frequency of forwarding done by each social platform user. As cellular network progresses and mobile smart devices popularizes, people can log in to social platforms to forward messages agilely whenever they can and wherever they want. Consequently, if there are urgent messages that must be known by the masses in a short time, transferring them through other users on social platforms or on the Internet will likely to induce higher probability of quick message propagation.

Delay Tolerant Network (DTN) is an instance of Opportunistic Network Environment, in which no single route serves as a persistent end-to-end connection, and it requires users carrying smart mobile devices to move between same geographic locations to have the opportunities of sending out messages. The majority of previous researches focus on exploring methods of message delivery and forwarding via geographic location encounters. We have also proposedan approach, NCCU Trace Data [16] which involves collecting the data of students' real geographical location movements in campus environment. When a student meets with other students, there are opportunities of forwarding messages further. However, this paper believes that current DTN researches' focuses on physical encounter scenarios rely on people's move to triggering encounters with other peoples, which no long conforms with the real-world environment.

With the advancements of Information and Communication Technology, we can easily connect to social media platforms (i.e. Facebook, Twitter, Whatsapp, etc.) via mobile networks to share information with other users. This way of transferring messages is similar to the case of DTN. The further messages are delivered, the less likely for receiving users to be online. Thus, if a user sends out messages when others happen to be online at the same time, he or she still needs to transfer messages to social media platforms before these messages are delivered. However, even if receiving users are offline, messages still remain on social media platforms waiting to be transmitted to the currently offline users when they get online.

1.2 Virtual Contact

This research believes that above-mentioned scenario is a case of *virtual contact*, as described by Figure 1 below. As illustrated by the figure, virtual contacts happen when social media users are physically located in distant locations where physical contacts are not possible, yet they can achieve virtual space encounters in cyberspace constructed on the Internet disregarding the boundaries of space and time as long as they were once online on common social medias. Virtual contact on social platform align with one of the characteristics of DTN, store-carry-and-forward. Because users are likely to get online at different time, they are able to deliver messages on social media platforms when online, and other receiving users will accept the messages once online. If both the message-sending user and the message-receiving user are online at the same time, real-time transferring will be carried out. Message delivery is no longer limited to the classic scenario that users must be located in the same geographic region. Owing to the convenience of social media platforms, users can make *virtual contact* with each other at different geographic locations, so that the message-carrying users will be able to deliver them to the others.

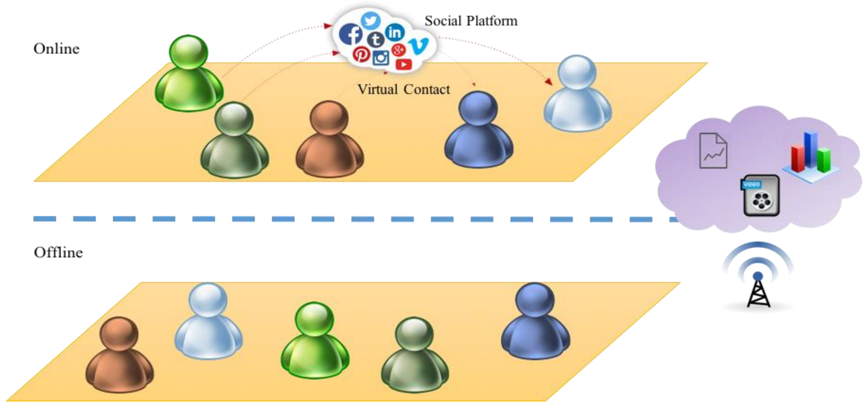

Fig.1. Overview of the virtual contact diagram

1.3 Motivation

Previous works on message delivery mostly neglect to consider message senders' delivering behavior. In a real-world environment, everyone acts out of his or her own free will and reacts to forward or receive messages. For example, some users tend to passively accept messages without any intent to share or forward them, while other users are more likely to share received messages in the hopes that even more people know about them.

This paper believes that the scenario of virtual contact is just like the characteristic of store-carry-and-forward. When users become online on social media platforms, the message-carrying users are granted with the opportunities of meeting other users, and making direct message delivery. Besides, we think people would behave differently according to their current moods and environments. For instance, one may prefer browsing messages to sharing them; one may also desire to share interesting messages to other internet users of similar interests, etc. Thus, this paper investigates how users' daily behavior patterns can influence the way messages are delivered. Considering a message delivery scenario, when an urgent message shall be sent in a way that the more people learn the better it is, we suggest a better accommodating message propagation scheme: transmitting messages in accordance with messages' properties or users' behaviors when using social media platforms.

1.4 Purpose

This research proposes an approach that message delivery should be determined based on users' individual behaviors in a network environment. Regardless of the message delivery and acceptance, everyone will ultimately choose to deliver or receive messages based on their own interests, which aligns with the real-world scenario on message propagation. Furthermore, messages categorized to be of similar interests would make carriers behave alike. This paper attempts to resolve the problem of finding the most appropriate person to propagate messages with interests in respective fields, thus allowing them to be received by as many people as possible within a fixed amount of time. Lastly, we utilize message propagation behaviors to conduct clustering, and, therefore, are able to efficiently find a model for fast message propagation. When there is a certain type of messages needed to be dispersed quickly, we can find groups most suitable for fast message propagation via the above-mentioned message propagation model based on behavior clustering to achieve better efficiency.

2. Related Work

Previous social network researches based on DTN define social community according nodal geographic locations and chances of making encounters with other nodes to decide their closeness in terms of social network distances, then develop strategies of message delivery based on it. SimBet[4] utilizes nodal betweenness to computer centrality and similarity for each node to help making decision on which nodes to deliver messages.

