Socio-Geo: Social Network Routing Protocol in Delay Tolerant Networks

Seungbae Kim, Mario Gerla
Department of Computer Science, University of California, Los Angeles
{sbkim, gerla}@cs.ucla.edu

***Abstract*—Social based routing has emerged as one of the most efficient routing solutions for Delay Tolerant Networks. It opportunistically relays data to more sociable nodes that have a higher probability of meeting the destination. Many researchers have tried to find the most appropriate metrics that can reflect the real world, such as frequency, freshness, and/or centrality. However, these metrics are mainly used to represent sociality between nodes without considering the actual social relationship between people. In this paper, we propose a novel social network routing protocol that exploits the human friendship information in order to perform routing. First, by collecting data from online social network services, we generated a new dataset trace that includes real social relation information in addition to users' mobility patterns. With the geolocation information that incorporates social relationship information, our protocol can improve the probability estimate of encountering destinations and deliver data packets.**

I. Introduction

Delay or disruption tolerant networks (DTNs) [1][2] have recently received much attention from both industry and academia due to the various applications of these networks such as mobile sensor networks, space communications, and military operations. In most of DTN scenarios, nodes are mobile and can communicate with each other with wireless connections, so that end-to-end paths cannot be sustainable and lead to large transmission delay due to the highly dynamic network connectivity. For this reason, traditional adhoc networking approaches [3][4] cannot be directly applied to DTNs since they need to establish end-to-end paths completely.

Social based routing strategies [9][17] have emerged as one of the most efficient routing solutions for DTNs, especially for mobile social networks [5]. With the concept of 'carry-and-forward' data dissemination, each node relays data to more sociable nodes that have a higher probability of meeting the destination whenever the nodes opportunistically encounter each other. In the last decade, several social based routing protocols such as BubbleRap [6], PeopleRank [7], and SocialCast [8] have been proposed. Their common goal is to construct a high-quality social graph exhibiting human behaviors by observing interactions between nodes. The early studies simplified the social graph as a contact graph; however, many researchers have tried to find the most appropriate and accurate social based metrics which can reflect the real world considering the human behaviors, such as combining of meeting frequency, freshness, and/or centrality.

However, such node-based social metrics cannot completely capture the human behaviors because they miss one important factor, the actual social relationship between people. In other words, their social metrics are considered as the sociality between two nodes. For example, if a node frequently encounters the other node, then these two nodes may have a high sociality value. In real life, however, the number of interactions does not necessarily lead to a social relationship between two people.

There are two reasons to consider the human social relationships in the social based DTN research. First, people spend much more time to communicate with their friends or family members than a stranger. In previous DTN research [6][7][8] used networks that a random node creates a message to a random destination node. In contrast, if we use the social relationship, then nodes will generate more messages to their friends. Consequently, researchers can evaluate their social metrics or routing protocols more accurately on top of the social networks. Second, we can define and use a new type of social metrics if we have such social relationship information. For example, we can state that people already know their friends' home, workplace, and frequently visited places. This knowledge can be used for message forwarding in the network.

Therefore, in this paper, we propose a method to build a human trace dataset that includes social relationship information between users in addition to users' contact information deriving from mobility patterns. Instead of distributing GPS devices to capture the human traces, we collected geolocation information from one of the largest online social network services, Instagram. From this service, we can gather both friend relationships between users and users' frequently visited places that are provided from their social media. By using this method, we created a dataset that contains 80 nodes with their own mobility patterns and 20 groups of friends. We also propose a novel social based routing protocol, named Socio-Geo, which exploits the social geolocation information to perform routing. Our routing protocol uses nodes' mobility patterns and encountering counts to decide to relay the message to an encountered node. Additionally, Socio-Geo limits the number of message copies to maximize the delivery ratio and to keep low network overhead. We conducted simulations on our dataset and the results show that the proposed routing protocol achieves a higher delivery probability than other DTN routing protocols with relatively low network overhead.

