Social Network Analysis for Routing in Disconnected Delay-Tolerant MANETs

Elizabeth Daly and Mads Haahr
Distributed Systems Group,
Computer Science Department,
Trinity College Dublin
Dublin, Ireland
{elizabeth.daly,mads.haahr}@cs.tcd.ie

ABSTRACT

Message delivery in sparse Mobile Ad hoc Networks (MANETs) is difficult due to the fact that the network graph is rarely (if ever) connected. A key challenge is to find a route that can provide good delivery performance and low end-to-end delay in a disconnected network graph where nodes may move freely. This paper presents a multidisciplinary solution based on the consideration of the so-called small world dynamics which have been proposed for economy and social studies and have recently revealed to be a successful approach to be exploited for characterising information propagation in wireless networks. To this purpose, some bridge nodes are identified based on their centrality characteristics, i.e., on their capability to broker information exchange among otherwise disconnected nodes. Due to the complexity of the centrality metrics in populated networks the concept of ego networks is exploited where nodes are not required to exchange information about the entire network topology, but only locally available information is considered. Then SimBet Routing is proposed which exploits the exchange of pre-estimated 'betweenness' centrality metrics and locally determined social 'similarity' to the destination node. We present simulations using real trace data to demonstrate that SimBet Routing results in delivery performance close to Epidemic Routing but with significantly reduced overhead. Additionally, we show that Sim-Bet Routing outperforms PRoPHET Routing, particularly when the sending and receiving nodes have low connectivity.

Categories and Subject Descriptors

C.2.1 [**Network and Architecture Design**]: Wireless Communication ; Store and forward networks.

General Terms

Algorithms, Design, Performance.

Keywords

Delay & Disruption Tolerant Networks, MANETs, Sparse Networks, Social Network Analysis, Ego Networks.

1. INTRODUCTION

A Mobile Ad hoc Network (MANET) is a dynamic wireless network with or without fixed infrastructure. Nodes may move freely and organise themselves arbitrarily [5]. Sparse Mobile Ad hoc Networks are a class of Ad hoc networks where node density is low, and contacts between the nodes in the network do not occur very frequently. As a result, the network graph is rarely, if ever, connected and message delivery must be delay-tolerant. Traditional MANET routing protocols such as AODV [33], DSR [17], DSDV [32] and LAR [19] make the assumption that the network graph is fully connected and fail to route messages if there is not a complete route from source to destination at the time of sending. For this reason traditional MANET routing protocols cannot be used in sparse MANETs. To overcome this issue, node mobility is exploited to physically carry messages between disconnected parts of the network. These schemes are sometimes referred to as *mobility-assisted routing* that employ the *store-carry-and-forward* model. Mobility-assisted routing consists of each node independently making forwarding decisions that take place when two nodes meet. A message gets forwarded to encountered nodes until it reaches its destination.

In this paper we propose the use of social network analysis techniques in order to forward data in a disconnected delay-tolerant MANET. Social networks exhibit the small world phenomenon which comes from the observation that individuals are often linked by a short chain of acquaintances. The classic example of a small world is Milgram's 1967 experiment, where 60 participants in Nebraska were asked to forward a letter to be delivered to a stockbroker in Boston [27]. The letters could only be forwarded to someone whom the current letter holder knew by first name and who was assumed to be more likely than the current holder to know the person to whom the letters were addressed. The results showed that the median chain length of intermediate letter holders was approximately 6, giving rise to the notion of 'six degrees of separation'. A more recent experiment is that conducted by Hsu and Helmy who performed an analysis on real world traces of different university campus wireless networks [15]. Their analysis found that node encounters are sufficient to build a connected relationship graph, which is a small world graph. Therefore social analysis techniques may be suitable for a number of disconnected delay-tolerant MANETs.

A number of solutions for routing in disconnected networks have been proposed based on a node's observed encounters where data is routed to nodes with the highest 'probability to deliver' to a destination node. Such metrics are typically based on either direct or indirect observed encounters [4, 6, 13, 18, 24].
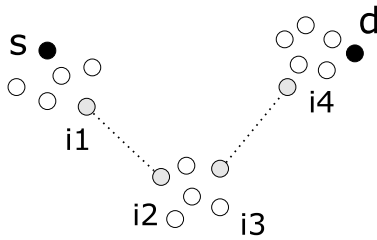
Figure 1: Disconnected Clusters

Some networks may consist of cliques where metrics based on direct or indirect encounters may not find a suitable carrier for the message. Consider three disconnected clusters in figure 1. Source node s wishes to send a message to destination node d. However node s is involved in a highly cliquish cluster in which none of the nodes have directly or indirectly met destination node d. This makes the decision of selecting a node to forward data difficult. The three clusters are linked by bridging ties from $i1$ to $i2$ and from $i3$ to $i4$. A path exists between the three clusters using intermediate nodes $i1$, $i2$, $i3$ and $i4$ which form bridges between the three clusters. Weak acquaintance ties of $i1$-$i2$ and $i3$-$i4$, illustrated by the dashed lines, become a crucial bridge between the three tightly connected groups, and these groups would not be connected if not for the existence of these weak ties. We propose to forward data based on the identification of these 'bridges' and the identification of nodes that reside within the same cluster as the destination node. Our main contribution of this paper is a new forwarding metric based on ego network analysis to locally determine a node's centrality within the network and the node's social similarity to the destination node. To the best of our knowledge, this is the first work to exploit social analysis techniques in this manner.

The remainder of this paper is organized as follows: Section 2 reviews related work in the area of message delivery in disconnected networks. Section 3 introduces the concept of node centrality in a network and shows how it can be used to forward information. Section 4 discusses social similarity and how it can be used to evaluate the strength of a relationship between two nodes. Section 5 presents SimBet Routing. Section 6 compares the performance of SimBet Routing to Epidemic Routing [38] and the PRoPHET Routing protocol [24] using real trace data from the MIT Reality Mining project [1, 7]. Finally, Section 7 concludes the paper.

