<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Exploring centrality for message forwarding in opportunistic networks</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Xu, K; Li, VOK; Chung, J</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2010</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/126111">http://hdl.handle.net/10722/126111</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>Creative Commons: Attribution 3.0 Hong Kong License</td>
</tr>
</tbody>
</table>
Exploring Centrality for Message Forwarding in Opportunistic Networks

Kuang Xu*, Victor O.K. Li*, Jaewoo Chung†
*The University of Hong Kong †Massachusetts Institute of Technology

Abstract—In opportunistic networks, centrality characterizes a node's capability to act as a communication hub. In this paper, we provide an in-depth study of choosing effective centrality metrics for message forwarding in bandwidth-limited opportunistic networks. Based on this study, we propose a destination-unaware forwarding algorithm that accounts for the popularity of a node and the contact durations between nodes. We evaluate the algorithm on two experimental human mobility traces. The simulation results show that the proposed algorithm achieves higher system throughput while maintaining a lower forwarding cost compared with several known destination-unaware forwarding schemes.

I. INTRODUCTION

Centrality, one of the important characteristics of network structure, addresses the question, “What is the relative importance of an element within a network?” The answer depends on what we mean by importance. In social networks, centrality may correspond to the authority or popularity of individuals, and it is important for promotion and management of word-of-mouth communications amongst people [23]; in biological networks, centrality may correspond to robust bridging characteristics of nodes, and it is important for the selection of druggable candidates with minimal side effects upon targeting [10]; and in opportunistic (mobile) networks, “centrality” nodes that actively make contacts (of physical co-location) with others may be important for facilitating communication between other nodes [3] [12] [13]. In this paper, we study the problem of utilizing the centrality of nodes for message forwarding in opportunistic networks.

An opportunistic network with human-carried mobile devices is a variant of delay tolerant network (DTN) [9], in which mobile nodes communicate with each other via store-copy-forward, hop-by-hop opportunistic data forwarding. As opposed to traditional communication that relies on an infrastructure and ad hoc networking that assumes contemporaneous paths between communication endpoints, end-to-end connectivity in opportunistic networks is not guaranteed due to the unpredictable node mobility, resource constraints (e.g., battery and storage), and short radio range. However, opportunistic networking applies to many practical scenarios such as communication between rural villages and in undeveloped regions [21], or when the communication infrastructure is destroyed by natural or man-made disasters.

For routing in opportunistic networks, a host of message forwarding schemes have been proposed in the literature [8] [15] [17] [18]. They use historical contact information to predict future contact opportunities, and rely on comparisons between per-node metrics to make forwarding decisions. The recent studies in [3] [13] show that message-forwarding strategies utilizing the social structure of the underlying network enjoy promising performance. One of the concepts they use is centrality, which measures the popularity of a node, or the influence a node has over the spread of information. They assume that popular nodes in a social graph are more efficient to act as communication hubs. Here, the social graph is an aggregated contact (physical co-location) graph over time. However, the centrality of a node in this static graph may not properly represent its actual capability to ferry messages between others since the social graph does not consider the dynamics of node mobility in the underlying temporal network. Along the same lines, [12] proposed an analytical centrality metric for multi-cast in opportunistic networks, based on a Poisson model of the contact process. However, in practice, no forwarding strategy can make use of the past contact information if inter-contact times are homogeneously and exponentially distributed [2]. On the other hand, most utility-based routing strategies that favor nodes inferred with high possibility to contact others in the future, such as a node with more encounters in the past [7] [19] or a node that contacted other nodes more recently [5], do not take into account the durations of contacts. Nevertheless, contact durations are important for message forwarding in practice since they are short in opportunistic networks and the actual network is bandwidth-limited. In this paper, we propose a centrality-based message forwarding (CMF) strategy that takes into consideration both the popularity of a node in the underlying social graph and the contact durations. To account for the temporal dynamics of node mobility, we utilize an aging method to update a node’s status. CMF is destination-unaware, and can be incorporated as a module into destination-aware forwarding algorithms. We evaluate the proposed algorithm on two experimental human mobility traces, and compare its performance with several known destination-unaware forwarding schemes. The simulation results show that CMF achieves higher message delivery throughput while maintaining a lower forwarding cost.

The purpose of this paper is not to propose a new complete message routing algorithm, rather, an in-depth study of choosing effective centrality metrics to characterize the capability of a node to relay messages, so as to provide a working module for a wide range of design choices. In contents to follow, we describe the proposed centrality metrics and message-forwarding strategy in Section II, followed by the performance
II. CENTRALITY-BASED FORWARDING

In this section, we first present the experimental human mobility traces that we utilize for the evaluation of both centrality metrics and forwarding protocols in this paper, followed by the proposed centrality metrics. Then, we present the centrality-based message-forwarding strategy.