II. Real Life Tracing Dataset

Since DTN research started, many DTN researchers have used human trace datasets from the *Cambridge Haggle project* [10] or the *MIT Reality Mining project* [11]. These datasets contain the human traces that were generated by experimental devices carried by the experiment participants. In most social-based DTN routing studies, the researchers have used these trace datasets because their routing metrics are calculated as sociality between nodes such as node centrality or meeting frequency. However, if some researchers propose routing protocols that exploit actual social relationships between people, then these datasets have not sufficient information.

The social relationship between people should be considered in some applications running on mobile social networks because it provides new types of information that could be a critical routing metric. In the real world, for example, people usually communicate with their friends or family members who may or may not be close in proximity. Additionally, people tend to follow a movement pattern in their daily activities, and this movement pattern is well known by their friends and relations in many cases. Therefore, to evaluate performance of routing protocols that exploit sociality between people, we need to create a new dataset that includes the social relationship between people and their movement patterns.

To obtain real-life tracing data, we collected data from one of the largest online social network services, Instagram. In this service, users share snapshots of their daily life with location information, and make a social relationship by following another user. We built a new tracing dataset on top of these two types of information: location and social relationship. First, we collected location information from social media. By collecting a series of locations from a user, we made a mobility pattern of the user. Second, we also collected users' follower IDs to obtain the social relationship information. We assume that a follower is a friend of the user, and they know each other as a friend or a relation. We generated followers' mobility patterns as well in our dataset.

In the dataset, we limited the scope of the region as Los Angeles. To this end, we first downloaded photos including its' user IDs that were taken at UCLA. We filtered out some users who have many locations out of LA by checking their social media. Then we collected all social media with locations in LA from each user, and defined this series of locations as a mobility pattern of the user. We also collected the users' follower IDs as social relationship data and their mobility patterns. Our dataset contains 80 users with their mobility patterns and social relationship information between them.

III. Social Geolocation (Socio-Geo) Routing Protocol

A. Socio-Geo Routing Metrics

We define two social routing metrics: **encounter probability** and **encounter count**. The encounter probability represents the chance to meet a destination node and it is calculated by comparing two mobility patterns of a node itself and the destination node. For example, two nodes, A and B, have a set of places, $P_A = \{a_1, a_2, \ldots, a_n\}$ and $P_B = \{b_1, b_2, \ldots, b_m\}$, respectively, as their movement models, and their transmission range is R_{TX}. To find common places of two nodes, we need to compare each pair of points from two movement models. If the distance between two points is less than the transmission range, these two points are in a common place. The number of common places can be formulated as an equation (1).

$$compare(a_i, b_j) = \begin{cases} 1, & \text{if } distance(a_i, b_j) \leq R_{TX} \\ 0, & \text{otherwise} \end{cases}$$

$$common(P_A, P_B) = \sum_{i=1}^{n} \sum_{j=1}^{m} compare(a_i, b_j) \tag{1}$$

Then, the encounter probability is calculated by dividing the number of common places by the total number of possible pairs of two points.

$$enPr(A, B) = \frac{common(P_A, P_B)}{n \times m} \tag{2}$$

Figure 1 shows mobility patterns of three nodes–A, B, and C– and places for each node are depicted in a different shape. The dotted circle indicates the transmission range of the node. The encounter probabilities

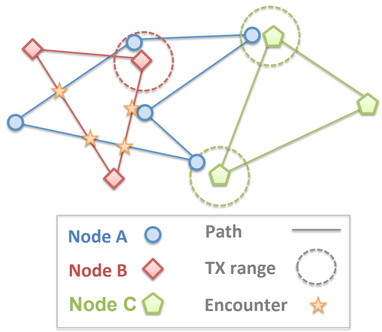

Fig. 1: Mobility patterns of three nodes

of each node pair are $enPr(A,B) = \frac{1}{15} = 0.067$, $enPr(A,C) = \frac{2}{15} = 0.134$, and $enPr(B,C) = \frac{0}{9} = 0.0$. It seems that node C has a higher chance to forward a message to node A than node B if we only see the encounter probability. However, node B can possibly encounter node A when it moves from one place to the other places. The possible encounter points are shown as a star in the figure. Though the encounter probability is low since node B has only one common place, it may meet node A more frequently than node C. That is because node B has four encounter points whereas node C does not have any encounter point except two common places.

