Search in Combined Social and Wireless Communication Networks: Delay and Success Analysis

Yalin E. Sagduyu, Member, IEEE, Yi Shi, Senior Member, IEEE, and Kartavya Neema Student Member, IEEE

Abstract—This paper models and analyzes the problem of search (navigation) with local information in combined social and wireless communication networks. Social networks are modeled with short-range and long-range connections representing small-world and scale-free network characteristics. By distinguishing the delay and success probability on different link types, the end-to-end delay distribution and success probability are first derived as functions of the social separation from the destination. New routing algorithms are then developed to improve the delay and chain completion success, and the effects of delay deadline on success probability are evaluated. The analysis is extended to the multi-layer combined social and communication network model, where wireless communication becomes the underlay to route information with the aid of social connections. The analytical results on delay and success probability are validated by comparing them with search results on a real-world social and communication network. Our results show how social connections can help reduce the search delay and increase the success probability in chain completion that runs on interdependent social and wireless communication network structures.

Index Terms—Social networks; wireless communication networks; interdependent networks; search; navigation; routing; delay; deadline; success probability.

I. INTRODUCTION

There has been an increasing interest in studying the methods of using social network connections, such as social network-aware routing, to improve the communication network performance [1]–[5]. This paper analyzes the delay of search (or end-to-end message delivery) and chain completion success between a source-destination pair in a combined social and communication network. Each node searches the next hop in the process of finding the destination node by using only local (one-hop) information about the neighbors. This problem has been extensively studied for social networks with real-world experiments [6]–[9] and the algorithmic aspects have been analyzed [10]–[20], typically with greedy routing scheme.

We leverage the Octopus model [21] to characterize the social and communication interactions in large-scale networks. Originally proposed for social networks, the Octopus model randomly deploys nodes in some geometric area with two types of links, i.e., short-range connections (SRCs) and long-range connections (LRCs). SRCs exist between a node and each of its neighbors that are socially separated within some range. LRCs are other social connections following some distributions, e.g., power-law distributions to model scale-free networks [22], and together with SRCs they model the small-world phenomenon [23].

For a social network generated by the Octopus model, the delay properties of the search problem have been developed in [21], [24] based on two assumptions that (i) the delay incurred in one hop is one time unit, independent of the type of links (SRC or LRC), and (ii) all chains are completed (i.e., there is no link failure in end-to-end searches). These two assumptions do not necessarily hold in real networks. We extend the Octopus model to capture heterogeneous delays on unreliable SRCs and LRCs, and after introducing probabilistic link failures, we develop new routing algorithms to improve the delay and end-to-end success probability in social search. This approach also serves as the first step to extend the Octopus model to cover the combination of social and communication networks with heterogeneous delays and success probabilities on different link types.

Information can be transferred through either social or communication links. For example, in cellular networks, social links on top of the 3G/4G infrastructure may constitute an overlay network, and at the same time WiFi or Bluetooth links of smart phones may establish a wireless ad hoc network with peer-to-peer communication. Such multi-layer network structures have many applications including emergency broadcasts, trusted communications, and secure key exchanges. The joint analysis on such a combined network is needed to reveal how one network affects the other in the context of information delivery in interdependent networks [25], [26]. Wireless communications with limited transmission range can be modeled as a random geometric graph, which is a special case of the Octopus model. Applying this property, we integrate communication networks with social networks induced by the Octopus model and obtain analytical expressions for the delay distribution and success probability (subject to link failures and delivery deadlines) in the combined network. The analysis is validated with large-scale search simulations as well as with search results on a real-world network data set [27]. Our results show that the combined network structure can support information transfer applications with better performance in...
terms of delay and/or success probability by using social links in communication network.

Our contributions can be summarized as follows:

1) we validate analytical result on social search delay by using real-world social network data;
2) we analyze the impact of errors in social separation knowledge on search delay;
3) we analyze search delay under different delay properties on SRCs and LRCs, and improve delay performance by a new routing algorithm;
4) we analyze the impact of link success probability on the end-to-end chain completion probability;
5) we apply our analytical results to combined social and wireless communication networks and validate search characteristics by using a real-world data set.
6) we quantify how the combination of social and wireless communication networks improves search performance.

The reminder of this paper is organized as follows. Section II presents the Octopus model for social and communication networks and validate social search delay analysis by real-world data. Section III analyzes the potential increase of the small-world phenomenon in social networks. In Section IV, we analyze the delay distribution with different delay properties on SRCs and LRCs, and develop a new routing algorithm to reduce delay. In Section V, we model the probabilistic success on each link and analyze the end-to-end chain completion success. In Section VI, we consider the problem of search on combined networks, and analyze the delay and chain completion characteristics with heterogeneous and unreliable (social/communication) links. In Section VII, we use a real-world (combined social and communication network) data set to validate the search characteristics. Section VIII summarizes the contributions.