This article believes that nodes make frequent encounters in terms of geographical locations, then there exists a social network relationship between them. It further proposes the more mobile nodes are, the nodes with better utilities can be to help deliver message. However, one drawback of such design is that messages are likely to be centralized on nodes with better utilities. If there exists less active nodes in the network, they, in the worst case scenario, might never get any message. Bubble rap[12] thinks it is better to take a prolonged observation on each node's encounters via geographical location and turn the observation results into a simulated social network. Each node should belong to at least one community or multiple communities, and each node should have information on global ranking and local ranking of the whole network system. Routing algorithm then utilizes estimated community and centrality of the social network. According the ranking done by the global community, messages are forwarded to the node with highest centrality in the community until the target node and the forwarding node belong to the same community. SANE routing [8] depends otherwise on interests and similarities as estimation basis for message delivery. The author believes that nodes with similar interests are likely to make encounters with each other. If a message is to be forwarded to the target node, forwarding the message to the nodes with similar interests will yield a better transmission efficiency. A previous research by us, NCCU Trace Data[16] is about collecting students' movements in a campus. As a laboratory collective effort, we develop a APP which can be utilized to track students' movement traces when attending classes at school, as showed in Figure 2, and propose a message delivery method via similarities between people's interests in a real-world campus environment. This method performs better than traditional routing algorithms.

Fig.2. NCCU Trace Data of the screenshot

Besides, our previous research [17] also suggests that people will often move, according to their interests, to buildings of similar interest properties. It may happen that people with similar interests are likely to have similar routes of geographical location displacements. The message delivery method based on interest properties has a better performance compared to other routing algorithms.

The aforementioned works are all about investigating the fact that when nodes make frequent encounters via geographical location movements, there may exist social network relation and exhibit close betweenness among nodes. If a message is to be delivered, this leads to better message delivery ratio. However, this kind of social network estimation method tends to be shallow. Frequent encounters via geographical location might not necessarily mean close betweenness, but only show that nodes happen to be neighbors or have the same moving directions. This paper believes that personal information should be utilized to confirm whether there exists social network relationships between nodes, then making estimates of nodes' social network relationships to develop a method of message delivery that aligns better with real-world environments. Table 1 is the comparison among different strategies adopted by referenced works.

In a conventional Delay Tolerant Network environment, encounters via geographical locations introduce opportunities for message propagation or dispersion. Compared to the conventional methods, this paper is different in the way that opportunities for message dissemination, as we proposed, occur in cyberspace, so methods suggested by previous works are merely inspirations for our approach for that there is temporarily no similar method comparable to this research.

Table 1. Comparison table of the related work

Related work	Forward Strategy	Mobility Model	Characteristic
SimBet[4]	Node Utility	MIT Reality	betweenness Centrality
Bubble rap[12]	Node Utility	MIT Reality Cambridge Infocom06	ego-central
SANE [8]	Node Utility	Infocom06	Social-aware
NCCU Trace Data [16]	Message Interest Utility	NCCU Trace Data	Interest-aware

3. Proposed Approach

We are continuing a previously proposed work, NCCU Trace Data [16]. This research is part of the project responsible of NCCU Trace Data collection from 115 participants' Facebook online histories and friend lists, except that we were unable to collect online histories properly from 11 of them. This makes a total of 104 participants' online histories, all of whom were notified and agreed with how we handle their data. Moreover, we asked each of the 104 experiment participants to fill out a questionnaire aiming to collect personal information, interests, and rankings of frequent Facebook activities, etc. These activities include: (1) sharing messages with friends on social media platforms (referred it as *Social*); (2) sharing articles in the groups on social media platforms (referred as *Interest*); (3) refusing to share messages while only receiving them (referred as *Read Only*). According the rankings done by the participants, we would assign the most frequent activities a weight of 3, the second most frequent one a weight of 2, and the lease frequent one a weight of 1. Each participant is required to fill out at least 1 frequent activity. Part of the original data are presented by Table 2.

Table 2. Behavioral ordering

User_ID	Social	Interest	Read Only
1	3	1	1
2	3	1	1
3	3	2	1
4	3	1	1
5	3	1	1
6	3	1	2
7	3	2	1
8	3	1	1
9	1	1	3
10	3	1	2
11	3	1	3
12	3	1	1

According to the answers filled by the participants, this research would utilize K-means algorithm to achieve clustering based on each one's behavior weights, divide them into 3 groups, and derive the classification scheme from each group's behavior weight features. The results are presented in Table 3.

Table 3.K-means algorithm classification

User_ID	Group	Distance	Social	Interest	Read Only
1	0	1.337	3	1	1
2	0	1.337	3	1	1
3	0	1.669	3	2	1
4	0	1.337	3	1	1
5	0	1.337	3	1	1
6	0	0.492	3	1	2
7	0	1.669	3	2	1
8	0	1.337	3	1	1
9	1	2.642	1	1	3
10	0	0.492	3	1	2
11	1	2.717	2	1	3
12	0	1.337	3	1	1
13	0	1.669	3	2	1
14	2	2.492	1	3	1

K-means algorithm classifies the participants into 3 different groups. The first group numbered with 0 consists of 88 people. The second group numbered with 1 consists of 6 people. The third group numbered with 2 consists of 10 people. Table 4 below displays part of the data for Group 0.

Table 4. K-means algorithm classification

User_ID	Group	Distance	Social	Interest	Read Only
1	0	1.337	3	1	1
2	0	1.337	3	1	1
3	0	1.669	3	2	1
4	0	1.337	3	1	1
5	0	1.337	3	1	1
6	0	0.492	3	1	2
7	0	1.669	3	2	1
8	0	1.337	3	1	1
10	0	0.492	3	1	2

As the data K-means algorithm classified to be group 0 show, participants prefer to share messages on social media platforms, described by their responses to the questionnaires. Thus, this research defines the first group as Social Group. Although classified as part of the Social Group, participants would obviously engage in activities other than sharing messages with friends on social media platforms. They could also choose to share articles in their groups or simply receive messages without sharing them. The second group classified by K-means algorithm, as showed by their responses to the questionnaires, all prefer to only receive messages without sharing, so they are defined as Read Only Group. The third group classified by K-means algorithm stated in their questionnaire responses that they all prefer to share articles in their groups, so they are defined as Interest Group.