To consider the chance of encounter on the way to the other places, we use encounter count as the second socio-geo routing metric. The encounter count is simply the number of meetings with another node. Both the encounter probability and the encounter count can be updated in an intermediate node without exchanging further information except the mobility pattern of a destination node.

$$enCount(A,B) = \text{Number of encounters } (A,B) \quad (3)$$

B. Forwarding Strategy

In this subsection, we describe the message forwarding strategy in Socio-Geo. Socio-Geo works on two sequential phases: **message dissemination** and **socio-geo forwarding**. In the message dissemination phase, a node aggressively transfers message copies to any available nodes until the node holds only one copy of the message. Then it moves into the socio-geo forwarding phase to selectively give the message copy to the other nodes that have a higher chance to meet the destination node than that of the node itself. The first phase helps the message be transferred to the other nodes that have high socio-geo values for the destination node. For example, if we do not use the first phase, only the source node tries to find other nodes with higher social values. However, when we use the first phase, a number of nodes that received the message copy can also help the source node find other nodes with higher social values to deliver the message.

Algorithm 1: Message forwarding algorithm

```
M = message; N = node that carries the message M;
E = encountered node; D = destination node;
M.c = number of copies of M;
α = encounter count threshold;
while N has more message in the buffer do
    M = next message in the buffer;
    if E is D then
        forward the M;
    else if E already has M then
        do not forward;
    else
        if M.c>1 then
            M.c = M.c / 2; then forward the M;
        else if M.c=1 then
            if enPr(E, D)>enPr(N, D) then
                forward the M;
            else if enCount(E, D)>α then
                forward the M;
            end
        end
    end
end
```

Algorithm 1 is the pseudo code of the message forwarding algorithm. In the first phase, a node forwards the half of the message copies to any encountered node without checking the socio-geo routing metrics. The message will be spread to the network quickly when a source node creates the message. Once the number of nodes that have the message copy increases in this phase, the chance to encounter the destination node or nodes that have high socio-geo values will also increase in the second phase.

In the second phase, a node forwards a message to an encountered node only if the encountered node has a higher encounter probability for a destination than that of the node itself or has a larger encounter count value than a threshold value α that is a certain number of times of encountering. When the node forwards the message, the number of the message copies is not changed from one on both the node and the encountered node. Therefore, the number of message copies in the network can keep increasing if newly encountered nodes have higher socio-geo metric values.

TABLE I: Simulation environment

Number of nodes	160 (80 nodes from the dataset + 80 random nodes)
Movement model	Mobility pattern from the dataset + Random way point
Simulation map size	1 km $\times$ 1 km (1:15 scale)
Message size / TTL	500 KB or 1 MB / 30 minutes
Buffer size	10 MB
Node speed	0.5 - 3 km/h
Tx range / speed	10 meters / 250 KBps
Simulation time	5 hours