II. OCTOPUS MODEL

We consider a network generated by the Octopus model [21] to model the small-world phenomenon in social networks [23]. There are \( n \) nodes randomly deployed on a disk with unit radius (the topology layout can also be assumed arbitrary [21]). A node has a SRC to another node if the distance between them is less than a range \( r \), and has \( n_{LRC} \) LRCs, which are chosen among nodes outside the range \( r \). Here, \( n_{LRC} \) can follow any arbitrary distribution, e.g., power law distribution to model scale-free networks [22]. The distance between two nodes refers to social separation when one node searches for the other node. This model has been studied in [21] for the asymptotic case (as \( n \) grows to infinity such that there is an infinite number of SRCs per node) with the unit delay over reliable SRCs or LRCs. Each node has only local information of the distance of its own and of its neighbors to the destination.

A. Greedy Routing

A greedy routing algorithm is used in [21] for social search. Each node \( i \) on the path to the destination node \( b \) chooses its next-hop node \( j \in N_i \) with the minimum distance \( H_{j,b} \) (or with the minimum social separation \( h_{j,b} \)) from the destination \( b \), where \( N_i \) is the neighborhood set of node \( i \), \( H_{j,b} \) is the distance between nodes \( j \) and \( b \), and \( h_{j,b} = \frac{H_{j,b}}{r} + 1 \) is the social separation (measured in hops) between nodes \( j \) and \( b \). Note that the social separation is equal to the number of hops to reach each other using SRCs only. When there is a tie between SRC and LRC, i.e., the best LRC decreases the separation by one, node \( i \) chooses a SRC to break the tie. We apply this greedy routing and extend it in Sections IV-B and V-B to improve delay and success probability, respectively.

B. Delay Properties of Octopus Model under Greedy Routing

Denote \( \varphi(t) = E[t^{n_{LRC}}] \) as the probability generating function of the number of LRCs, \( n_{LRC} \), per node, where the expectation is taken with respect to the distribution \( Q \) of \( n_{LRC} \), and \( \beta_i = 1 - \frac{e^{-x}}{x} \) as the probability that a given LRC lies outside the disk of radius \( (i - 1)r \) centered at the destination node. Note that \( \varphi(t) = t^{n_{LRC}} \) if the number of LRCs is fixed and equal to \( n_{LRC} \). The delay on each hop is assumed to be one unit of time. Denote \( T_k \) as the average delay to travel from any node \( x \) to destination \( b \) with hop separation \( h_{x,b} = k, M_1 \) as the location of the next hop node, \( X_b \) as the location of the destination and \( \rho(\cdot, \cdot) \) as the distance between any two locations inside the network domain. In [21], it has been shown that \( P(i - 1)r \leq \rho(M_1, X_b) < ir = \varphi(\beta_i) - \varphi(\beta_{i+1}), \) \( 1 \leq i \leq k - 2 \), for a LRC and \( P((k - 2)r \leq \rho(M_1, X_b) < (k - 1)r) = \varphi(\beta_{k-1}) \) for a SRC.

We can analyze the average delay by considering the first hop event. With probability \( \varphi(\beta_{k-1}) \), the first hop is a SRC and the social separation is reduced by \( 1 \) from \( k \) to \( k - 1 \). Otherwise, the first hop is a LRC and the social separation is reduced from \( k \) to \( i \) with probability \( \varphi(\beta_i) - \varphi(\beta_{i+1}) \), where \( i = 1, 2, \ldots, k - 2 \) (based on the routing algorithm described in Section II-A). As derived in [21], the delay for social separation \( k \geq 2 \) is recursively expressed as

\[
T_k = 1 + T_{k-1} \varphi(\beta_{u(k)+1}) + \sum_{i=1}^{u(k)} T_i (\varphi(\beta_i) - \varphi(\beta_{i+1})), \tag{1}
\]

where the first term “1” is the delay for the next hop, \( \varphi(\beta_{u(k)+1}) \) is the probability the next hop is a SRC and \( T_{k-1} \) is the remaining delay if the next hop is a SRC. \( \varphi(\beta_i) - \varphi(\beta_{i+1}) \) is the probability the next hop is a LRC that decreases hop distance to \( i \) and \( T_i \) is the remaining delay. The initial condition is \( T_1 = 1 \) and

\[
u(k) = k - 2. \tag{2}\]

Similarly, we can find the probability distribution of delay by conditioning on the first hop, which is either a SRC or a LRC from a node \( x \) with social separation \( h_{x,b} = k \) to another node \( y \) with \( h_{y,b} = i \). For given social separation \( k \geq 2 \), the probability that the delay is \( m \) \( (1 \leq m \leq k) \) is recursively expressed as

\[
P_k(m) = P_{k-1}(m-1) \varphi(\beta_{u(k)+1}) + \sum_{i=1}^{u(k)} P_i(m-1) (\varphi(\beta_i) - \varphi(\beta_{i+1})), \tag{3}
\]

with initial conditions \( P_k(0) = 0, k \geq 1 \), and \( P_1(1) = 1 \).
C. Communication Network: Special Case of Octopus Model

An ad hoc communication network can be modeled as a random geometric graph, where nodes are uniformly and independently distributed on a region (e.g., disk), and two nodes i and j are connected if and only if the distance between them is less than a threshold r, i.e., \( H_{i,j} < r \). Here, r is equal to the transmission/reception range \( R_C \) in a wireless network. Random geometric graph is a special case of the Octopus model without LRCs (i.e., \( n_{LRC} = 0 \) and \( \varphi(\beta_i) = 1 \)) such that the delay properties are reduced to \( T_k = k \) and \( P_k(m) = 1 \), if \( m = k \), or \( P_k(m) = 0 \), otherwise. In Section VI, we will combine a wireless communication network with the social network induced by Octopus model. This combination requires different delay properties assigned to SRCs and LRCs. We will start with analyzing heterogeneous link delays in Section IV.