2. RELATED WORK

A number of projects attempt to enable message delivery by using a virtual backbone with nodes carrying the data through disconnected parts of the network [11, 34]. The Data MULE project uses mobile nodes to collect data from sensors which is then delivered to a base station [34]. The Data MULEs are assumed to have sufficient buffer space to hold all data until they pass a base station. The approach is similar to the technique used in [2, 11, 12]. These projects study opportunistic forwarding of information from mobile nodes to a fixed destination. However, they do not consider opportunistic forwarding between the mobile nodes.

'Active' schemes go further in using nodes to deliver data by assuming control or influence over node movements. Li et al. [22] explore message delivery in disconnected MANETs where nodes can be instructed to move in order to transmit messages in the most efficient manner. The message ferrying project [40] proposes proactively changing the motion of nodes to help deliver data. They investigate both 'node initiated' mobility, where the nodes move to meet a known message ferry trajectory, or 'ferry initiated' mobility,

where the nodes signal communication requests via a long range radio, and the message ferry moves to meet them. Both assume control over node movements and in the case of message ferries, knowledge of the paths to be taken by these message ferry nodes. Other work utilises a time-dependent network graph in order to efficiently route messages. Jain et al. [16] assume knowledge of connectivity patterns where exact timing information of contacts is known, and then modifies Dijkstra's algorithm to compute the cost edges and routes accordingly. Merugu et al. [26] likewise make the assumption of detailed knowledge of node future movements. They route messages over a time-dependent graph with knowledge of when each node will next be encountered. Handorean et al. [14] take a similar approach with knowledge of connectivity. However, they do relax this assumption where only partial information is known. This information is time-dependent and routes are computed over the time-varying paths available. However, if nodes do not move in a predictable manner, or are delayed, then the path is broken. Additionally, if a path to the destination is not available using the time-dependent graph, the message is flooded.

Epidemic Routing [38] provides message delivery in disconnected environments where no assumptions are made in regards to control over node movements or knowledge of the network's future topology. Each host maintains a buffer containing messages. Upon meeting, the two nodes exchange summary vectors to determine which messages held by the other have not been seen before. They then initiate a transfer of new messages. In this way, messages are propagated throughout the network. This method guarantees delivery if a route is available but is expensive in terms of resources since the network is essentially flooded. Attempts to reduce the number of copies of the message are explored in [31] and [36]. Ni et al. [31] take a simple approach to reduce the overhead of flooding by only forwarding a copy with some probability $p < 1$, which is essentially randomized flooding. The Spray-and-Wait solution presented by Spyropoulos et al. [36] assigns a replication number to a message and distributes message copies to a number carrying nodes and then waits until a carrying node meets the destination.

A number of solutions employ some form of 'probability to deliver' metric in order to further reduce the overhead associated with Epidemic Routing by preferentially routing to nodes deemed most likely to deliver. These metrics are based on either contact history, location information or utility metrics.

Burgess et al. [4] transmit messages to encountered nodes in the order of probability for delivery, which is based on contact information. However, if the connection lasts long enough, all messages are transmitted, thus turning into standard Epidemic Routing. Acknowledgments are sent to all nodes upon delivery, and the delivered messages are then deleted from the buffers. PRoPHET Routing [24] is also probability-based, using past encounters to predict the probability of meeting a node again, nodes that are encountered frequently have an increased probability whereas older contacts are degraded over time. Additionally, the transitive nature of encounters is exploited where nodes exchange encounter probabilities and the probability of indirectly encountering the destination node is evaluated. Similarly [18] and [37] define probability based on node encounters in order to calculate the cost of the route. [6] and [13] use the so-called 'time elapsed since last encounter' or the 'last encounter age' to route messages to destinations. In order to route a message to a destination, the message is forwarded to the neighbour who encountered the destination more recently than the source and other neighbours.

Lebrun et al. [20] propose a location-based delay-tolerant routing scheme that uses the trajectories of mobile nodes to predict their future distance to the destination and passes messages to nodes that

are moving in the direction of the destination. Leguay et al. [21] present a virtual coordinate system where the node coordinates are composed of a set of probabilities, each representing the chance that a node will be found in a specific location. This information is then used to compute the best available route.

Musolesi et al. [28] introduce a generic method that uses Kalman filters to combine and evaluate the multiple dimensions of a node's context in order to make routing decisions. The context is created from measurements that nodes perform periodically, which can be related to connectivity. The approach only uses a single copy of a message, which is passed from one node to a node with a higher 'delivery metric'. The authors propose passing messages for unknown destinations using a 'default route' which is the 'most mobile' node available. Spyropoulos et al. [35] use a combination of random walk and utility-based forwarding. Random walk is used until a node with a sufficiently high utility metric is found after which the utility metric is used to route to the destination node.

Our work is distinct in that the SimBet Routing metric is comprised of both a node's centrality and its social similarity. Consequently, if the destination node is unknown to the sending node or its contacts, the message is routed to a structurally more central node where the potential of finding a suitable carrier is dramatically increased. We make no assumptions of control of node movements as in [22, 40] or knowledge of node future movements as in [14, 16, 26]. Unlike multi-copy strategies, we assume the existence of a single copy of each message in the network which reduces resource consumption compared to [31, 36, 38]. We will show that SimBet Routing improves upon encounter-based strategies where direct or indirect encounters may not be available [4, 6, 13, 18, 24, 37].

3. CENTRALITY

We estimate a node's centrality in the network in order to identify bridges. Centrality in graph theory and network analysis is a quantification of the relative importance of a vertex within the graph (e.g., how important a person is within a social network). The centrality of a node in a network is a measure of the structural importance of the node. A central node, typically, has a stronger capability of connecting other network members. There are several ways to measure centrality. The three most widely used centrality measures are Freeman's degree, closeness, and betweenness measures [9, 10].