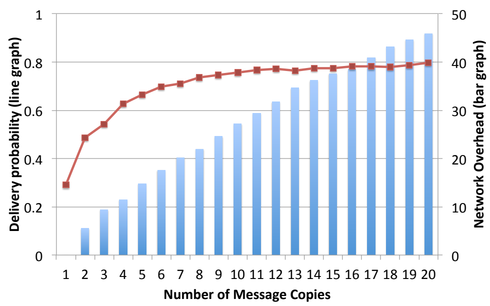

Fig. 2: Performance on different number of copies

IV. Simulation

A. Simulation Environment

To evaluate the performance of the routing protocols, we use ONE simulator [15] that is for the opportunistic network environment simulation. Table I shows default input parameters for the simulation. The dataset includes the 80 nodes with 20 friends groups and each node has a unique movement pattern that is composed of multiple places. A message is created by a node from these 80 nodes, and the destination node of the message is one of friend nodes of the source node. We also put additional 80 nodes that have random way point movement patterns because we believe that there are many people around us that we do not know who they are and also their mobility patterns. These random nodes work as intermediate nodes that carry messages to destination nodes and do not create any message. The map of Los Angeles city is scaled down for the simulation, and all nodes move with various speeds.

B. Message Dissemination Phase

The Socio-geo router disseminates message copies in the first forwarding phase. If the initial number of message copies is too small, nodes will quickly turn into the socio-geo forwarding phase and it causes the message to be disseminated slowly to the network. If we use a too large message copy value, however, too many copies will be spread out and it triggers a severe network overhead and a large number of packet drops. Therefore, to determine the initial number of message copies is important to have sufficient number of copies to rapidly spread out the message as well as to keep the network not to be over flooded. To find an optimal number of message copies in our simulation environment, we conduct simulations on a different number of initial message copies.

Figure 2 shows the message delivery probability and the network overhead. The message delivery probability is calculated as the total number of delivered messages over the total number of messages, and the network overhead is the number of delivered messages subtracted from the number of message relays divided by the number of delivered messages. The delivery probability is only about 30% when initial message copy is 1. The delivery probability increases rapidly when the number of copies meets a certain number, and then it increases slightly beyond this point. That is because the certain number of message copies is sufficient to disseminate the message in the network. Also, the router can still forward the message in the second phase by using the social metrics even though the first phase is done. However, the network overhead grows linearly while the message copies increase because the message is still relayed by many nodes in the second phase. Therefore, we need to pick a number of message copies that leads to having high delivery probability while minimizing the network overhead. This value is highly dependent on the density of the network, the message TTL, and the buffer size. So that we need to propose a method to find an optimal value by a node itself as our future work.

C. Social-geo Phase

The Socio-Geo routing protocol uses two social metrics, the encounter probability and the encounter count, in the second phase of message forwarding. Since the second phase helps to achieve high delivery probability, we analyze the effectiveness of the social metrics on the performance and understand the relationship between two metrics.

We define two social metrics because the encounter probability does not consider the contacts on a path. As we introduced an example in section III-A, a node can have high encounter count value even though it has a very low encounter probability because this node may have a few common places with a destination node but it has many contact points on its routes (figure 1). To attest that this example is true, we compare values of the two metrics of all pairs of node sets. Figure 3 shows the relationship between two social-geo metrics. A single point in the figure indicates the encounter probability and the encounter count value of any

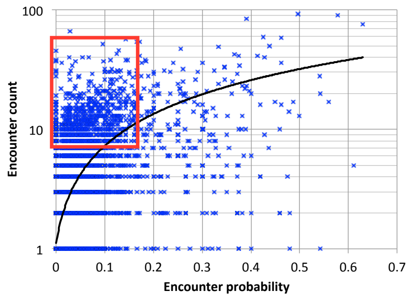

Fig. 3: Encounter probability vs. Encounter count

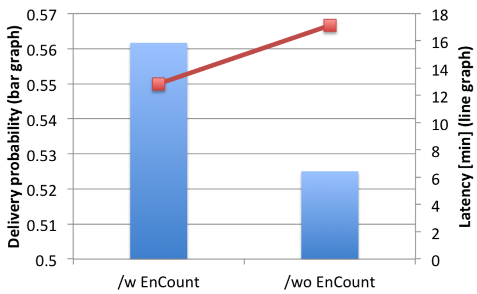

Fig. 4: Routing performances on different scenarios

two nodes. Though some nodes with high encounter probability tend to have high encounter count values, many points are scattered in the highlighted box on the graph. This result represents that a large number of nodes actually meet its destination nodes frequently, but have less than 10% of the encounter probability. Thus, the encounter count should be used for message forwarding in the socio-geo phase.