D. Validation with Real-World Network Data

First, we verify Octopus model and preliminaries for search properties with real-world social network data. The data used for this purpose is the Arxiv HEP-TH (high energy physics theory) citation network [28] that has been used in [13] for a general network search problem. We select all papers in 1995-2000 and build the connectivity graph, where there is an undirected edge from i to j, if a paper i cites paper j. This graph has 21,097 nodes and 224,999 edges. Connections in citation network are undirected and highly topic related. These types of connections are similar to mentions, replies or retweets in an interaction graph built from Twitter network.

We perform social search by finding paths between randomly selected source-destination pairs in the citation graph. Ideally, social distance should be defined as the minimum hop distance between two nodes. But such distances are not available to individual nodes since nodes have only one-hop local information. If there is an edge between two nodes i and j, the distance between them is set to \( h_{i,j}^1 = 1 \). Otherwise, \( h_{i,j}^2 = 2 \). We can also use the similarity (in terms of paper content) to define the distance assuming two similar papers should have small hop distance. We define a set \( S_i \) to include all words that appear in the title and abstract of paper i (some common words, e.g., “an”, “an”, may exist in any paper and are excluded from \( S_i \)). The Jaccard similarity between papers i and j is defined as \( s_{i,j} = \frac{|S_i \cap S_j|}{|S_i \cup S_j|} \). Then, we can define distance as \( h_{i,j}^2 = 1 - s_{i,j} \). We observed that greedy routing based on either \( h_{i,j}^1 \) or \( h_{i,j}^2 \) usually finds long paths, i.e., they alone do not reflect shortest path characteristics. Therefore, we define a distance \( h_{i,j} = h_{i,j}^2 + (1 - \delta) h_{i,j}^1 \) that combines local connectivity information and similarity, where \( \delta \) is a small positive number.

We fit the citation graph to the Octopus model via exhaustive search to minimize the difference of average node degree in the (real) citation graph and the (synthetic) Octopus model that we generate with different parameters \( (r, \alpha) \), where r is the range of the SRCs and \( \alpha \) is the power law exponent of the LRCs. We run the social search on both the fitted graph and the original citation graph. Both graphs have 21,097 nodes and the average node degree is 21.33 in citation graph and 22.26 in Octopus model with selected \( r = 0.0225 \) and \( \alpha = 2.01 \).

### Table I

<table>
<thead>
<tr>
<th>Error ( \epsilon )</th>
<th>( \epsilon = 0 )</th>
<th>( \epsilon = 0.05 )</th>
<th>( \epsilon = 0.1 )</th>
<th>( \epsilon = 0.15 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Delay</td>
<td>7.68</td>
<td>14.31</td>
<td>26.55</td>
<td>34.01</td>
</tr>
</tbody>
</table>

However, citation graph is subject to a larger average search delay (28.30 hops) compared to the Octopus model (with average search delay of 7.68 hops). The reason is that the local estimation of the distance does not accurately represent the actual distance (i.e., shortest path length) and includes estimation errors, although Octopus model assumes that each node accurately knows the distance of itself and its neighbors to the destination.

For more realistic modeling, we can add an error term \( e_{i,j} \) (chosen uniformly random from the range \([ -\epsilon, +\epsilon ]\)) to each element of distance matrix in Octopus model. Table I shows that as the error magnitude \( \epsilon \) increases, the social search delay under Octopus model increases as well. The error is between 0.1 and 0.15 to fit the social search delay close to citation graph. Next, we will analyze the effects of distance estimation error on the social search delay under the Octopus model.

III. EFFECTS OF ESTIMATION ERROR ON SEARCH DELAY

Greedy routing assumes that each node knows the distance (social separation) of itself and its neighbors to the destination. However, a node may have only incomplete local information and it has been noted in [8] that the small-world paths are significantly longer (40%) than the shortest paths because people often make the wrong small-world choices in social search. To quantify this phenomenon, we model the local uncertainty of individual users on the global social separation by introducing two error probabilities: 1) \( e_1(k, \hat{k}) \) is the probability that a node x with actual social separation \( h_{y,b} = k \) from the destination b estimates its separation as \( \hat{k} \); 2) \( e_2(i, \hat{i}) \) is the probability that a node estimates that neighbor node y with actual social separation \( h_{y,b} = i \) to the destination b has social separation \( \hat{i} \). We assume that Octopus model continues to hold with the same parameters (SRC range \( r \) and LRC distribution \( Q \)) in the estimated distance graph. Also we assume \( T_1 = 1 \) (i.e., no error when the destination is one hop away).