'Degree' centrality is measured as the number of direct ties that involve a given node [10]. A node with high degree centrality maintains contacts with numerous other network nodes. Such nodes can be seen as popular nodes with large numbers of links to others. As such, a central node occupies a structural position (network location) that may act as a conduit for information exchange. In contrast, peripheral nodes maintain few or no relations and thus are located at the margins of the network. Degree centrality for a given node p_i is calculated as:

$$C_D(p_i) = \sum_{k=1}^{N} a(p_i, p_k) \tag{1}$$

where $a(p_i, p_k) = 1$ if a direct link exists between p_i and p_k and $i \neq k$.

'Closeness' centrality measures the reciprocal of the mean geodesic distance $d(p_i, p_k)$, which is the shortest path between a node p_i and all other reachable nodes [10]. Closeness centrality can be regarded as a measure of how long it will take information to spread from a given node to other nodes in the network [29]. Closeness centrality for a given node is calculated as:

$$C_C(p_i) = \frac{N - 1}{\sum_{k=1}^{N} d(p_i, p_k)} \tag{2}$$

where N is the number of nodes in the network and $i \neq k$.

'Betweenness' centrality measures the extent to which a node lies on the paths linking other nodes [9, 10]. Betweenness centrality can be regarded as a measure of the extent to which a node has control over information flowing between others [29]. A node with a high betweenness centrality has a capacity to facilitate interactions between the nodes that it links. In our case it can be regarded as how well a node can facilitate communication to other nodes in the network. Betweenness centrality is calculated as:

$$C_B(p_i) = \sum_{j=1}^{N} \sum_{k=1}^{j-1} \frac{g_{jk}(p_i)}{g_{jk}} \tag{3}$$

where g_{jk} is the total number of geodesic paths linking p_j and p_k, and $g_{jk}(p_i)$ is the number of those geodesic paths that include p_i. Freeman's centrality metrics are based on analysis of a complete and bounded network which is sometimes referred to as a sociocentric network. These metrics become difficult to evaluate in networks with a large node population because they require complete knowledge of the network topology. For this reason the concept of 'ego networks' has been introduced. Ego networks can be defined as a network consisting of a single actor (ego) together with the actors they are connected to (alters) and all the links among those alters. Consequently, ego network analysis can be performed locally by individual nodes without complete knowledge of the entire network. Marsden introduces centrality measures calculated using ego networks and compares these to Freeman's centrality measures of a sociocentric network [25]. Degree centrality can easily be measured for an ego network where it is a simple count of the number of contacts. Closeness centrality is uninformative in an ego network, since by definition an ego network only considers nodes to which the ego node is directly related and then by definition the distance from the ego node to all other nodes considered in the ego network is 1. On the other hand, betweenness centrality in ego networks has shown to be quite a good measure when compared to that of the sociocentric measure. Marsden calculates the egocentric and the sociocentric betweenness centrality measure for the network shown in figure 2.
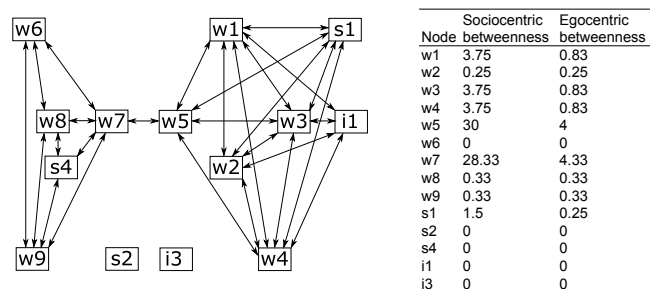

Node	Sociocentric betweenness	Egocentric betweenness
w1	3.75	0.83
w2	0.25	0.25
w3	3.75	0.83
w4	3.75	0.83
w5	30	4
w6	0	0
w7	28.33	4.33
w8	0.33	0.33
w9	0.33	0.33
s1	1.5	0.25
s2	0	0
s4	0	0
i1	0	0
i3	0	0

Figure 2: Bank Wiring Room network sociocentric and egocentric betweenness [25]

The betweenness centrality $C_B(p_i)$ based on the egocentric measures does not correspond perfectly to the sociocentric measures. However, it can be seen that the ranking of nodes based on the two measures of betweenness are identical in this network. This means that two nodes may compare their own locally calculated between-

ness value, and the node with the higher betweenness value may be determined. In effect, the betweenness value captures 'how much a node connects nodes that are themselves not directly connected'. For example in the network shown in figure 2, $w9$ has no connection with $w4$. The node with the highest betweenness value connected to $w9$ is $w7$, so if a message is forwarded to $w7$, the message can then be forwarded to $w5$ which has a direct connection with $w4$. In this way, betweenness centrality may be used to forward messages in a network. Marsden compared sociocentric and egocentric betweenness for 15 other sample networks and found that the two values correlate well in all scenarios [25].

4. SIMILARITY

Sociologists have long known that social networks (e.g., networks of personal acquaintances) display a high degree of transitivity, showing that there is a heightened probability of two people being acquainted if they have one or more other acquaintances in common. In physics literature this phenomenon is called 'clustering'. Watts and Strogatz showed that real-world networks exhibit strong clustering or network transitivity [39]. A network is said to show 'clustering' if the probability of two nodes being connected by a link is higher when the nodes in question have a common neighbour.

The degree of contact between nodes has an important effect in terms of information dissemination. When the neighbours of nodes are unlikely to be in contact with each other, diffusion can be expected to take longer than when the degree of separation is lower. Consequently, nodes with a lower degree of separation from a given node are good candidates for information dissemination to that node. The degree of separation can be measured by the ratio of common neighbours between individuals in social networks.

Newman analysed the time evolution of scientific collaborations and observed that the use of examining neighbours, in this case co-authors of authors, could help predict future collaborations [30]. From this analysis Newman determined that the probability of two individuals collaborating increases as the number m of their previous mutual co-authors goes up. A pair of scientists who have five mutual previous collaborators, for instance, are about twice as likely to collaborate as a pair with only two, and about 200 times as likely as a pair with none. Additionally Newman determined that the probability of collaboration increases with the number of times one has collaborated before, meaning that past collaborations are a good indicator of future ones.