A. Experimental dataset

To base our study on realistic settings, we utilize two experimental datasets of human mobility traces collected by two research projects, Haggle [4] at Infocom2006 conference and Reality Mining [6] at MIT campus. In these experiments, responders carry Bluetooth-enabled devices that periodically discover others in the vicinity and log contacts. These datasets are popularly utilized in the literature of opportunistic networking, and cover diverse environments from conference site to university campus, with experimental periods from a few days (Infocom2006) to an academic year (Reality Mining).

**Infocom2006.** In this four-day experiment, there are 98 Bluetooth devices (iMote), with 78 iMotes (short radio range) carried by attendees to the student workshop and 20 stationary iMotes (long radio range) distributed throughout the conference venue (i.e. conference hall, bar, and lift). The Bluetooth scanning rate of nodes is set to once per 120 seconds. For our study, we only consider the short range iMotes carried by people, and the traces we use are from 24 April 2006 8:00am to 25 April 2006 12:00pm. Since the contacts in the raw logs we access are not synchronized (the devices were not started at the same time), we manually synchronize the traces by looking for mutual scans between devices [20].

**Reality Mining.** This project comprises 100 MIT students and faculty carrying smart phones with pre-installed programme over the course of 9 months. The collected information includes call logs, Bluetooth devices in proximity, cell tower IDs. We only use the Bluetooth contact logs for our study. Since the time granularity between consecutive Bluetooth scanings is five minutes, many short contacts were not logged, and the resulting (temporal) network of contacts is quite sparse. In this paper, we select a session of four weeks from 1 March 2005 to 28 March 2005, which does not contain extended holiday periods and the contacts in the traces are relatively stable.

Table I summarizes the post-processed datasets we utilize in this paper. Figure 2 shows the total duration of each node contacting all other nodes, divided by the duration of the entire period under observation. We see highly heterogeneous contact durations in each dataset, leading to very different abilities to ferry messages for others. We are interested in utilizing nodes with better abilities of ferrying messages.

<table>
<thead>
<tr>
<th>Experimental dataset</th>
<th>Infocom2006</th>
<th>Reality Mining</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device type</td>
<td>iMote</td>
<td>Phone</td>
</tr>
<tr>
<td>Duration</td>
<td>28 hours</td>
<td>28 days</td>
</tr>
<tr>
<td>Scanning Granularity</td>
<td>120</td>
<td>300</td>
</tr>
<tr>
<td>Number of participants</td>
<td>77</td>
<td>80</td>
</tr>
<tr>
<td>Number of contacts</td>
<td>23,478</td>
<td>4123</td>
</tr>
<tr>
<td>Pairwise contact freq (per day)</td>
<td>6,878</td>
<td>0.047</td>
</tr>
<tr>
<td>Average contact duration (secs)</td>
<td>216</td>
<td>2281</td>
</tr>
</tbody>
</table>

Table I

**POST-PROCESSED MOBILITY TRACES SUMMARY**

evaluation in Section III. We conclude with suggestions for future work in Section IV.

B. Centrality metric

As discussed in Section I, many centrality metrics are calculated based on the aggregated network contact graph. They do not consider the strength (contact time) of contacts between nodes and the dynamics of node mobility. To effectively represent the probability and capability of a node to contact other nodes, and to ferry messages between them when the transmission capacity of the network is limited, we propose a centrality metric taking into account contact time between participating nodes.

Consider an opportunistic network consisting of $N$ mobile devices carried by people moving around. We denote this temporal network by $G = (V, E)$, and we consider discrete-time model and assume the time unit is 1, such that

- $V$ is the set of participating nodes. We assume the network size is finite, and $N = |V|$.
- Each contact event $(u, v) \in E$, for all pairs $(u, v)$ and all time $t$ at which the contact starts is independent of others.
- All pairs $(u, v)$ are undirected, and each contact event $(u, v)_t$ has a weight $w_{uv_t} > 0$ representing the duration of the contact (contact time).

We also denote by $C_{uv} = \sum_{t<T} w_{uv_t}$ the cumulative contact time between nodes $u$ and $v$ within time $T$.