To see the effectiveness of the encounter count on the routing performance, we conduct simulations on two different cases; In the first case, we use both social metrics, and we only use the encounter probability in the second case. Figure 4 shows routing performances of the two cases. The result graph shows that we can achieve higher delivery probability and lower latency if we use two routing metrics rather than one. When we use encounter count, the delivery probability increased by 7% and the delivery latency decreased by 25% from 17 minutes to 12 minutes.

D. Performance Evaluation

We evaluate routing performance of Socio-Geo with other routing protocols: MaxProp [12], Spray-and-Wait [13], and Epidemic [14]. For simulation, we used same parameters that are shown in table I. We changed the message TTL to evaluate the performance of the routing protocols. Additionally, we set the number of message copies as 20 for both Socio-Geo and Spray-and-Wait protocols to see the saturated delivery probabilities.

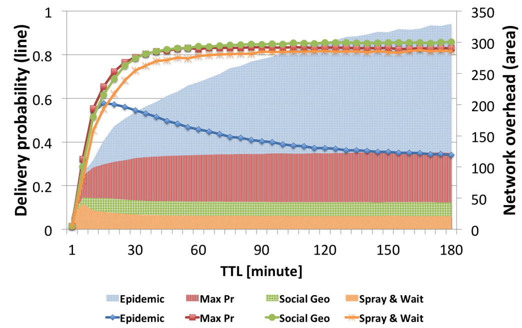

Fig. 5: Delivery probability and network overhead

Figure 5 shows the delivery probability (line graph) and network overhead (area graph). Socio-Geo, Max-Prop, and Spray-and-Wait protocols have around 80% delivery ratio while the performance of the other two protocols degraded when TTL increases. Socio-Geo has 5% larger average delivery probability than that of Spray-and-Wait. That is because Socio-Geo can spread more message copies to other nodes even if a node only has a single copy of the message.

The overhead of Epidemic increase linearly when TTL increases. Since Epidemic routing protocol relays too many messages, the delivery probability decreases because of packet drops. Other three routing protocols have stable ratio of the number of relayed messages over the number of delivered messages. However, Max-Prop has about three times larger network overhead than that of Socio-Geo because it uses the flooding method on the encounter estimation values. Socio-Geo can keep the lower overhead than MaxProp because of the limited number of message copies in the first message forwarding phase. The network overhead is about 20 for Spray-and-Wait since it relays only 20 copies to the network.

Overall, Socio-Geo routing protocol has high delivery probability and low network overhead compared with other routing protocols.

V. Related Works

To exploit the sociality between nodes in DTNs, several social based routing protocols have been developed. BubbleRap [6] uses both node centrality and node community knowledge to forward messages. It detects hierarchical communities in the network and calculates centrality for each community. A message is forwarded through the hierarchical communities until it reaches the destination node. SocialCast [8] assumes that nodes tend to interact frequently if they share the same interest. It predicts the probability of node interactions and change of the degree of connectivity to make a decision whether forward a message or not. PeopleRank [7] makes a social graph and ranks the nodes in the graph by using the social relationships based on the PageRank algorithm. It forwards a message to a node with a higher PeopleRank value which represents a more sociable node in the network. Though they proposed various social properties among nodes, these properties have a limitation in describing the social relationship among people because of the lack of information such as human mobility pattern with friend relationship.

Human mobility pattern has been shown to be useful in link prediction and routing in the mobile opportunistic networks [16]. In [16], the authors found that the similarity between two individuals' movements are strongly correlated with their proximity in the social network. Thus, mobility pattern can be used to predict the formation of new links. [17] shows the mobility patterns of mobile devices are closely related to users' social relationships and behaviors, which lays a foundation for applying mobility pattern to routing protocols. However, the trace dataset used in their research is difficult to reflect daily mobility patterns because it was captured on a campus WiFi network which is a limited space.