Liben-Nowell and Kleinberg explored this theory by using this common neighbour metric in order to predict future collaborations on an author database [23]. The probability of a future collaboration $P(x, y)$ between authors x and y was calculated by:

$$P(x, y) = |N(x) \cap N(y)| \qquad (4)$$

Where $N(x)$ and $N(y)$ are the set of neighbours of author x and y respectively. This probability captures the 'similarity' between nodes x and y, relative to the network topology. The authors analysed a number of different metrics to capture the similarity between nodes. The results were promising where links were predicted, using the common neighbours metric, by a factor of up to 47 improvement compared to that of random prediction.

5. ROUTING BASED ON BETWEENNESS CENTRALITY AND SIMILARITY

In this section we present SimBet Routing, a forwarding algorithm based on betweenness centrality and similarity as described in sec-

tions 3 and 4 respectively. The algorithm makes no assumptions of global knowledge and forwarding decisions are based solely on local calculations.

5.1 Betweenness Calculation

Betweenness centrality is calculated using an ego network representation of the nodes with which the ego node has come into contact. Mathematically, node contacts can be represented by an adjacency matrix A, which is an $n \times n$ symmetric matrix, where n is the number of contacts a given node has encountered. The adjacency matrix has elements:

$$A_{ij} = \begin{cases} 1 & \text{if there is a contact between } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$$

We consider contacts to be bidirectional, so if a contact exists between i and j then there is also a contact between j and i. The betweenness centrality is calculated by computing the number of nodes that are indirectly connected through the ego node. The betweenness of the ego node is the sum of the reciprocals of the entries of $A^2 [1 - A]_{i,j}$ [8]. The following is a a matrix representation of the two-hop neighbourhood contacts of node $w8$ from figure 2.

$$w8 = \begin{matrix} & w8 & w6 & w7 & w9 & s4 \\ w8 & \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ w6 & 1 & 0 & 1 & 1 & 0 \\ w7 & 1 & 1 & 0 & 1 & 1 \\ w9 & 1 & 1 & 1 & 0 & 1 \\ s4 & 1 & 0 & 1 & 1 & 0 \end{bmatrix} \end{matrix}$$

$$w8^2[1 - w8] = \begin{matrix} & w8 & w6 & w7 & w9 & s4 \\ w8 & \begin{bmatrix} \star & \star & \star & \star & \star \\ w6 & \star & \star & \star & \star & 3 \\ w7 & \star & \star & \star & \star & \star \\ w9 & \star & \star & \star & \star & \star \\ s4 & \star & \star & \star & \star & \star \end{bmatrix} \end{matrix}$$

Since the matrix is symmetrical, only the non-zero entries above the diagonal need to be considered. In this case the only remaining entry of $w8^2 [1 - w8]$ is 3 and the reciprocal of the value is 0.33 which gives us the egocentric betweenness value for the node. When a new node is encountered, the new node sends a list of nodes it has encountered. A new entry is made in the $n \times n$ matrix. As an ego network only considers the contacts between nodes that it has directly encountered only the entries for the contacts that the newly encountered node shares in common with the ego node are inserted into the matrix.

5.2 Similarity Calculation

Node similarity is calculated using the same $n \times n$ matrix discussed in section 5.1. The number of common neighbours between the current node i and destination node j is a simple count of the non-zero equivalent row entries in the matrix. Consider the example matrix representing the contacts of node $w8$ in section 5.1. Node $w8$ has a similarity with nodes $w6$, $w7$, $w9$ and $s4$ of 2, 3, 3 and 2 respectively. This example only allows for the calculation of similarity for nodes that have been met directly, but when nodes exchange a list of nodes it has encountered as described in section 5.1 we may obtain useful information in regards to nodes that we have not yet encountered. As discussed in section 4 the number of common neighbours may be useful for ranking known contacts but also for predicting future contacts. It may also represent the possibility of routing to

an indirect node through a contact. Hence we maintain a list of indirect encounters and maintain a separate $n \times m$ matrix where n is the number of nodes that have been met directly and m is the number of nodes that have not directly been encountered, but may be indirectly accessible through a direct contact.

$$
w8 \;=\; \begin{array}{c} \\ w8 \\ w6 \\ w7 \\ w9 \\ s4 \end{array}
\begin{array}{c} \begin{array}{cccccc} w8 & w6 & w7 & w9 & s4 & w5 \end{array} \\
\left[\begin{array}{ccccc} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{array} \right] \left[\begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{array} \right] \end{array}
$$

The example above shows the inclusion of an indirect contact node $w5$ in the similarity calculation. The fact that node $w7$ has a contact with node $w5$ was learnt during an information exchange between node $w8$ and node $w7$. Since node $w8$ has no direct contact with node $w5$, it is added to the indirect contact matrix. Node $w8$ and node $w5$ share node $w7$ as a common neighbour and therefore have a similarity 1.

5.3 SimBet Utility Calculation

The SimBet utility is a value between 0 and 1 and is based on two components: similarity utility and betweenness utility. Selecting which node represents the best carrier for the message becomes a multiple attribute decision problem, where we wish to select the node that provides the maximum utility for carrying the message. This is achieved using a pairwise comparison matrix on the normalised relative weights of the attributes. The similarity utility $SimUtil_n$ and the betweenness utility $BetUtil_n$ of node n for delivering a message to destination node d compared to node m is given by:

$$ SimUtil_n(d) = \frac{Sim_n(d)}{Sim_n(d) + Sim_m(d)} \tag{5} $$

$$ BetUtil_n = \frac{Bet_n}{Bet_n + Bet_m} \tag{6} $$

The $SimBetUtil_n(d)$ is given by combining the normalised relative weights of the attributes given by:

$$ SimBetUtil_n(d) = \alpha SimUtil_n(d) + \beta BetUtil_n \tag{7} $$

where α and β are tunable parameters and $\alpha + \beta = 1$. Consequently these parameters allow for the adjustment of the relative importance of the two utility values.