The mobility characteristics of nodes in opportunistic networks vary within one scenario and across different ones. Here, we consider two typical cases of opportunistic networks, dense network and sparse network. Dense network means, from the perspective of a single node, that a node always has concurrent contacts with multiple peers in the network. In reality, this could be scenarios in which all the participants move about within a relatively small area (e.g. conference venue, department store, etc.) with frequent physical co-location of people. We use the traces of Infocom2006 as a representative of this type of network since we can see (in Section II-A), for more than half of the nodes $u$ in the network, $\sum_{v=1, v\neq u}^N C_{uv}$ is greater than $T$. On the other
hand, sparse network is the case in which overlapping contact events with a single node rarely happen. This could refer to scenarios in which the participants move around in a larger area (e.g. university campus, residential community, etc.), and thus frequent physical co-location of two people takes place with low probability. The network of the Reality Mining project corresponds to this case, and we observe that for all nodes \( u \) in this network, \( \sum_{v=1, v \neq u}^{N} C_{uv}^{T} < T \). We will further explain our classification later.

Since the influence a node has over the spread of information is related to how many different nodes this node can make contacts with, we define a random variable \( A_{uv} \) as

\[
A_{uv} = \begin{cases} 
1 & \text{if node } u \text{ and node } v \text{ meet within } T, \\
0 & \text{otherwise.}
\end{cases}
\]

and let \( D_u = \sum_{v=1, v \neq u}^{N} \frac{A_{uv}}{N-1} \), which represents the number of unique peers a node \( u \) encounters within \( T \), normalized by \( N-1 \). For simplicity, we call this the popularity of node \( u \). In reality, a popular person may refer to a person who frequently visits others, such as a mailman.

**Definition 1.** For a dense opportunistic network of \( N \) nodes, and the popularity of a node \( u \) within \( T \) is \( D_u \), the centrality of \( u \) is defined as

\[
Q_u = D_u \cdot \frac{1}{T} \sum_{v=1, v \neq u}^{N} C_{uv}^{T} \quad (1)
\]

where \( Q_u \) takes into consideration both the popularity of node \( u \) and the cumulative contact duration between \( u \) and all of its contacted nodes within \( T \). We consider the cumulative contact duration since in a bandwidth-limited network, the volume of messages a node can ferry is closely related to the time that the node stays within communication range with other nodes. We carry out a number of message dissemination emulations using Epidemic Routing [24] with finite bandwidth in the Infocom2006 traces, in which every node creates messages destined to all other nodes. We count the number of times a node acts as relays to delivered messages. Fig. 2(a) shows the correlations between three centrality metrics and the number of times a node acts as relays. We can see that the results (represented by dots) of centrality \( Q_u \) form a straight line while those of other two metrics (Greedy\(^1\) and Betweenness\(^2\)) are scattered throughout the figure. This indicates that our proposed centrality metric is more effective at characterizing the ability of a node to ferry messages when considering the limited bandwidth of a network, and that nodes with higher values of \( Q_u \) are more helpful in ferrying messages.

However, for an opportunistic network of low contact rate between mobile devices, in particular a network in which a node does not have concurrent co-locations with multiple other nodes, such that \( \sum_{v=1, v \neq u}^{N} C_{uv}^{T} < T \), Definition 1 can not appropriately reflect the possibility of a node to contact others. In fact, for a node \( u \), the larger the duration of a single contact with a node \( v \) (i.e. \( w_{uv} \)), the less opportunity for node \( u \) to come into contact with other nodes. In this case, we prefer nodes with high rate of change of connectivity with others. We propose the following centrality metric for sparse networks.

**Definition 2.** For a sparse opportunistic network of \( N \) nodes, and the popularity of a node \( u \) within \( T \) is \( D_u \), the centrality of \( u \) is defined as

\[
Q_u = D_u \cdot \frac{2}{T} \sum_{v=1, v \neq u}^{N} \sum_{t<T} \frac{1}{w_{uv}} \quad (2)
\]

The centrality metric \( Q_u \) for sparse opportunistic network takes into account both the popularity of a node and the duration of all single contacts with this node within \( T \). Similar to the case of dense network, we carry out a number of emulations, and the result in Fig. 2(b) indicates that for sparse networks \( Q_u \) performs better than other centrality metrics.

For the classification of the two types of opportunistic networks, [14] showed that when varying the proportion of mobile nodes that is willing to participate in opportunistic forwarding, phase transitions of system performance of message delivery were observed. Since our aim in the paper is to show that node centrality considering contact time has significant effect on the performance of opportunistic communication, we do not investigate in detail how to exactly define the two types of networks, and we defer it as future work.