VI. Conclusion

In this paper, we introduce a social network routing protocol, named Socio-Geo, that exploits human mobility patterns to predict the probability of encountering destination node. Socio-Geo has the two-phased message forwarding strategy which is working on the number of message copies to maximize the number of successful transmissions. We also create a human trace dataset that includes the social relationship between people as well as their mobility patterns by collecting data from an online social network service. On top of the dataset, Socio-Geo achieves higher delivery ratio and lower network overhead than typical DTN routing algorithms.

The results presented in this paper also lead to many interesting directions for future research. First, our perspective on the social network properties over DTNs expands the possibility of bringing new types of social metrics to routing algorithms. Besides to use the users' mobility patterns, we can also exploit famous locations as rendezvous points to forward messages, or build a location graph for each node to find a path to reach a destination node. Second, the human mobility dataset can be improved with other factors. A location may have a weight value so that people stay longer or visit more frequently to the location with higher weight value than low weight value.

References

[1] K. Fall, *"A delay-tolerant network architecture for challenged internets,"* Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications. ACM, 2003.

[2] S. Jain, *et al.* *"Routing in a delay tolerant network,"* Vol. 34. No. 4. ACM, 2004.

[3] C. E Perkins, *"Ad hoc networking,"* Addison-Wesley Professional, 2008.

[4] J. Broch, *et al.* *"A performance comparison of multi-hop wireless ad hoc network routing protocols,"* In Proceedings of the 4th annual ACM/IEEE international conference on Mobile computing and networking, 1998.

[5] M. Xiao, *et al.* *"Community-aware opportunistic routing in mobile social networks,"* IEEE Transactions on 63, no. 7 (2014).

[6] P. Hui, *et al.* *"Bubble rap: Social-based forwarding in delay-tolerant networks,"* Mobile Computing, IEEE Transactions on 10, no. 11 (2011).

[7] A. Mtibaa, *et al.* *"Peoplerank: Social opportunistic forwarding,"* INFOCOM 2010

[8] P. Costa, *et al.* *"Socially-aware routing for publish-subscribe in delay-tolerant mobile ad hoc networks,"* IEEE Journal on 26, no. 5 (2008).

[9] K. Wei, *et al.* *"A survey of social-aware routing protocols in delay tolerant networks: applications, taxonomy and design-related issues,"* Communications Surveys & Tutorials, IEEE 16, no. 1, 2014.

[10] J. Scott, *et al.* *"CRAWDAD dataset cambridge/haggle, traceset: imote,"* http://crawdad.org/cambridge/haggle/20090529/imote, May 2009.

[11] N. Eagle, *et al.* *"Reality mining: sensing complex social systems,"* Personal and ubiquitous computing 10, no. 4 (2006).

[12] J. Burgess, *et al.* *"MaxProp: Routing for Vehicle-Based Disruption-Tolerant Networks,"* INFOCOM 2006.

[13] T. Spyropoulos, *et al.* *"Spray and wait: an efficient routing scheme for intermittently connected mobile networks,"* ACM SIGCOMM workshop on Delay-tolerant networking, 2005.

[14] A. Vahdat, *et al.* *"Epidemic routing for partially connected ad hoc networks,"* Technical Report CS-2000-06, Department of Computer Science, Duke University, 2000.

[15] A. Kernen, *et al.* *"The ONE simulator for DTN protocol evaluation,"* International conference on simulation tools and techniques. ICST, 2009.

[16] D. Karamshuk, *et al.* *"Human mobility models for opportunistic networks,"* Communications Magazine, IEEE 49, no. 12 (2011).

[17] F. Xia, *et al.* *"Socially aware networking: A survey,"* Systems Journal, IEEE 9, no. 3 (2015).