5.4 SimBet Routing

This section describes SimBet Routing outlined in algorithm 1. The algorithm represents the communication between nodes n and m. Upon reception of a Hello message node n verifies that node m is a new neighbour. If this is the case, any messages destined for node m are delivered and an encounter request is sent. Node m replies with a list of nodes it has encountered. This list of contacts is then used to update the betweenness value on node n and the similarity value as described in sections 5.1 and 5.2 respectively. The two nodes then exchange a summary vector containing a list of destination nodes they are currently carrying messages for along with their own locally determined betweenness value and the similarity value for each destination. For each destination in the summary vector, node n calculates the SimBet utility of node n and node m as described in section 5.3. If node n has a higher SimBet utility

---

Algorithm 1 SimBet Routing Algorithm, pseudo-code of node n

1: **upon** reception of Hello message h from node m **do**
2:   **if** newNeighbour(m) == true
3:     **if** $msgQueue$.hasMsgsForDest(m) == true
4:       deliverMsgs(m)
5:     requestEncounters(m)
6:
7: **upon** reception of encounter vector ev from node m **do**
8:   addNodeEncounters(m, ev)
9:   updateBetweenness()
10:   updateSimilarity()
11:   exchangeSummaryVector(m)
12:
13: **upon** reception of summary vector sv from node m **do**
14:   Vector $requestMsgs$
15:   **for all** $destinations \in sv$ **do**
16:     **if** m.simBet(d) < simBet(d)
17:       $requestMsgs$.add(d)
18:   sendMsgRequest(m, $requestMsgs$)
19:
20: **upon** reception of message request vector mrv from node m **do**
21:   Vector $transferMsgs$
22:   **for all** $messages \in mrv$ **do**
23:     $transferMsgs$.add($msgQueue$.getMsgs(d))
24:   sendTransferMsgs(m, $transferMsgs$)
25:
26: **upon** reception of transfer message tm from node m **do**
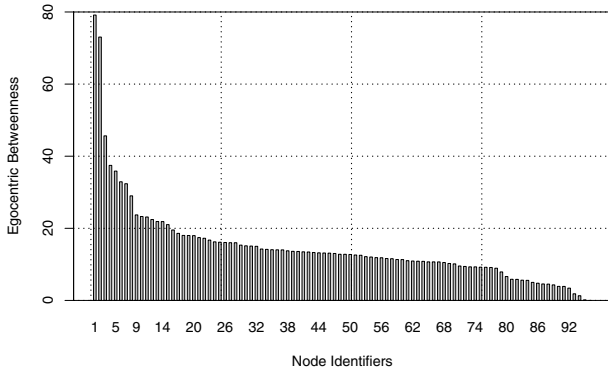27: $msgQueue$.add(tm)

---

for a given destination, the destination is added to a vector of destinations for which messages are requested. When all destinations in the summary vector has been compared, node n sends the message request list to node m. Node m then removes all messages destined for the destination node from its queue and forwards them to node n. Upon receiving a transfer message from node m the message is added to the message queue of node n.
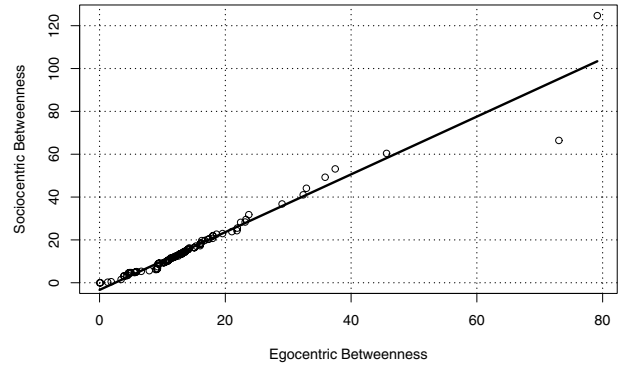
6. SIMULATION RESULTS

In this section we describe the simulations used to evaluate SimBet Routing and compare its cost and performance to Epidemic Routing and PRoPHET Routing. Our first experiment examines inter-node communication between the entire node population in order to evaluate the overall performance. In our second experiment we highlight the conditions where SimBet Routing succeeds in finding a route while PRoPHET fails, by limiting inter-node communication to the nodes least connected in the network.

6.1 Simulation Setup

In order to evaluate the premise of routing based on centrality and similarity we utilised a trace of node contacts from the MIT Reality Mining project [1, 7]. The study consisted of 100 users carrying Nokia 6600 smart phones over the course of nine months. They collected information using call logs, Bluetooth devices in proximity, cell tower IDs, application usage and phone status. For our purposes we use Bluetooth sightings in order to identify direct contacts between nodes where data transfer could have taken place. The trace file of these sightings was used to generate an event based simulation. Each time a contact was observed, nodes exchange encounters and update their locally calculated ego betweenness and social similarity values.

(a) Betweenness Distribution of nodes from MIT trace

(b) Scatterplot of sociocentric vs. egocentric betweenness for MIT trace

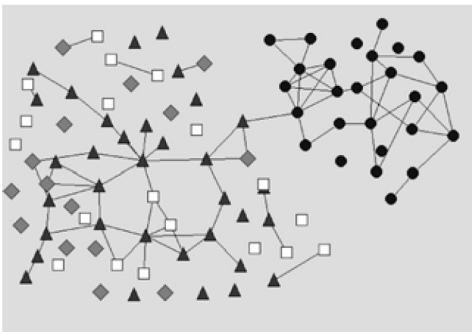**Figure 3: Betweenness Distribution**

Figure 4: Friendship network Eagle and Pentland [7]

Figure 3 a) shows the distribution of ego betweenness values calculated for all nodes at the end of the simulation. As can be seen there are two primary nodes that have a high ego betweenness. These nodes may represent highly social people who serve to link some of the less sociable individuals. Interestingly, this corresponds well to the social structure shown in figure 4 which the MIT Reality Mining team derived from interviews with the participants as to who they spent time with, both in the workplace and out of the workplace, and who they would consider to be in their circle of friends [1, 7]. They observed two distinct cliques with two nodes linking them.