Conditioned on the first hop, the average delay is

\[
T_k = 1 + \sum_{k=1}^{k_{max}} \sum_{i=1}^{k_{max}} e_1(k, \hat{k}) \varphi_{\hat{k}, i} e_2(i, \hat{i}) T_i \]

\[
= 1 + \sum_{k=1}^{k_{max}} e_1(k, \hat{k}) \sum_{i=1}^{k_{max}} \varphi_{\hat{k}, i} \sum_{i=1}^{k_{max}} e_2(i, \hat{i}) T_i, \quad (4)
\]

for \( 2 \leq k \leq k_{max} \), where depending on whether the first hop is SRC or LRC, \( \varphi_{\hat{k}, i} \) is \( \varphi(\beta_{\hat{k}}) \) for \( i = \hat{k} - 1 \) and \( \varphi(\beta_{\hat{k}}) = \varphi(\beta_{\hat{i} + 1}) \) for \( i = 1, \ldots, \hat{k} - 2 \). We have \( k_{max} \) - 1 equations by (4) on variables \( T_k \). \( 2 \leq k \leq k_{max} \) and can compute the values for \( T_k \)'s. The only requirement is that \( \sum_{k=1}^{k_{max}} e_{1,k}(k, \hat{k}) = 1 \) and \( \sum_{k=1}^{k_{max}} e_{2,i}(i, \hat{i}) = 1 \), where \( k_{max} = \lfloor \frac{1}{\alpha} \rfloor + 1 \) is the maximum separation between any two nodes. If there is no error, we have \( e_{1,k}(k) = 1 \), if \( k = \hat{k} \), and 0, otherwise, \( e_{2,i}(i) = 1 \), if \( i = \hat{i} \), otherwise, \( e_{2,i}(i) = 0 \).
and 0, otherwise, and (4) is reduced to (1). If the distances estimated for other nodes are independent from the actual distances, \( e_2(i, k) \) is given by \( (2i-1)^2 \), if \( i = 1, \ldots, k_{\text{max}} - 1 \), and by \( 1 - (k_{\text{max}} - 1)^2 \), if \( i = k_{\text{max}} \). Then, the average delay is \( T_1 = 1 \) and \( T_k = 1 + \frac{1}{2} k, k \geq 2 \).

For numerical results in this paper, we set \( r = 0.05 \) and assume a deterministic number of LRCs, \( n_{\text{LRC}} = 3 \), per node (unless stated otherwise). We assume no error for \( e_1(k, k) \) and uniform error pattern over one or two hops for \( e_2(i, i) \). Figure 1 shows for \( k_{\text{max}} = 20 \) how the delay increases with errors in estimating social separation. At the saturation (as \( k \) grows), the one- and two-hop errors increase the social search delay by 40% and 247%, respectively. We can use the delay in (4) to model realistic social search with longer delays under potential errors. The average delay saturates as the social separation \( k \) grows, pointing at the “small-world” phenomenon (even under estimation errors) in the sense that the delay between source-destination pairs remains bounded as the number of nodes grows to infinity. Results show that this trend holds irrespective of estimation errors.

Another potential source for delay discrepancy is that different link types may have different delays such as LRCs incurring longer delays than SRCs. We need such heterogeneous link delays to construct combined social and communication network. Therefore we will focus next on this case and extend Octopus model to heterogeneous link delays.

IV. OCTOPUS MODEL WITH HETEROGENEOUS LINK DELAYS

A. Delay under Greedy Routing

To combine social and communication links, we need to assign heterogeneous delay properties to different links (SRCs and LRCs in the combined network). The original Octopus model assumed the same delay for SRCs and LRCs. We now distinguish \( D_S \) and \( D_L \) as the delay of a SRC and LRC, respectively. For example, \( D_S \) may correspond to one-hop search delay between close friends and \( D_L \) may correspond to one-hop search delay between acquaintances (e.g., Facebook social network allows users to organize their friends into categories of close friends and acquaintances as well as into restricted and custom lists).

We consider a source-destination pair with social separation \( k \) and obtain the delay distribution by conditioning on the first hop in social search. The first hop is a SRC with probability \( \varphi(\beta_{u(k)}+1) \) such that the social separation of \( u(k) + 1 = k - 1 \) hops is left towards the destination, and this SRC incurs delay \( D_S \). Similarly, the first hop is LRC with probability \( \varphi(\beta_l) - \varphi(\beta_{l+1}) \) such that the social separation of \( i \) hops is left towards the destination, and this LRC incurs delay \( D_L \). Conditioned on this first hop event, the probability that the delay is \( d \in D_k = \{sD_S + lD_L: s \geq 0, l \geq 0, s + l = k\} \) for given social separation \( k \geq 2 \) is recursively expressed as

\[
P_k(d) = P_{k-1}(d - D_S) \varphi(\beta_{u(k)}+1) + \sum_{i=1}^{u(k)} P_i(d - D_L) (\varphi(\beta_i) - \varphi(\beta_{i+1})),
\]

where \( \varphi(\beta_{u(k)}+1) \) is the probability that the next hop is a SRC and \( P_{k-1}(d - D_S) \) is the success probability for the remaining path, \( (\varphi(\beta_i) - \varphi(\beta_{i+1})) \) is the probability that the next hop is a LRC that decreases hop distance to \( i \) and \( P_i(d - D_L) \) is the success probability for the remaining path. The initial conditions are \( P_k(d) = 0 \) if \( d \notin D_k \) and \( P_1(D_S) = 1 \).