Figure 3 below is a visualization of Social Group which this research attempts to represent with a tree diagram. Each route consists of three nodes representing different behavior preferences, as the participants can only engage in one certain activity during a fixed amount of time. The weights of the activity nodes would change according to the routes connecting them. The first route shows that 88 of the participants would prioritize to share messages with friends on social media platforms, 37 of them would secondly prioritize to share messages with their groups, and these 37 people are least likely to only receive messages without sharing on social media platforms. The chance of traveling along the first route is 42%. The second route shows that 88 of the participants would prioritize to share messages with friends on social media platforms, 51 of them are least likely to share messages with groups on social media platforms, and 16 of them secondly prioritize to receive messages without sharing them on social media platforms. The chance of traveling along the second route is 18%. The third route shows that 88 of the participants would prioritize to share messages with friends on social media platforms, 51 of them are least likely to share message with groups on social media platforms, and 35 of them are also least likely to only receive messages without sharing on social media platforms. The chance of traveling along the third route is 40%. Based on the descriptions above, we can conclude that if there is a message to be received or delivered, users would have 42% chance to choose the first route, 18% chance to choose the second route, and 40% chance to choose the third route while each route has different weights according to the activity preferences it represents. Each activity carries a weight. The higher the weight it is, the more likely the corresponding activity to be performed. The actual resulting route will randomly select one of the activities to be the basis of propagating or receiving messages. Figure 4 and Figure 5 are the visualizations of Read Only Group and Interest Group, respectively.

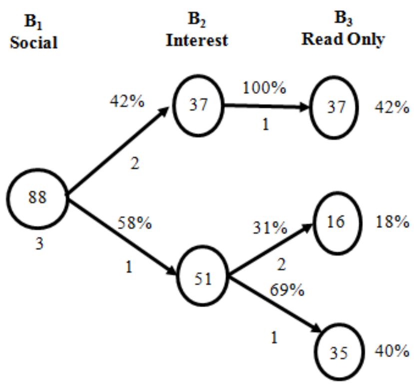

Fig.3. Social Group

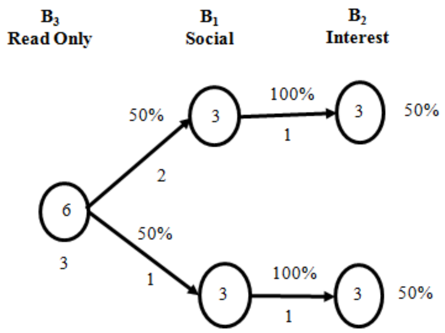

Fig.4. Read Only Group

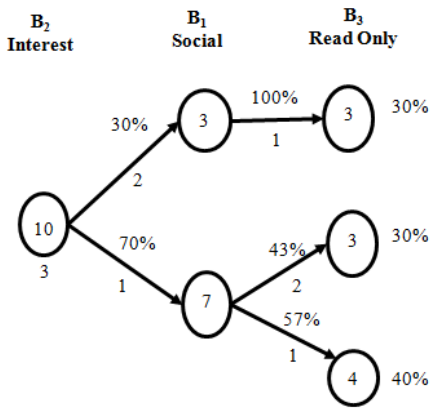

Fig.5. Interest Group

Like explained above, this research classifies the behavior patterns into three categories: (1) receiving messages from or propagating messages to social media friends, referred as Social Behavior; (2) receiving interest-provoking messages or propagate messages to interested users, referred as Interested Behavior; (3) only receiving messages without sharing to other users, referred as Read-Only Behavior.

(1) Message receiving and propagating for Social Behavior: users log into social media platforms, gain access to their friend lists of respective platforms, and invoke actions to receive or deliver messages according to the lists. This behavior pattern captures the scenario that whether users are interested at the messages or not, they will receive messages from their friends and propagate received messages to others. Such scenario is visualized by Figure 6.

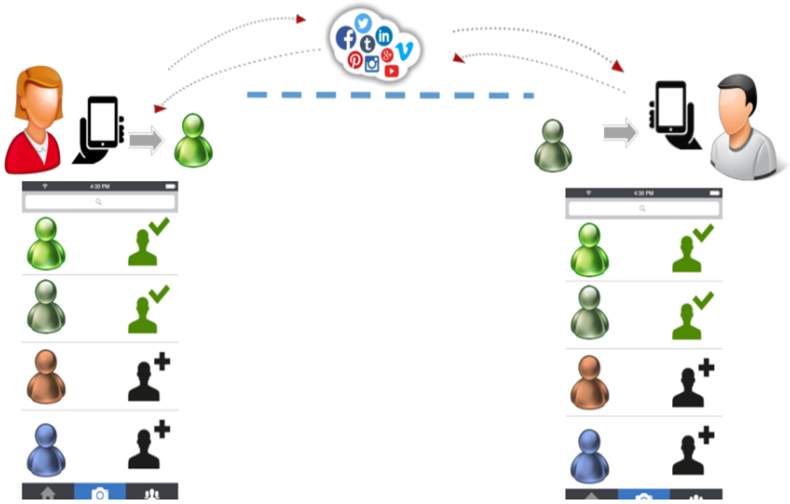

Fig.6. Message receiving and propagating for Social Behavior

(2) Message receiving and propagating for Interested Behavior: When a user U_X gets online on social media platforms, this behavior pattern captures the scenario that no matter users are friends of each other or not, they would receive interested messages or deliver interested messages to other users. Possible interest properties include: sports, reading, social activities, artistic events, community services. When a user receives messages, he or she will determine whether messages' interest properties $M_K(I_V)$ correlate with his or her own interests. To calculate its correlation with the interests, we rely on Cosine similarity as the basis. On the other hand, message delivery also depends on Cosine similarity as the basis to calculate its correlation with the interests. As shown in formula (1) below:

$$\mathrm{Cos}\big(U_X(I_V), M_K(I_V)\big) = \frac{U_X(I_V) \cdot M_K(I_V)}{||U_X(I_V)|| \cdot ||M_K(I_V)||}. \tag{1}$$

(3) Message receiving and propagating for Read-Only Behavior: When users of this behavior pattern gets online on social media platforms, they will receive messages delivered from their friends or interested messages, but they would not deliver any message at all to others.