In order to validate how accurately the locally calculated ego betweenness value correlates to a global view of the network we utilised the UCINET software for Social Network Analysis in order to calculate the sociocentric betweenness value using global information of the topology [3]. Figure 3 b) shows a scatterplot of the egocentric betweenness vs. the sociocentric betweenness which exhibit a very close correlation. The two outlier points represent the two most central nodes, and although the egocentric value does not directly map to the sociocentric value in these cases, the relative ranking of the two nodes is identical. We used Pearson's correlation of egocentric and sociocentric betweenness in order to evaluate the quality of the correlation. The Pearson's correlation value is 0.971 and thus egocentric betweenness reflects the comparative centrality of nodes for this network very well.

The simulation is event-based where a Bluetooth sighting in the MIT trace is assumed to be a contact where nodes can exchange in-

formation. Using these contacts we explored routing based on three protocols, Epidemic Routing [38], PRoPHET Routing [24] and the SimBet Routing described in section 5. The default parameters for PRoPHET Routing were used as defined in [24]. The parameters for the SimBet utility in equation 7 are set to $\alpha = \beta = 0.5$ which assigns an equal importance to the similarity and betweenness utility. We simulated each sending node generating one message for each receiving node. We assume that a complete information transfer as outlined in algorithm 1 may occur when a contact between two nodes is established.
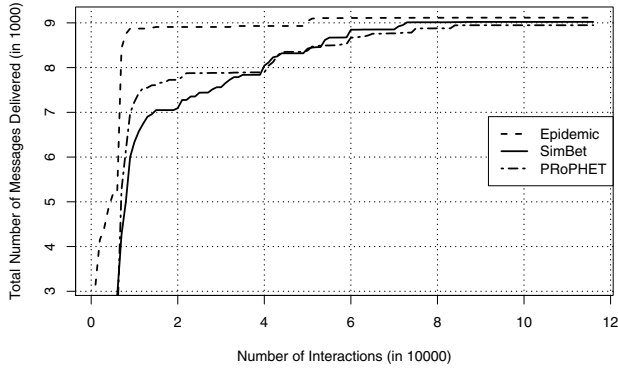
6.2 Performance Comparison

In the first test, all nodes are considered to be sending and receiving nodes. Each sending node generates a single message for all other nodes. We compare the three different routing protocols based on the following criteria.

Total Number of Messages Delivered: The ultimate goal of the SimBet Routing design is to achieve delivery performance as close to Epidemic Routing as possible. This is because Epidemic Routing always finds the best possible path to the destination and therefore represents the baseline for the best possible delivery performance.
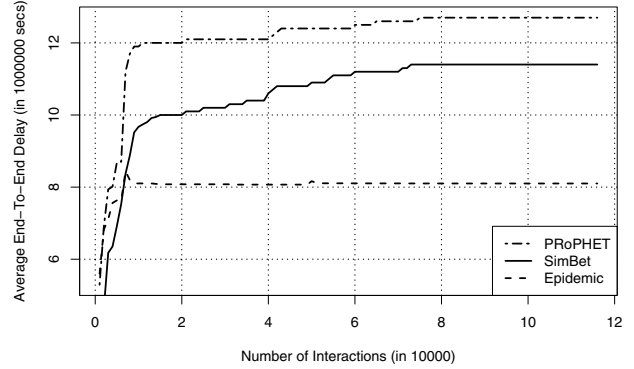
Average End-to-End Delay: End-to-End delay is an important concern in SimBet Routing design. Long end-to-end delays means the message must occupy valuable buffer space for longer, and consequently a low end-to-end delay is desirable. Again Epidemic Routing presents a good baseline for the minimum end-to-end delay possible.

Average Number of Hops per Message: It is desirable to minimise the number of hops a message must take in order to reach the destination. Wireless communication is costly in terms of battery power and as a result minimising the number of hops also minimises the battery power expended in forwarding the message.
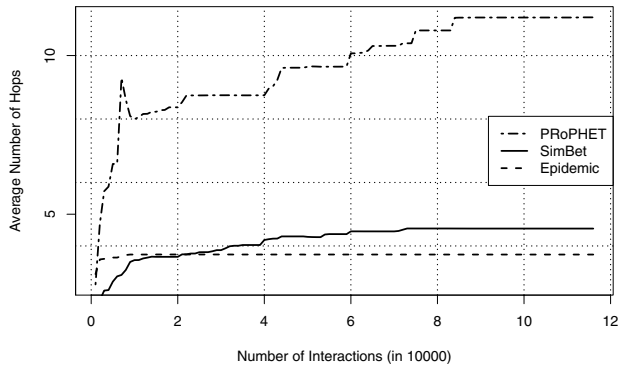
Total Number of Forwards: This value represents the overhead in the network in terms of how many times a message forward occurs in the network. PRoPHET and SimBet are expected to perform similarly in this respect, as both only assume the existence of one copy of the message on the network. Epidemic Routing, however, assumes the existence of multiple copies and continues forwarding a given message until each node is carrying a copy. This means Epidemic Routing is costly in terms of the number of transmissions required along with the amount of buffer space required on each node.
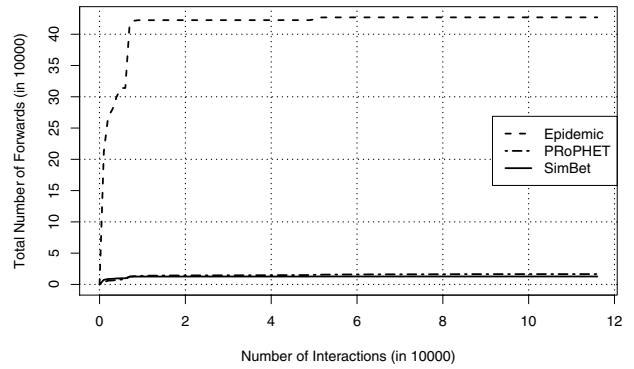
37

(a) Total Number of Messages Delivered

(b) End-to-End Delay (secs)

(c) Average Number of Hops per Message

(d) Total Number of Forwards

Figure 5: Simulation Results: Each node sending a message to each other node in the network