**C. Message forwarding**

Opportunistic network is by nature non-static, in which the centralities of nodes may change over time. To track a node’s centrality, each node \( u \) maintains two pieces of information, \( S_u \) and \( Q_u \). \( S_u \) represents the utility of centrality for node \( u \) to carry messages, and \( Q_u \) is the centrality of node \( u \) in the current time interval \( n \). Each node is assigned an aging factor \( \tau \), which indicates the aging granularity of a node’s centrality in the past. Node \( u \) periodically updates \( S_u \) to obtain its most recent utility for carrying messages, in the following way:

\[
S_u^{(n)} = \frac{\tau \cdot S_u^{(n-1)} + q Q_u}{\tau + 1},
\]

where \( q \) is a normalization factor. \( S_u \) takes into account all the contact histories up to time interval \( n \). This utility function
needs to be able to forget out-dated status of a node, yet it has to be resilient to the temporal fluctuations. Consider that after time \( n_e \) a previously active node \( u \) becomes stationary (or not contacting any other nodes), then \( S_u \) ages exponentially according to:

\[
S_u^{(n)} = S_u^{(n-1)} \cdot \frac{\tau}{(\tau + 1)} \\
= S_u^{(n_e)} \cdot \frac{(\tau/\tau + 1)^{(n-n_e)}}{(\tau + 1)^{(n-n_e)}} \\
= S_u^{(n_e)} \cdot e^{-\gamma(n-n_e)},
\]

and the half-life of the decay is thus \( \frac{\ln 2}{\gamma} + n_e \).

When the aging factor \( \tau \) is less than 1, the utility \( S_u \) ages rapidly since for each new stage, the influence of the entire contact history is less than that of the new contact status. On the other hand, if \( \tau \) is set greater than 1, the resulting utility will be more resilient to temporal fluctuations as more emphases are put on the history than the current status, yet the influence of past contacts will still fade away over time. The aging factor \( \tau \) and the time scale of the update interval can be adapted in response to the changes in contact patterns. [16] proposed a method to find appropriate values of parameters by considering the time needed for a network to reach steady state and the time the system takes for its predictability values to adapt to a different mobility pattern. The method also applies to our study.

The pseudo-code in Algorithm 1 depicts the basic operations of the centrality-based message forwarding strategy from a node’s perspective, where \( T_u \) represents the update interval parameter. When two nodes meet, they compare their centralities, and the node with lower centrality value replicates the messages it carries and transfers them to the node with higher centrality value. When the two nodes get out of communication range with each other, both of them record the time they were in contact, and update their centralities accordingly. The pseudo-code is for destination-unaware communication scenarios, and it could be easily incorporated as a module into destination-aware algorithms (see Section III).

### III. Evaluation

In this section, we evaluate the performance of our proposed centrality-based message-forwarding strategy. We implement a set of message-forwarding algorithms in a trace-driven simulator\(^3\), utilizing the human mobility traces of Infocom2006 and Reality Mining (see Section II-A) as the inputs. We compare the performance of these message-forwarding algorithms based on the following two metrics:

**System throughput.** We study the cumulative number of messages that reach their destinations. Maximizing the throughput is the goal of any message-forwarding algorithm.

**Forwarding cost.** Forwarding cost is the number of message replicas generated by nodes in order to deliver a message. Here, we study the expected number of replicas of delivered messages, which also indicates the required number of forwards (or relays) of a message and its replicas for one of them to reach the destination. It is desirable for a message forwarding algorithm to minimize the forwarding cost so as to reduce network resource (e.g. bandwidth) consumption and save battery power and storage of mobile devices.

Each of the forwarding algorithms we select for comparison is used as a module in some known algorithms in the literature. All of them has the same properties: destination-unaware, working in a distributed manner (with local information), and not requiring future knowledge of the system. The forwarding decisions in all the algorithms are made based on comparisons of per node metrics, which are proposed to predict the possibility of a node to contact others.

- **Encounter rate (Frequency).** Node \( u \) forwards a message to node \( v \) upon contact if \( v \) made more total contacts with all other nodes in the past. This algorithm is used as a module in [19].
- **Last encounter time (LETime).** Node \( u \) forwards a message to node \( v \) upon contact if \( v \) contacted any other node more recently than did \( u \). This concept is explored by FRESH [5] for destination-aware forwarding.
- **Egocentric betweenness (EgoB).** This is a variant of betweenness centrality utilized in SimBet [3], which is calculated based on a node’s ego network and accounts for all the past contacts of a node. In our implementation, node \( u \) forwards a message to node \( v \) if \( v \) has higher value of egocentric betweenness than \( u \).
- **Epidemic** [24]. Node \( u \) always forwards a message to \( v \) upon contact unless \( v \) already has a replica of this message in its buffer. Epidemic forwarding always finds the best possible paths to a destination and therefore achieves the best delivery performance in throughput. However, it also generates the highest forwarding cost. Since all

\(^3\)Our simulator does not consider transmission contentions and scheduling policies in the network MAC layer since they do not affect the results of performance comparisons of different algorithms used in this paper.
the algorithms we study are destination-unaware, we do not intend to compare their performance with that of Epidemic forwarding, but we include its performance in the results to serve as the delivery upper bound of the system.