This research summarizes the mechanism of how virtual contact on social media platforms triggering receiving and delivering message as showed by Figure 7. It is further explained below:

(1) A user U_x relies on his or her hobby when using social media platforms to decide the moment of time he or she gets online.
(2) The user then confirms he or she shall be categorized as social group, interest group, or read-only group.
(3) At the message-receiving stage, the user will decide the action to be taken at this point of time according to a randomized probability.

(4) After receiving a message, the users then switches to the message-propagating stage.
(5) At the message-propagating stage, the user will decide again the action to be taken at this point of time according to a randomized probability.
(6) After executing both message-receiving and message-propagating stages, the users then gets offline on the social media platform.

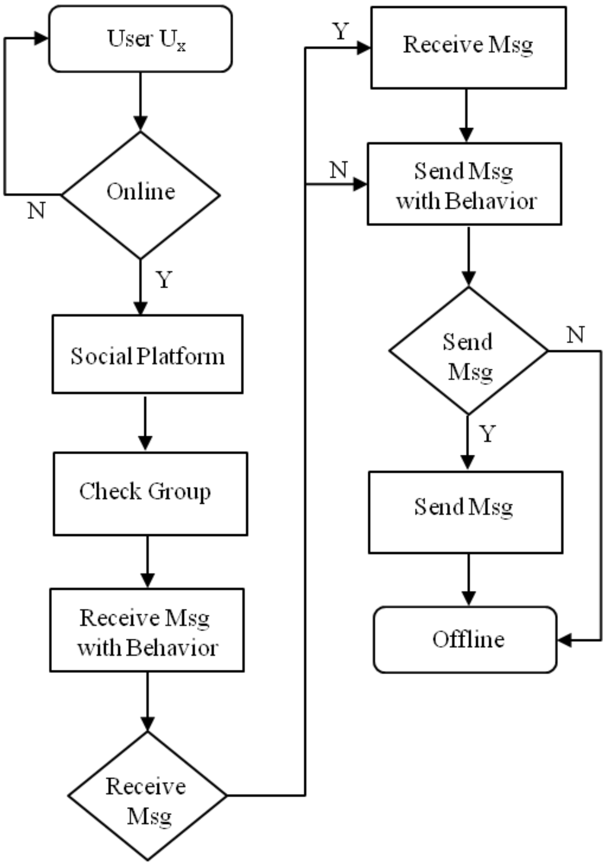

Fig.7. The Approach

In this research, we propose a message receiving and propagating mechanism constructed using the user behavior algorithm described below. When user goes online, the user will decide the action to be taken at this point of time according to a randomized probability, which is described by Algorithm 1. Message receiving and propagating mechanism is presented in Algorithm 2:

Algorithm 1 Checkout the Behavior.

Input: a dataset of User U=$\{u_1,\dots,u_n\}$, Behavior B=$\{b_1,\dots,b_3\}$, User Behavior Tree T=$\{t_1,\dots,t_3\}$, Group G =$\{g_1,\dots,g_3\}$

Output: User Behavior

foreach u_iUdo /* Receive Message*/
for (x, 1 to |G |)
if $u_i g_x$
select one of route from Tby
if random(1,n)<$t^{r1}_{gx}.\text{leaf}\lor$
switch (random(1,j))
case <$t^{r1}_{gx}.\text{weightofb}_1$
u_i belongs to ($t^{r1}_{gx}.\text{behaviorofb}_1$)
case <$t^{r1}_{gx}.\text{weightofb}_x + t^{r1}_{gx}.\text{weightofb}_2$
u_i belongs to ($t^{r1}_{gx}.\text{behaviorofb}_2$)
case <$t^{r1}_{gx}.\text{weightofb}_1 + \cdots + t^{r1}_{gx}.\text{weightofb}_3$
u_i belongs to ($t^{r1}_{gx}.\text{behaviorofb}_3$)
else if random(1,n) <=$t^{r1}_{gx}.\text{leaf}\lor + t^{r2}_{gx}.\text{leaf}\lor$
switch (random(1,j))
case <$t^{r2}_{gx}.\text{weightofb}_2$
u_i belongs to ($t^{r2}_{gx}.\text{behaviorofb}_2$)
case <$t^{r2}_{gx}.\text{weightofb}_2 + t^{r2}_{gx}.\text{weightofb}_1$
u_i belongs to ($t^{r2}_{gx}.\text{behaviorofb}_1$)
case <$t^{r2}_{gx}.\text{weightofb}_2 + \dots + t^{r2}_{gx}.\text{weightofb}_3$
u_i belongs to ($t^{r2}_{gx}.\text{behaviorofb}_3$)
elseif random(1,n) <=$t^{r1}_{gx}.\text{leaf}\big| + t^{r2}_{gx}.\text{leaf}\big| + t^{r3}_{gx}.\text{leaf}\lor$
switch (random(1,j))
case <$t^{r3}_{gx}.\text{weightof}(t^{r3}_{gx}.\text{behaviorofb}_3)$
u_i belongs to ($t^{r3}_{gx}.\text{behaviorofb}_3$)
case <$t^{r3}_{gx}.\text{weightofb}_3 + t^{r3}_{gx}.\text{weightofb}_1$
u_i belongs to ($t^{r3}_{gx}.\text{behaviorofb}_1$)
case <$t^{r3}_{gx}.\text{weightofb}_3 + \dots + t^{r3}_{gx}.\text{weightofb}_2$
u_i belongs to ($t^{r3}_{gx}.\text{behaviorofb}_2$)

Algorithm 2 Proposed Algorithm.