Figure 5 shows the results of each node sending a single message to every other node in the network. In figure 5 a) Epidemic Routing achieves the best delivery performance as expected, delivering 9116 messages. SimBet Routing performs quite close to Epidemic Routing delivering 9022 messages and better than PRoPHET Routing which delivers 8948 messages. Figure 5 b) shows that Epidemic Routing achieves the best performance in terms of average end-to-end delay, also as expected. PRoPHET takes 60% longer than Epidemic Routing with SimBet achieving results that fall in between that of Epidemic Routing and PRoPHET with a delay increase of 40% over Epidemic Routing. Figure 5 c) shows the average number of hops per message. The average number of hops per message of 4.5 achieved by SimBet Routing is very close to that of Epidemic Routing which is 3.7, whereas PRoPHET leads to longer routing paths resulting in an average hop value of 11. Figure 5 d) shows the total number of forwards. In this case, SimBet Routing performs best and the obvious disadvantage of Epidemic Routing becomes clear where the message continues to be forwarded throughout the network until each node has a copy, causing 33 times more message forwards than SimBet. SimBet results in a total of 12739 forwards which is lower when compared to PRoPHET which results in a total of 16519 forwards. The reduced number of forwards is due to the shorter paths found by SimBet.

As a result, we conclude that SimBet Routing comes close in terms of performance to that of Epidemic Routing, without the additional overhead of redundantly forwarding the message. Additionally, SimBet Routing performs better than PRoPHET in all metrics.

6.3 Delivery Performance Between Least Connected Nodes

As discussed in section 1 message delivery between nodes that have no direct or indirect contacts to the destination is problematic if based purely on past node encounters. The goal of the second experiment is to simulate message delivery between the least connected nodes in the network. To do this we vary the subset of sending and receiving nodes for each simulation. The first simulation in this experiment included only the least connected nodes as defined by the betweenness distribution shown in figure 3. For subsequent simulations the subset of nodes is increased to include the next two nodes in terms of increasing betweenness. A single message is generated between each node included in the subset.

Figure 6 shows the delivery performance of Epidemic Routing, PRoPHET and SimBet Routing for each simulation. The betweenness series shows the normalised maximum betweenness value of the subset of nodes. As can be seen from the figure, the performance of PRoPHET increases as the betweenness value of the subset of nodes increases. When considering message delivery between the 20 least connected nodes the percentage of message delivery for PRoPHET is low. In contrast SimBet Routing performs better with an improvement as high as 50 percentage points when considering the 2 least connected nodes and 25 percentage points when considering the 6 least connected nodes. This is achieved by routing from disconnected nodes to a more central node in order to find a good carrier for the message. As we have demonstrated

SimBet Routing clearly outperforms PRoPHET specifically in its ability to send messages among nodes with the lowest betweenness values which are the least central nodes in the network.
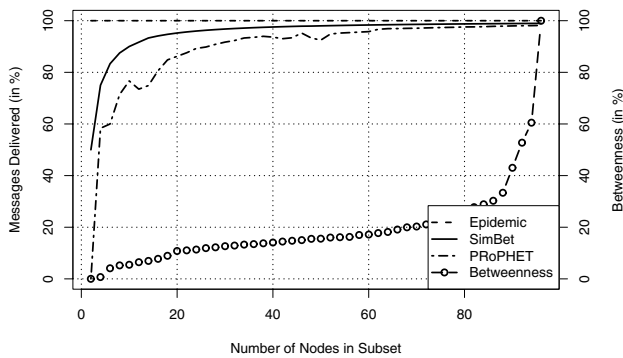

Figure 6: Percentage of message delivered between a subset of nodes increasing based on increased Betweenness

7. CONCLUSION

We have described SimBet Routing, a novel algorithm for routing in disconnected delay-tolerant MANETs based on social network analysis techniques. In the course of our research we have not come across a similar application of social network analysis. The SimBet routing metric is comprised of both a node's egocentric betweenness centrality and a node's social similarity. As such, if the destination node is unknown to the sending node or its contacts, the message is routed to a structurally more central node where the potential of finding a suitable carrier is dramatically increased. We have demonstrated through simulation using real trace data that SimBet Routing achieves delivery performance comparable to Epidemic Routing, without the additional overhead. We have also demonstrated that SimBet Routing succeeds in finding a route where PRoPHET fails due to the low connectivity of the sending and receiving nodes. We feel that these forwarding metrics may prove useful in other distributed systems where global topology information is unavailable, especially where the underlying networks exhibit small-world characteristics.