The average delay can be directly derived from \( D_k = \sum_{d \in D_k} P_k(d) d \), or can be recursively found by conditioning on the first hop as follows:

\[
D_k = (D_S + D_{k-1}) \varphi(\beta_{u(k)}+1) + \sum_{i=1}^{u(k)} (D_L + D_i) (\varphi(\beta_i) - \varphi(\beta_{i+1})),
\]

for \( k \geq 2 \), where \( (D_S + D_{k-1}) \) is the delay for the remaining path if the next hop is a SRC, \( (D_L + D_i) \) is the delay for the remaining path if the next hop is a LRC that decreases hop distance to \( i \). The initial condition is \( D_1 = D_S \).

The average delay of social search under heterogeneous delay properties is shown in Figure 2 for different values of social separation \( k \), (deterministic) number of LRCs \( n_{\text{LRC}} \), and LRC delay \( D_L \), where \( r = 0.05 \) and \( D_S = 1 \). The delay \( D_k \) is close to linear (namely, close to \( k \) achieved by SRCs only) for a small social separation \( k \) but saturates as \( k \) increases such that nodes can find each other with a bounded delay irrespective of how far they are from each other. This saturation point of social separation increases with \( D_L \) and decreases with \( n_{\text{LRC}} \).
Next, we compare the analytical delay results with simulations, where we generate finite-size random networks of $n = 5000, 10,000$ and $15,000$ nodes according to the Octopus model with parameters $r = 0.05$ and $n_{LRC} = 3$. Then, greedy routing is run by varying $D_L$ and fixing $D_S = 1$. The average number of SRCs per node would go to infinity in the asymptotic model of the analysis and simulations can only approximate the analysis by achieving the average delay of $12.29, 24.45$ and $36.65$ SRCs per node, respectively, for $n = 5000, 10,000$ and $15,000$ nodes. Figure 3 shows that as $n$ increases, the average delay of simulated greedy search approaches the analytical values computed from (6).

B. New Routing for Delay Improvement

The original greedy routing algorithm selects the next node $i$ to be closest (in terms of social separation) to the destination $b$ independent of the delay caused by that hop (SRC or LRC). By distinguishing delays on SRCs and LRCs, we modify the greedy routing algorithm such that any node $i$ selects neighbor $j$ with the maximum value of $D_{i,j}^{-1}$ as the next hop, where $D_{i,j}$ is the delay of link from node $i$ to node $j$ and is either $D_S$ or $D_L$ depending on whether the link $(i, j)$ is SRC or LRC, respectively. The derivation of delay characteristics is similar to the original greedy routing. For a link $(i, j)$, we have

$$\frac{h_{i,b} - h_{j,b}}{D_{i,j}} = \begin{cases} \frac{1}{D_S} & \text{for SRC } (i, j) \\ \frac{1}{D_L} & \text{for LRC } (i, j) \end{cases}.$$  

The LRC link $(i, j)$ is chosen compared to SRC link if $\frac{h_{i,b} - h_{j,b}}{D_{i,j}} \geq \frac{1}{D_S}$. This condition can be rewritten as $h_{j,b} \leq u(h_{i,b})$, where

$$u(k) = k - \lfloor D_L/D_S \rfloor - 1. \quad (7)$$

Then, $P_k(d)$ and $D_k$ follow from (5) and (6), respectively, with the only change that $u(k)$ is defined by (7) instead of (2).

Figure 4 shows the average delay for $r = 0.05$, $n_{LRC} = 3$ and $D_S = 1$. If $D_L = D_S$, the two routing algorithms are the same. The delay gain of modified routing increases with $D_L$. For large value of $D_L$, the original greedy routing may perform worse than using SRCs only (e.g., for $D_L = 5$ in Figure 4), whereas the new routing prevents such cases and improves the delay performance. For instance, the measured delay reduction is $14\%$ for $D_L = 5$ and $k = 12$.

V. CHAIN COMPLETION SUCCESS

A chain (a search path) of social search may not be completed because of the failure of links on that path. For instance, only 18 out of 96 message chains were completed in a real-world mail experiment (i.e., the success probability is 0.1875) [6] and only 384 out of 24,163 message chains were completed in an online e-mail experiment (i.e., the success probability is 0.0158) [7]. There are several reasons for the dead end of a message chain before reaching the destination, e.g., (i) the receiving side of the link may drop the message (e.g., the message is considered spam) or (ii) greedy routing may not find a link with positive progress towards the destination (because of the finite nature of the underlying connectivity graph). We aim to model failures in chain completion by assigning a forwarding probability to each link.