Input: a dataset of User Behavior U={$u_1, \dots, u_n$}, a message M with Interest i M={$m_1^i, \dots, m_j^i$}, Behavior B={$b_1, \dots, b_3$}, Group G={$g_1, \dots, g_3$}

Output: send and receive message behavior

Initial:

    Checkout the Behavior$B$={$b_1, \dots, b_3$} , for all User U={$u_1, \dots, u_n$}

foreach$u_i U$**do** /* Receive Message*/

    while check all message m_j^i from source node

            if $u_i b_1$

              if source in the friend list of u_i

    receive the message m_j^i to user u_i

            if $u_i b_2$

              if $S_c(m_j^i, u_1)$> *Threshold Constant*

            receive the message m_j^i to user u_i

            if $u_i b_3$

            receive the message m_j^i to user u_i

foreach$u_i U$**do** /* Send Message*/

    while check all received message m_l^i from source

            if $u_i b_1$

              if source in the friend list of u_i

            send the message m_l^i to source node

            if $u_i b_2$

              if $S_c(m_l^i, u_1)$>*Threshold Constant*

            send the message m_l^i to source node

            if $u_i b_3$

                do nothing

4. SimulationResult

4.1. Simulation Setting

The simulation conducted by this research has been run in a campus environment, by importing students' online histories on social media platforms and utilizing those as inputs a program written to realize the probabilistic model of students' daily behavior pattern on the platforms. This adopts The ONE Simulator to simulate the number of the actual experiment participants as show in Figure 8:

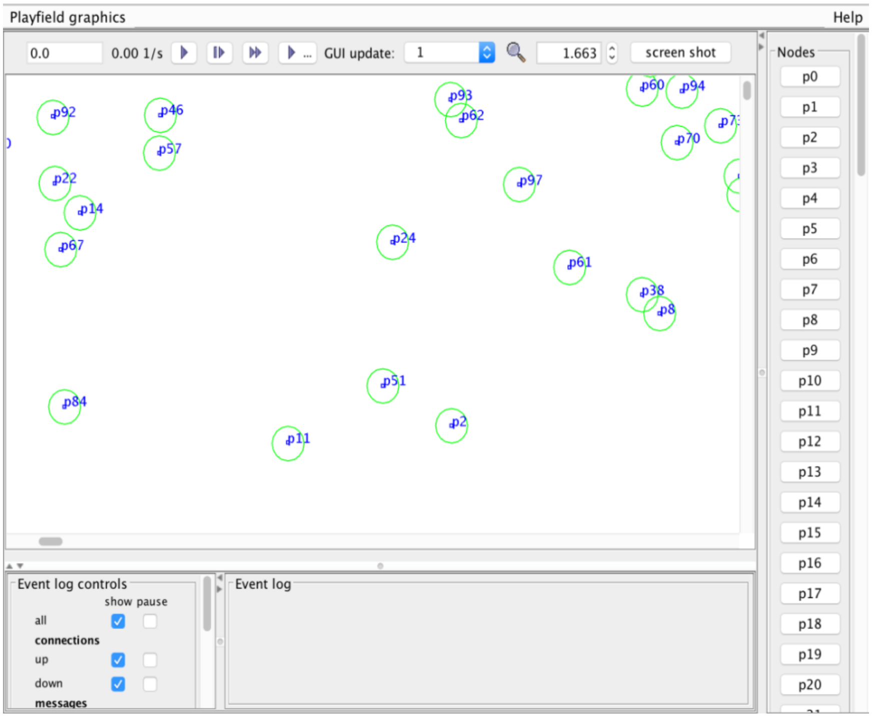

Fig.8. The ONESimulator

Using 104 nodes as the representation. The simulation duration is set to be one day, and only one message to be delivered during the session. The message is to be generated by randomly choose one node from the three groups including Social Group, Interest Group, and Read-Only Group. Nodes and message are attached with properties of interests. Each node has a friend list from collected participant data. The experiment parameters set are listed in Table 5 below:

Table 5. Simulation setting

Simulation parameters	Description
Simulation times	86400 sec
Number of nodes	104
Message Size	500K~1MB
Number of message creation	1
Buffer size	500MB
Time To Live	Unlimited
Virtual Contact	The Users behavior

4.2. Simulation Results

This research adopts K-means clustering as the method to conduct clustering based on behavior patterns. In the experimental simulation, we aim to incorporate different clustering algorithms to repeat behavior pattern clustering, which allows us to compare and analyze performances achieved by varied clustering algorithms. Furthermore, the result will help us to evaluate whether using a different clustering algorithm impacts the performance of message propagation. The final simulation result will include results obtained from employing Gaussian mixture model and Mean shift clustering algorithms in respective experiments.

The simulation result is presented in Figure 9. We can clearly interpret the message dissemination speeds from the graph that the result suggests the fastest one is Interest Group, the second fastest one is Social Group, and the last one is Read-Only Group. In the result, it is easy to find out no matter which group a student node is part of, the message propagation ratio is steadily increasing. Of the three groups, Interest Group has a better dissemination ratio because messages contain properties of interests. When a message is interesting to most people, it is easily disseminated at a speed even faster than between friends. This kind of situation resembles how the masses utilize social media platforms as we tend to join groups or follow fan pages according to our interests, such as star chasing, group buying, technical news, etc. Thus, when obtaining information about relevant interests, we are more likely to share articles in those groups. If friends of us happen to share similar interests, it is possible for us to share those messages to them.

In Read-Only Group, the students assigned to this group tend to passively accept messages without sharing. This type of behavior is called by the general public as "Lurker." It is relatively easy to infer from the simulation result that the message propagation for the Lurker group was not as active. We can conclude that in order for a message to be disseminated within a short amount of time, it is necessary to forward it first to the students highly favoring the message's interest property, thus speeding up the message dissemination.