Acknowledgments

8. REFERENCES

[1] http://reality.media.mit.edu/.
[2] BEAUFOUR, A., LEOPOLD, M., AND BONNET, P. Smart-tag based data dissemination. In *proc. WSNA '02* (2002), ACM Press, pp. 68–77.
[3] BORGATTI, S. P., EVERETT, M. G., AND FREEMAN, L. C. Ucinet 6 for windows: Software for social network analysis, 2002.
[4] BURGESS, J., GALLAGHER, B., JENSEN, D., AND LEVINE, B. N. Maxprop: Routing for vehicle-based disruption-tolerant networking. In *proc. Infocom 2006* (April 2006), vol. 4, IEEE, pp. 1688–1698.
[5] CORSON, S., AND MACKER, J. Mobile ad hoc networking (manet): Routing protocol performance issues and evaluation considerations, rfc 2501, 1999.
[6] DUBOIS-FERRIERE, H., GROSSGLAUSER, M., AND VETTERLI, M. Age matters: efficient route discovery in mobile ad hoc networks using encounter ages. In *proc. MobiHoc '03* (2003), ACM Press, pp. 257–266.
[7] EAGLE, N., AND PENTLAND, A. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing V10*, 4 (May 2006), 255–268.
[8] EVERETT, M., AND BORGATTI, S. P. Ego network betweenness. *Social networks (Soc. networks) 27*, 1 (2005), 31–38.
[9] FREEMAN, L. C. A set of measures of centrality based on betweenness. *Sociometry* (1977), 35–41.
[10] FREEMAN, L. C. Centrality in social networks conceptual clarification. *Social networks (Soc. networks)* (1979), 215–239.
[11] FRENKIEL, R. H., BADRINATH, B. R., BORRES, J., AND YATES, R. D. The infostations challenge: balancing cost and ubiquity in delivering wireless data. *Personal Communications, IEEE 7*, 2 (2000), 66–71.
[12] GLANCE, N., SNOWDON, D., AND MEUNIER, J.-L. Pollen: using people as a communication medium. *Comput. Networks 35*, 4 (March 2001), 429–442.
[13] GROSSGLAUSER, M., AND VETTERLI, M. Locating nodes with ease: last encounter routing in ad hoc networks through mobility diffusion. In *proc. INFOCOM '03* (2003), vol. 3, IEEE, pp. 1954–1964 vol.3.
[14] HANDOREAN, R., GILL, C., AND ROMAN, G.-C. Accommodating transient connectivity in ad hoc and mobile settings. *Lecture Notes in Computer Science 3001* (March 2004), 305–322.
[15] HSU, W., AND HELMY, A. On nodal encounter patterns in wireless LAN traces. In *proc. WiNMee '06* (2006), IEEE.
[16] JAIN, S., FALL, K., AND PATRA, R. Routing in a delay tolerant network. *SIGCOMM Comput. Commun. Rev. 34*, 4 (October 2004), 145–158.
[17] JOHNSON, D., AND MALTZ, D. Dynamic source routing in ad-hoc wireless networks. *Mobile Computing* (1996), 152–181.
[18] KHELIL, A., MARRON, P. J., AND ROTHERMEL, K. Contact-based mobility metrics for delay-tolerant ad hoc networking. In *proc. MASCOTS '05* (2005), IEEE, pp. 435–444.
[19] KO, Y.-B., AND VAIDYA, N. H. Location-aided routing (lar) in mobile ad hoc networks. *Wirel. Netw. 6*, 4 (July 2000), 307–321.
[20] LEBRUN, J., CHUAH, C.-N., GHOSAL, D., AND ZHANG, M. Knowledge-based opportunistic forwarding in vehicular wireless ad hoc networks. In *proc. VTC '05* (2005), vol. 4, pp. 2289–2293.
[21] LEGUAY, J., FRIEDMAN, T., AND CONAN, V. Evaluating mobility pattern space routing for DTNs. In *proc. IEEE Infocom 2006* (April 2006), vol. 5, IEEE, pp. 2540–2549.
[22] LI, Q., AND RUS, D. Sending messages to mobile users in disconnected ad-hoc wireless networks. In *proc. MobiCom '00* (2000), ACM Press, pp. 44–55.
[23] LIBEN-NOWELL, D., AND KLEINBERG, J. The link prediction problem for social networks. In *proc. CIKM '03* (2003), ACM Press, pp. 556–559.

[24] LINDGREN, A., DORIA, A., AND SCHELÉN, O. Probabilistic routing in intermittently connected networks. *Lecture Notes in Computer Science 3126* (2004), 239–254.

[25] MARSDEN, P. V. Egocentric and sociocentric measures of network centrality. *Social networks (Soc. networks) 24* (October 2002), 407–422.

[26] MERUGU, S., AMMAR, M., AND ZEGURA, E. Routing in space and time in networks with predictable mobility. *Technical Report GIT-CC-04-7, Georgia Institute of Technology*.

[27] MILGRAM, S. The small world problem. *Psychology Today 1* (May 1967), 60–67.

[28] MUSOLESI, M., HAILES, S., AND MASCOLO, C. Adaptive routing for intermittently connected mobile ad hoc networks. In *proc. WoWMoM '05* (2005), IEEE, pp. 183–189.

[29] NEWMAN, M. E. J. A measure of betweenness centrality based on random walks. *Technical Report cond-mat/0309045, arXiv.*.

[30] NEWMAN, M. E. J. Clustering and preferential attachment in growing networks. *Phys. Rev. E, 64(025102)* (2001).

[31] NI, S.-Y., TSENG, Y.-C., CHEN, Y.-S., AND SHEU, J.-P. The broadcast storm problem in a mobile ad hoc network. In *proc. MobiCom '99* (1999), ACM Press, pp. 151–162.

[32] PERKINS, C. E., AND BHAGWAT, P. Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. In *proc. SIGCOMM '94* (October 1994), vol. 24, ACM Press, pp. 234–244.

[33] PERKINS, C. E., AND ROYER, E. M. Ad-hoc on-demand distance vector routing. In *proc. WMCSA '99* (1999), IEEE, pp. 90–100.

[34] SHAH, R. C., ROY, S., JAIN, S., AND BRUNETTE, W. Data mules: modeling a three-tier architecture for sparse sensor networks. In *proc. SNPA '03* (2003), IEEE, pp. 30–41.

[35] SPYROPOULOS, T., PSOUNIS, K., AND RAGHAVENDRA, C. S. Single-copy routing in intermittently connected mobile networks. In *proc. SECON '04* (2004), IEEE, pp. 235–244.

[36] SPYROPOULOS, T., PSOUNIS, K., AND RAGHAVENDRA, C. S. Spray and wait: an efficient routing scheme for intermittently connected mobile networks. In *proc. WDTN '05* (2005), ACM Press, pp. 252–259.

[37] TAN, K., ZHANG, Q., AND ZHU, W. Shortest path routing in partially connected ad hoc networks. In *proc. GLOBECOM '03* (2003), vol. 2, IEEE, pp. 1038–1042 Vol.2.

[38] VAHDAT, A., AND BECKER, D. Epidemic routing for partially connected ad hoc networks. *Technical Report CS-200006, Duke University* (2000).

[39] WATTS, D. J., AND STROGATZ, S. H. Collective dynamics of 'small-world' networks. *Nature 393*, 6684 (June 1998), 440–442.

[40] ZHAO, W., AMMAR, M., AND ZEGURA, E. A message ferrying approach for data delivery in sparse mobile ad hoc networks. In *proc. MobiHoc '04* (2004), ACM Press, pp. 187–198.