A. Success Probability under Greedy Routing

We distinguish the forwarding probabilities $P_{fs}$ and $P_{fl}$ for SRCs and LRCs, respectively. The first hop is a SRC with probability $\varphi(\beta_{u(k)}+1)$ such that the social separation of $k-1$ hops is left towards the destination, and this SRC is successful with probability $P_{fs}$. Similarly, the first hop is a LRC with probability $\varphi(\beta_{i})-\varphi(\beta_{i+1})$ such that the social separation of $i$ hops is left towards the destination, and this LRC is successful with probability $P_{fl}$. Conditioned on this first hop event, the end-to-end success probability is recursively expressed as

$$S_k = P_{fs}S_{k-1}\varphi(\beta_{u(k)}+1) + \sum_{i=1}^{u(k)} P_{fl}S_i(\varphi(\beta_{i}) - \varphi(\beta_{i+1})) \quad (8)$$

for $k \geq 2$, where the initial condition is $S_1 = P_{fs}$ and $u(k)$ is defined by either (2) or (7) based on the greedy routing used. Similarly, the delay distribution is recursively expressed as

$$P_k(d) = \frac{1}{S_k}P_{fs}P_{k-1}(d-D_S)\varphi(\beta_{u(k)}+1) + \sum_{i=1}^{u(k)} P_{fl}P_{k-1}(d-D_L)(\varphi(\beta_{i}) - \varphi(\beta_{i+1})) \quad (9)$$

for $k \geq 2, d \in D_k$, where the initial conditions are $P_k(d) = 0$, if $d \notin D_k$, and $P_k(D_S) = 1$. In (9), we only consider the successful end-to-end searches by normalizing the distribution.
with respect to the end-to-end success probability $S_k$. With this distribution, we can calculate the average delay $D_k$.

We compare the analytical results of success probability with simulations, where we generate random networks according to the Octopus model with $r = 0.05$ and $n_{LRC} = 3$. Figure 5 shows that the success probability $S_k$ is close to $(P_{fs})^k$ (achieved by using SRCs only) for small values of $k$ and saturates to a finite value as $k$ increases, i.e., nodes can find each other with a bounded success probability independent of their separation. This end-to-end success probability increases with increasing link success probabilities $P_{fl}$ and $P_{fl}$. As the number of nodes increases, the success probability in simulated greedy searches approaches the analytical values under heterogeneous SRC and LRC link forwarding probabilities.

### B. New Routing for Success Probability Improvement

We introduce a new routing algorithm to improve the successful chain completion. Define $P_{fl}$ as the probability of success on a link from node $i$ to node $j$. Note that $P_{fl}$ is either $P_{fs}$ or $P_{fl}$ depending on whether the link $(i, j)$ is SRC or LRC, respectively. Node $i$ selects neighbor $j \in N_i$ with the maximum value of $P_{fl}(P_{fs})^{h_{i,j}} b_{i,j}$ as the next hop subject to the condition $h_{i,j} < h_{i,b}$ that ensures positive progress towards the destination. The first term $P_{fl}$ in $P_{fl}(P_{fs})^{h_{i,j}} b_{i,j}$ corresponds to the probability of success in the next hop and the second term $(P_{fs})^{h_{i,j}} b_{i,j}$ corresponds to the probability of success on the remaining path to the destination by using SRCs. With local information only, node $i$ cannot know the number and type of hops in the remaining path and simply estimates this path by $h_{i,j}$ SRCs as done explicitly in the original greedy routing algorithm. Note that this new success-based routing algorithm is reduced to the original greedy routing for $P_{fl} \leq P_{fl}$. For a link $(i, j)$, we have

$$P_{fl}(P_{fs})^{h_{i,j}} b_{i,j} = \begin{cases} (P_{fs})^{h_{i,j}+1} & \text{for SRC } (i, j) \\ (P_{fl})(P_{fs})^{h_{i,j}} & \text{for LRC } (i, j) \end{cases}.$$

Then, LRC link $(i, j)$ is chosen compared to SRC link, if $(P_{fl})(P_{fs})^{h_{i,j}} b_{i,j} > (P_{fs})^{h_{i,j}} b_{i,j}$. This condition can be rewritten as $h_{i,j} \leq u(h_{i,j})$, where

$$u(k) = k - \lfloor \log(P_{fl})/ \log(P_{fs}) \rfloor - 1. \tag{10}$$

With this $u(k)$, the success probability follows from (8) and the delay distribution follows from (9).

### C. Chain Completion Success under Deadlines

Another reason of unreliability in social search is that messages may arrive at the destination later than some deadline $\tau$ and therefore they expire and cannot contribute to the end-to-end success probability. Then, the probability of successfully finding a destination with social separation $k \geq 1$ is given by

$$S_k = \sum_{d \in \mathcal{D}_k, d \leq \tau} P_{k}(d),$$

where $P_{k}(d)$ is the delay distribution under either the original or new algorithm. The success probability $S_k$ under different deadlines and the original greedy algorithm is shown in Figure 7 for $r = 0.05$, $n_{LRC} = 3$, $D_S = 1$, $D_L = 2$, $P_{fl} = 0.75$ and $P_{fl} = 0.9$. Note that $S_k$ decreases as $\tau$ decreases or $k$ increases. With deadlines, there is a significant drop in success probability because the deadline expiration may contribute to most of the failures in chain completion compared to link failures.
VI. COMBINED SOCIAL-COMMUNICATION NETWORK

In previous sections, we build the preliminary tools to analyze combined social and wireless communication networks. We consider a wireless network as the underlay communication network. In addition to the wireless network, there is a social network, where nodes can communicate via social links. For the analysis of search problem in combined social and communication networks, we extend the model of different social link delay and success properties, and map social links to a communication network. Rather than modeling each network separately, we model them as one combined network and distinguish the delay and success properties of communication and social links that coexist with different levels of geographic and social separation. We consider the information transfer through the combined network, where the information can be carried between source-destination pairs by either communication or social links.