Figure 10 is the visualization obtained from behavior pattern clustering by Gaussian mixture model (GMM) algorithm. We can tell from the simulation result that Interest Group still achieves the best dissemination message ratio, close to 80 % of people received the message. Read Only Group remains the most poorly performed one, gaining nearly 40% of people receiving the message. After message propagation lasted for about a day, Social Group and Interest Group eventually share very similar final dissemination ratios, which shows that adopting Gaussian mixture model algorithm as the approach for clustering would make it more difficult to highlight the obvious gap among different message dissemination ratios. However, employing K-means algorithm as the approach for behavior pattern clustering will make it easier to tell Interest Group has a significantly better message dissemination ratio when compared to Social Group and Read Only Group.

Figure 11 shows behavior pattern clustering by Mean shift algorithm. The simulation result indicates that Interest Group has the best message dissemination ratio, close to 70% of people received the message. Social Group has a message dissemination ratio slightly lower than Interest Group, and Read Only Group remains the worst performed one. The simulation result suggests that no matter which clustering algorithm is used, the behavior pattern of Read Only Group is comparatively the most difficult one for facilitate fast message propagation. As respective simulation results obtained from three different clustering algorithms adopted for behavior clustering, if there is a message characterized by multiple interest properties to be quickly dispersed, K-means clustering algorithm is a better option to distinguish clearly the differences of message dissemination ratios achieved by each behavior pattern clustering, thus identifying the most appropriate group for fast message propagation.

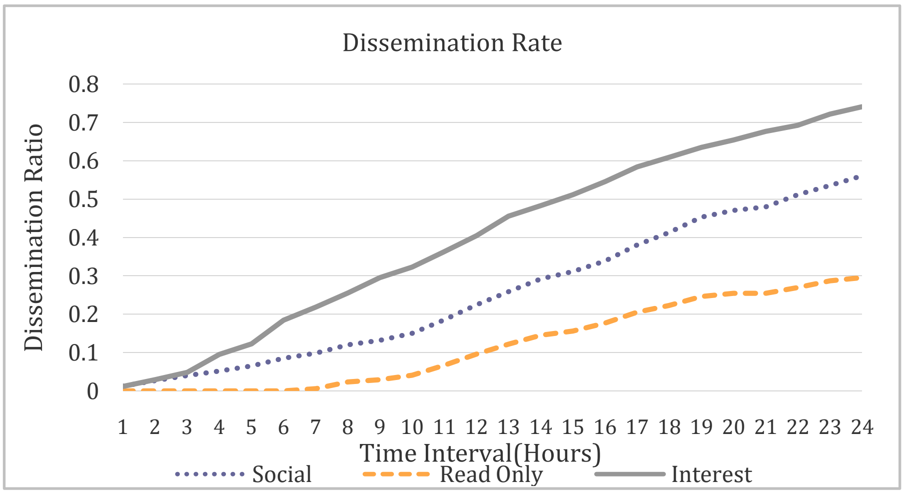

Fig.9. Dissemination Rate (K-means)

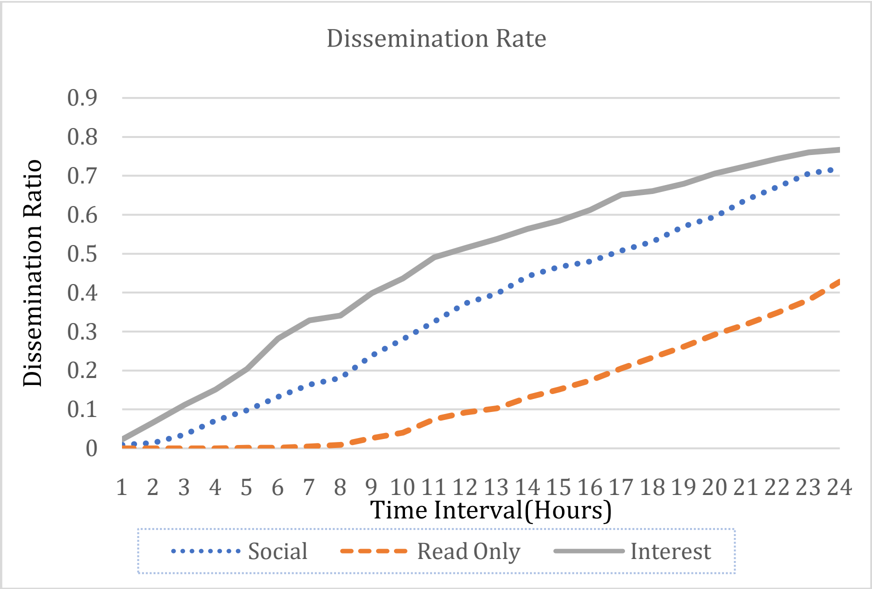

Fig.10. Dissemination Rate (GMM)