- **Direct.** Direct transmission is the simplest forwarding strategy where node $u$ forwards a message to node $v$ upon contact only if $v$ is the destination of this message. We include its performance in the results to serve as the delivery lower bound.

We start the simulations with a warm up period, after which each node sends a message to every other node in the system. The transmission capacity of the system is set to one message per 50 seconds. For all algorithms in the simulations, we assume there is no limit of message replicas, such that a node always replicates a message and keeps it in the buffer when forwarding this message to others. We also assume each node has infinite buffers and carries a message replica until the end of the simulation if this message is not delivered, but upon delivering a message, all replicas of the message are deleted from the system. In practice, the notification of delivered messages can be handled with certain acknowledgment mechanisms [1]. Since our study scenario is for bandwidth-limited network, we define the order in which messages in the buffer are transmitted:

- Forward the message with the lowest forwarded hop count. This assigns a higher priority to new messages [1].
- Forwarding the message with the largest forwarded hop count. This assigns a higher priority to older messages.
- Forward a random message from the buffer.
In the results, we find the message-forwarding algorithms that use the first transmission policy above always have slightly better performance than others. Due to space limitations, we do not present the performance based on the other two transmission policies.

Figure 3 shows the message delivery performance of different algorithms based on the Infcom2006 traces. We set the time interval to 2 hours and $\tau = 0.5$ in CMF during the simulations. We observe two transitions (at $5 \times 10^4$ and $9 \times 10^4$ seconds) in the performance of all the algorithms, which are reasonable due to the fact that a day ends (around 10:00pm) and the next day begins (around 9:00am). Figure 3(a) clearly shows that CMF outperforms the other three destination-unaware algorithms (i.e., Frequency, LETime, and EgoB) in terms of the system throughput. In particular, the improvement is about 3% at the end of the simulation. Though the throughput improvement that CMF achieves is not significant, it maintains a low forwarding cost. From Figure 3(b) we can see CMF maintains a similar forwarding cost with Frequency during the entire simulation while achieving up to 8.6% and 19% less cost compared with EgoB and LETime, respectively.

Figure 4 presents the simulation results based on the Reality Mining traces, where the time interval and $\tau$ are set to 6 hours and 0.5, respectively, in CMF. The figure exhibits a staircase effect in the system throughput and the forwarding cost (it is presented more clearly in the curve of Epidemic forwarding). This corresponds to the alternations of day and night. The phases of the larger flat portions (at $0.4 \times 10^6$, $1 \times 10^6$, $1.6 \times 10^6$, and $2.2 \times 10^6$ seconds) in the performance curves correspond to the weekends, when most of the participants stay at home). We can clearly observe in Figure 4(a) that CMF achieves the highest delivery throughput among the four algorithms, with improvements up to 12.3% (of EgoB) at the end of the simulations. Again, Figure 4(b) shows our proposed CMF achieve lower forwarding cost than Frequency, LETime, and EgoB throughout the simulations, with up to 31.4% less cost (compared with LETime) at the end of the simulations.

IV. CONCLUSION

In this paper, we study the problem of choosing effective centrality metrics for message forwarding in opportunistic networks. We propose a destination-unaware message-forwarding strategy that takes into consideration both the popularity of a node in the underlying social graph and contact durations. We compare its message delivery performance with several known destination-unaware forwarding schemes, and the simulation results show that it achieves higher message delivery throughput while maintaining a lower forwarding cost, which indicates that the proposed centrality metric is more effective in characterizing the capability of a node to relay messages. However, forwarding schemes that tend to use only the most preferred nodes will cause unfair loading [22] in the system (in our simulations, we find that 63% of the traffic is handled by only 10% of the nodes). In addition, due to the resource constraint of human-carried mobile devices, a person may not be totally altruistic to relay messages for others (even if he has high capability to contact others) [25]. It would be interesting to investigate these issues.

ACKNOWLEDGEMENT

This research is supported in part by the University of Hong Kong Strategic Research Theme of Information Technology.

REFERENCES