An ad hoc communication network can be modeled as a random geometric graph with communication range $r_C$ (see Section II-C), where nodes are uniformly and independently distributed on a region (e.g., a disk with radius $r$), and two nodes are connected if and only if the distance between them is less than a threshold $r_C$. The threshold $r_C$ is the transmission/reception distance and typically depends on the transmission power, the signal-to-noise-ratio (SNR) requirement of wireless radios and the wireless channel characteristics, such as path loss, RMS delay spread, and interference.

We merge this random geometric graph with the social network induced by the Octopus model. There is a one-to-one mapping from nodes in social network to nodes in communication network. Based on this mapping, we integrate SRCs and LRCs in social network to the communication network and construct the combined network. Let $D_C$ denote the communication link delay, which corresponds to the transmission/reception delay, and $D_S$ and $D_L$ denote the SRC and LRC social link delays, which correspond to the processing delay by people when forwarding messages. We assume $D_C \leq D_S \leq D_L$.

In the social network, the distribution of the number of SRCs, $Q_S$, follows a Binomial distribution, which becomes a Poisson distribution, as the number of nodes goes to infinity, whereas the number of LRCs follows a general distribution $Q_L$. We map social links to the communication network. Then some of these links coincide with communication links, whereas some of them become LRCs in the combined graph. The number of SRCs in the social network that become LRCs in the combined graph follows the distribution

$$Q_{C,S}(n') = \sum_{n=n'}^{\infty} Q_S(n) B(n', n, 1 - r_C^2), n' \geq 0, \quad (11)$$

where $B(n', n, 1 - r_C^2) = (\binom{n'}{n})(1 - r_C^2)^{n-n'} (r_C^2)^n$ is the probability that $n'$ out of $n$ nodes are outside the communication range and the rest are in the communication range. Similarly, the number of LRCs in the social network that become LRCs in the combined graph follows the distribution

$$Q_{C,L}(n') = \sum_{n=n'}^{\infty} Q_L(n) B(n', n, 1 - r_C^2), n' \geq 0. \quad (12)$$

In the combined (social and communication) network, the total number of LRCs follows the distribution

$$Q_C(n') = \sum_{i=0}^{n'} Q_{C,S}(i) Q_{C,L}(n' - i), n' \geq 0. \quad (13)$$

To analyze the success probability, let $\varphi_C(t)$, $\varphi_{C,S}(t)$ and $\varphi_{C,L}(t)$ denote the probability generating function of the number of LRCs, $n_{LRC}$, per node, where the expectation is taken with respect to the distribution $Q_C$, $Q_{C,S}$ and $Q_{C,L}$, respectively. We condition on the first hop in the combined graph. Then $\tilde{\varphi}_C(n) = \varphi_C(\beta_{u(k)}+1)$ is the probability that a communication link is chosen, $\tilde{\varphi}_{C,S} = (\varphi_C(\beta_i) - \varphi_{C,S}(\beta_i+1))\varphi_{C,L}(\beta_i)$ is the probability that no social link can reduce social separation to $i - 1$ and that at least one a social SRC reduces social separation to $i$. Then some of these links coincide with communication links, and $\tilde{\varphi}_{C,L} = (\varphi_{C,L}(\beta_i) - \varphi_{C,L}(\beta_{i+1}))\varphi_{C,S}(\beta_{i+1})$ is the probability that no social link can reduce social separation to $i - 1$ and that at least one a social LRC reduces social separation to $i$.

By using one of the greedy routing algorithms (with the proper choice of $(u(k))$, the success probability for given separation $k$ on the combined network is

$$S_k = \frac{P_{fc} S_{k-1} \tilde{\varphi}_{C}(n') + \sum_{i=1}^{u(k)} (P_{fs} \tilde{\varphi}_{C,S} + P_{fl} \tilde{\varphi}_{C,L}) S_i}{S_k} \quad \text{for } k \geq 2, \text{ where the initial condition is } S_1 = P_{fc}. \quad (14)$$

The probability that the delay is $d \in D_k = \{c D_C + s D_S + l D_L, c \geq 0, s \geq 0, l \geq 0, c + s + l = k\}$ for given separation $k \geq 2$ is recursively expressed as

$$P_k(d) = \frac{1}{S_k} \left[ P_{fc} P_{k-1}(d - D_C) \tilde{\varphi}_{C}(n') + \sum_{i=1}^{u(k)} (P_{fs} P_i (d - D_S) \tilde{\varphi}_{C,S} + P_{fl} P_i (d - D_L) \tilde{\varphi}_{C,L}) \right], \quad (15)$$

where the initial conditions are $P_k(d) = 0$, if $d \notin D_k$, and $P_1(D_C) = 1$. Then the average delay $\bar{D}_k$ can be calculated from (15), or can be recursively expressed as

$$\bar{D}_k = (D_C + D_k - 1) \tilde{\varphi}_{C,S} + \sum_{i=1}^{u(k)} (D_L + D_i) \tilde{\varphi}_{C,L} + \sum_{i=1}^{u(k)} (D_S + D_i) \tilde{\varphi}_{C,L} \quad \text{for } k \geq 2, \text{ where the initial condition is } D_1 = D_S. \quad (16)$$

For performance evaluation, we assume that the number of LRCs for social links is Poisson distributed with mean 3, the number of SRCs for social links is Binomial distributed with mean 4, and $r_C = 0.05$. Figures 8 and 9 show the average search delay and success probability in the combined network under the original greedy routing. When the network uses social links as part of routing, the average delay can be reduced (as shown in Figure 8) and this gain increases as the social link delays $D_L$ and $D_S$ decrease. Similarly, the success probability increases by the use of social links (as shown in Figure 9) and this gain increases as the social link success probabilities $P_{fs}$ and $P_{fl}$ increase.