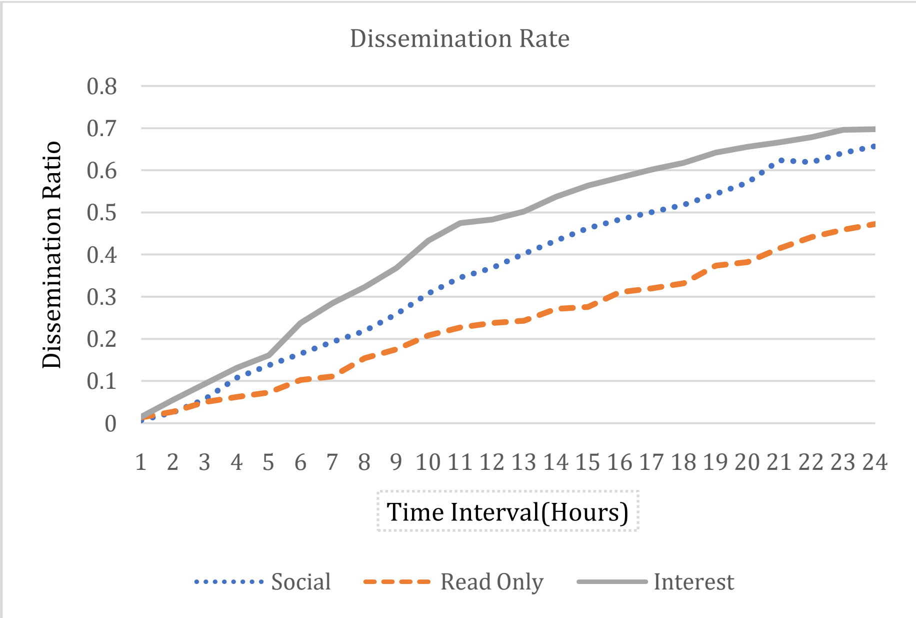

Fig.11. Dissemination Rate (Mean shift)

The simulation results from three different clustering algorithms indicate that people classified as Interest Group tend to have a faster dissemination rate. The major reason is probably due to the fact that multiple interest properties carried by messages will attract more people into dispersing and receiving, and every one's friend list shall have a limit. If a person wanting to disperse messages has few friends, the propagation speed of messages will tend to be slower when compared to messages interested the masses. This simulation result aligns very closely with the real scenario, in which messages are quickly dispersed because either receivers follow commonly interested community groups or fan pages rather than simply shared by friends.

We can compare utilizing different clustering algorithms on behavior patterns to evaluate the simulation of message propagation. From the simulation result, it can be told that people classified as Interest Group will achieve the best dissemination ratio when engaging in message propagation. This indicates the if there is a message carrying multiple interest properties to be quickly dispersed that most of the others obtain information, we can easily target people classified as Interest Group to initiate dispersion, therefore achieve great message dissemination ratio.

The simulation scenario mentioned above requires randomly assigning 5 different interest properties for each dispersed message. Consider another simulation scenario as described by Figure 12, the simulation result is conducted by dispersing messages whose interest properties are characterized as only few of the masses may pay attention to. The simulation result indicates that message dissemination ratio for Social Group is much better compared to Interest Group. We think it may be attributed to that messages' interest properties are not concerned by the masses, thus these messages cannot be easily dispersed for Interest Group. In this simulation scenario, the message dissemination ratio for Interest Group tends to be worse, while the message dissemination ratio for Social Group maintains approximately similar efficiency like the above-mentioned simulation result. Based on this simulation result, we can tell that if messages carrying obvious interest properties are to be dispersed, finding a member of Interest Group to disperse messages will result in a better message dissemination ratio. On the other hand, if messages are of implicit interest properties, finding a member of Social Group to disperse messages will yield a more stable message dissemination ratio.

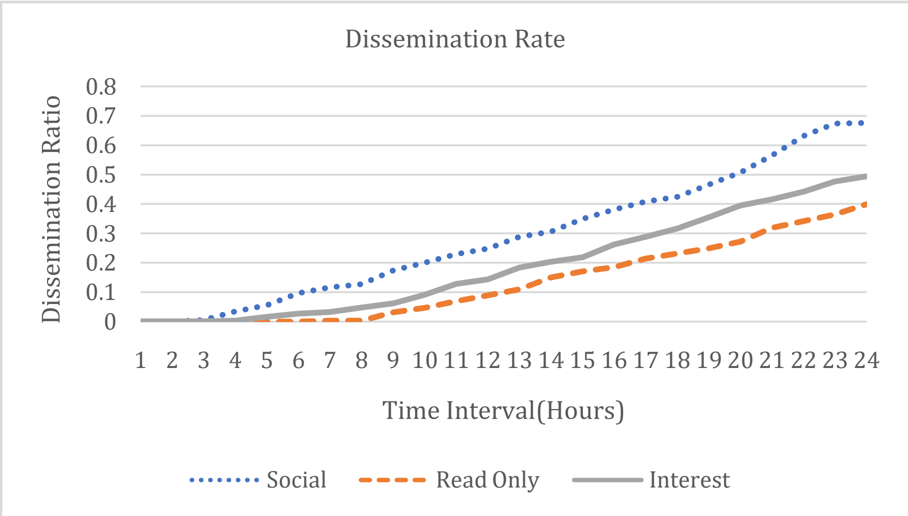

Fig.12. Dissemination Rate

We utilize different clustering algorithms to perform clustering on message dissemination behaviors. It can be told from the simulation result that when dispersing messages with clear interest properties, the message dissemination model obtained from behavior clustering shows that finding Interest Group to perform message propagation will yield better message dissemination ratio. In contrast, when messages' interest properties are not as clear, having Social Group to perform message propagation will result better efficiencies compared to Interest Group. The simulation result shows the method this paper proposed will be able to find an appropriate message dissemination group for messages being dispersed, achieving a satisfactory message dissemination ratio. Further, if the data collected were any different, situations describing message propagation in different environments would be possible. This paper proposed a message propagation model based on virtual contact of personal behavior in DTN. Future work can focus on models developed with other behavior pattern groups in varied environments.

5. Conclusions and Future work

This research, as the continuation of previous work of NCCU TRACE DATA, proposes an approach to propagate messages via user behaviors in a virtual environment. We believe users' encounters on social media platforms are just like the case of Delay

Tolerant Network. Both of them depend the mechanism of store-carry-and-forward. This research takes users' online histories on social media platforms and their hobbies when using the platforms into consideration. By conducting a simulation resembling users' real-world activities on social media platforms, it can be clearly told from the simulation result that if a message is to be propagated in a campus, relying on a group of people sharing similar interests to disperse them will surely result in more effective coverages.

We also point out two future research directions. The first is the integration of message propagation via physical and virtual contact. In addition to users' geographical location encounters with other users, it is possible to adopt users' message exchanges with other users on social media platform in a way that both physical and virtual contact are employed are the same time to propagate messages in a more realistic way. The second is to upscale the simulation model of message propagation. So far this paper has come up with a message propagation model in a campus. If the dissemination model is to be verified in different environments, only the data of to-be-verified environment are required to be collected as input of the presented simulation in order to find out the dissemination model for users in different environments.

References

1. A. Mtibaa, M. May, M. Ammar, and C. Diot.: PeopleRank: Combining Social and Contact Information for Opportunistic Forwarding. In Proc. IEEE Infocom 2010 Mini Conference, 1-5(2010)
2. Bulut. E, Szymanski, B.K.: Friendship Based Routing in Delay Tolerant Mobile Social Networks, in Global Telecommunications Conference (GLOBECOM) (2010)
3. E. M. Daly and M. Haahr.: Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs. in Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing(MobiHoc'07), Montreal, Quebec, Canada, 32-40 (2007)
4. E. M. Daly and M. Haahr.: Social Network Analysis for Information Flow in Disconnected Delay-Tolerant MANETs. IEEE Transactions on Mobile Computing, vol. 8, no. 5:606-621(2009)
5. K. Jahanbakhsh, G.C. Shoja, V. King.: Social-greedy: A Socially-Based Greedy Routing Algorithm for Delay Tolerant Networks. MobiOpp'10: Proceedings of the Second International Workshop on Mobile Opportunistic Networking, ACM, New York, NY, USA, 159–162 (2010)
6. Kopecky M., VojtasP.: Visual E-Commerce Values Filtering Frameworkwith Spatial Database metric. Computer Science and Information Systems (ComSIS) 17(3):983–1006 (2018)
7. Khattak Hasan Ali, Ameer Zoobia, Din UdIkram, Khan Muhammad Khurram.: Cross-layer design and optimization techniques in wireless multimedia sensor networks for smart cities.Computer Science and Information Systems (ComSIS) 16(1):1–17 (2019)
8. Mei, A., Morabito, G., Santi, P., Stefa, J.: Social-Aware Stateless Forwarding in Pocket Switched Networks. In: Proc. IEEE Infocom, Mini Conference (2011)
9. M. C. Chuah. "Social Network Aided Multicast Delivery Scheme for Human Contact-Based Networks." In Proceedings of the 1st Workshop on Simplifying Complex Network for Practitioners (Simplex) (2009)
10. N. Eagle and A. Pentland.: Reality mining: sensing complex social systems. Personal and Ubiquitous Computing, Vol 10(4):255–268 (2006)

11. N. Eagle, A. Pentland, and D. Lazer.: Inferring Social Network Structure using Mobile Phone Data. Proceedings of the National Academy of Sciences (PNAS),106(36), 15274-15278 (2009)
12. P. Hui, J. Crowcroft, and E. Yoneki.: Bubble Rap: Social-Based Forwarding in Delay Tolerant Networks. in Proc. ACM MobiHoc, 241–250 (2008)
13. P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, and C. Diot.: Pocket Switched Networks and the Consequences of Human Mobility in Conference Environments. in WDTN ’05: Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking (2005)
14. Spyropoulos, T., Psounis, K., AND Raghavebdra, C. S.: Spray and Wait: An Efficient Routing Scheme for Intermittently Connected mobile networks. In proc. WDTN ’05, ACM Press, 252–259(2005)
15. S.C. Nelson, M. Bakht, R. Kravets.: Contact–Based Routing in DTNs. CS, University of Illinois at Urbana-Champaign, USA, In Proc. INFOCOM, April (2009)
16. Tsai, Tzu-Chieh, Chan, Ho-Hsiang.: NCCU Trace: social-network-aware mobility trace. IEEE Communications Magazine, vol. 53, no. 10: 144-149 (2015)
17. Tsai, Tzu-Chieh, Chan, Ho-Hsiang, Han, Chien Chun, Po-Chi Chen.: A Social Behavior based Interest-Message Dissemination Approach in Delay Tolerant Networks.inProc. Int. Conf. Future Netw. Syst. Secur.:62–80 (2016)
18. Ying Zhu, Bin Xu, Xinghua Shi, and Yu Wang.:A Survey of Social-based Routing in Delay Tolerant Networks: Positive and Negative Social Effects. IEEE Communications Surveys & Tutorial, Issue: 9 (2012)
19. Infocom’06. Retrieved June 5, 2013, from http://crawdad.cs.dartmouth.edu/meta.php?name=cambridge/haggle#N10116
20. Zhensheng Zhang.: Routing in Intermittently Connected Mobile Ad Hoc Networks and Delay Tolerant Networks: Overview and Challenges. IEEE Communications Surveys & Tutorials, 8(1):24–37 (2006)

Ho-Hsiang Chan received his Master's degree from Shih Hsin University, Taiwan, in 2010. He is currently working toward his Ph.D. degree in the Department of Computer Science at National Chengchi University. His research interests are delay-tolerant-networks and wireless communication technology. His recent focus is message propagation based on people's behavior.

Tzu-Chieh Tsai received his BS and MS degrees both in Electrical Engineering from National Taiwan University in 1988, and from University of Southern California in 1991, respectively. After that, he got a PhD degree in Computer Science at UCLA in 1996. During 2005/08~2008/07, he was the Chair of Department of Computer Science, at National Chengchi University (NCCU), Taipei, Taiwan. During 2015/09~2018/11, he serves as Vice President of Office of Research & Development at NCCU. Currently, he is a professor in Computer Science Department at NCCU. He has served as the technical committee for several international conferences, and as one of editorial board for "International Journal on Advances in Networks and Services". He is also the CEO of NCCU Pepper &IoT Lab in the NCCU Fintech Center. His recent research works include Ad hoc Networks, P2P Data Management and Computing, Delay Tolerant Networks, 5G MEC Smart Computing, and Digital Performing Arts.

Received: January 28, 2020; Accepted: November 15, 2020.