Figure 10 shows the ratio of the average number of social links used (in comparison with communication links) for
different delays under delay-based routing and for different success probabilities under success-based routing. As the delay increases or success probability decreases over communication links, the network starts using more social links, and this adaptation capability between social and communication links allows the effective use of combined social and communication network resources, and improves the combined network performance (including the end-to-end delay and the end-to-end success probability).

VII. REAL-WORLD DATA VALIDATION

We use Gowalla dataset [27] to validate the Octopus model and search results in the combined social and communication network. Gowalla was a location-based social network, where users were able to check in to visited locations using their mobile device. This dataset provides friendship links and user locations over time. First, we filter the data with respect to location and maintain 4,395 users, who have checked in within 500m radius of the Austin city center at least once. We compute the average node locations in this range and select the user closest to the center as the destination and the rest of users as the sources of potential searches. We assume 25m as the communication range between users. There are 21,948 communication links and 24,173 social links. In this setup, we assume that users can use both social (friendship) links (e.g., through cellular backbone) and communication links (e.g., through a multi-hop ad hoc network) to reach the destination.

We build the Octopus model by fitting the communication range and average degree of social links to the Gowalla data.

This way, we obtain that $r_C = 0.05$ (for network radius of 1), the number of LRCs for social links is Poisson distributed with mean 5, and the number of SRCs for social links is Binomial distributed with mean 6. By assuming $D_S = D_L = 6$, $P_{fS} = P_{fL} = 0.5$, $P_{LC} = 0.5^{D/6}$ and varying $D_C$ from 1 to 3, we run delay-based routing on two networks, induced by real-world data and fitted Octopus model graph. Figure 11 shows the delay in the combined network under real-world data compared with analytical results. Note that the analysis assumes infinite number of nodes while the dataset has a large but finite number of nodes, which causes the gap between analytical and dataset results. Real dataset results are close to the analytical delay computation and match the analytical trends, (i) search delay increases with communication link delay, (ii) search delay increases with separation and first it remains close to linear but then quickly saturates. Similar results will follow for success probability.

VIII. CONCLUSION

We analyzed the search delay and chain completion success in combined social and communication networks. First, we evaluated the distribution of social search delay under heterogeneous link delays and validated results with real-world data under distance estimation error models. Then, we analyzed the end-to-end search success probability under unreliable links and delivery deadlines. We extended the search problem to combined social and communication networks and analyzed the search delay and chain completion characteristics on the generalized Octopus model with social and communication networks.
links. We validated the analysis via search simulations on a real-world dataset. Our results showed how the efficient use of social links can reduce the search delay and increase the success probability on combined structures of social and communication networks.

ACKNOWLEDGMENTS
We would like to thank Prof. Mung Chiang for his valuable information on the Octopus model, comments and suggestions.

REFERENCES

Yalin E. Sagduyu (S’02–M’08) received the B.S. degree from Bogazici University, Turkey, in Electrical and Electronics Engineering and the M.S. and Ph.D. degrees in Electrical and Computer Engineering from the University of Maryland at College Park. Dr. Sagduyu is a Program Manager at Intelligent Automation Inc., Rockville MD. His research interests are in the areas of design and optimization of wireless networks, cognitive radio networks, network coding, information theory, security, game theory, social networks, and big data analytics. Dr. Sagduyu co-chaired several IEEE and ACM Workshops on cognitive radio networks. He also served as the “MAC and Cross-Layer Design” Track Co-Chair at IEEE PIMRC, 2014.

Yi Shi (S’02–M’08–SM’13) is a Senior Research Scientist in Intelligent Automation Inc. His current research focuses on algorithms and optimization for social networks, satellite communication system, cognitive radio networks, MIMO, cooperative, sensor and ad hoc networks, and network coding. He has co-organized four IEEE and ACM workshops and has been a TPC member of more than 50 major IEEE and ACM conferences. He is an Editor of IEEE Communications Surveys and Tutorials. He authored one book, five book chapters and more than 100 papers on network design and analysis. He was a recipient of IEEE INFOCOM 2008 Best Paper Award, IEEE INFOCOM 2011 Best Paper Award Runner-Up, and ACM WUWNet 2014 Best Student Paper Award.

Kartavya Neema (S’15) is a Ph.D candidate in the School of Aeronautics and Astronautics, Purdue University. His current research interest includes consensus control, target tracking, distributed real-time optimization, analysis of communication and social networks, and UAV design. He is a student member of IEEE and American Institute of Aeronautics and Astronautics (AIAA).