# Delay of Social Search on Small-world Random Geometric Graphs

Hazer Inaltekin [a,*] Mung Chiang [a] H. Vincent Poor [a]
[a] *Princeton University, Princeton, NJ, United States*

**Abstract**

This paper studies the delay of social search by considering messages traveling between source and target individuals in small-world random geometric graphs. In particular, by considering such graphs constructed on different network domains such as, rectangular, circular and spherical network domains, an exact characterization of the average social search delay is obtained as a function of source-target separation, distribution of the number of long-range connections and geometrical properties of network domains. Derived analytical formulas for the average social search delay are first verified by agent-based simulations, and then compared and contrasted with empirical observations in small-world experiments. It is shown that individuals tend to communicate with one another only through their short-range contacts, and the average social search delay rises linearly, when the separation between the source and target is small. On the other hand, as this separation increases, long-range connections are more commonly used, and the average social search delay rapidly saturates to a constant value and stays almost the same for all large values of the separation. These results are consistent with experimental observations made by Travers and Milgram in 1969, as well as by others. Other somewhat surprising conclusions of the paper are that hubs have limited effect in reducing the delay of social search and the degree of social inequality existing in a society adversely affects this delay.

*Key words:* Small-world networks, Social networks, Social search, Delay

⋆ This research was supported in part by the National Science Foundation under Grants ANI-03-58807, CNS-06-25637, CCF-0448012, and by ONR Grant N00014-07-1-0864.
* Correspondence address: F310 Engineering Quadrangle, Olden Street, Princeton, NJ 08544, USA. Tel.: +1 607 379 1582
*Email address:* `hinaltek@princeton.edu` (Hazer Inaltekin).

## 1. Introduction

Many social processes, such as the search for resources (Granovetter, 1973, 1983; Lee, 1969), mobilization and organization of communities for common goals (Granovetter, 1973, 1983), spread of infectious diseases (Bearman et al., 2004; Klovdahl, 1985), adoption of innovations (Coleman et al., 1957), emergence of social norms (Centola et al., 2005) and spread of rumors (Donovan, 2007), depend critically on the structure of social networks among people. The goal of much small-world research (e.g., Milgram, 1967; Erickson and Kringas, 1975; Guiot, 1976; Pool and Kochen, 1978; Kochen, 1985; Watts and Strogatz, 1998; Kleinberg, 2000a,b; Newman, 2001; Dodds et al., 2003; Watts, 2004; Schenettler, 2009a,b) is to unravel intrinsic structural properties of social networks, such as geodesics, to shed light on this diverse set of social processes.

Milgram and his colleagues conducted the well-known letter-referral experiments, which are also known as small-world experiments, to estimate the average number of steps needed to connect two individuals in the United States (Milgram, 1967; Travers and Milgram, 1969; Korte and Milgram, 1970). They found this number to be around six, which forms the first empirical evidence for the common notion that any two individuals are separated by *six degrees of separation* or the so-called *small world phenomenon*. Subsequently, objections such as sample selection bias and low chain completion rates were raised against these findings (Kleinfeld, 2002). However, a similar experiment was repeated by Dodds et al. (2003) at the global scale by means of e-mail messages, and they also obtained results similar to those of Milgram's experiments by taking the chain attrition rates into account. Therefore, in spite of some drawbacks of these empirical studies, it is still believed that people on the world are connected to each other through a small number of acquaintances (Schenettler, 2009b).

Despite considerable research on the small-world phenomenon spanning more than 40 years, precise analytical expressions, apart from some bounds (Kleinberg, 2000a,b; Franceschetti and Meester, 2006), explaining the hidden dynamics of the search in social networks are still unknown. In this paper, we consider the problem of search in social networks analytically, seeking a mathematical confirmation of the above-noted experimental phenomena. We obtain precise analytical expressions for the time required for social search by modeling the time required for such search in terms of the delay experienced by messages to travel from source to target in a small-world network. We will refer to this quantity as *the delay of social search*, or alternatively as *the length of acquaintance chains connecting two individuals*.

Our results show that individuals tend to use their short-range connections, and the delay of social search rises linearly, when the social separation between source and target individuals is small. On the other hand, as this separation increases, long-range connections are more frequently used, and the delay of social search rapidly saturates to a constant value and stays almost the same for all large values of the separation. These results are consistent with the empirical observations of Travers and Milgram (1969) and Dodds et al. (2003). Furthermore, they clearly reveal the dynamics of the social search, and the range of source-target separations for which long-range connections positively contribute to such search. Our expressions also quantify the effect of the number of long-range connections on the time required for social search. In particular, we show that the time required benefits from social equality, and in contrast to intuition and some previous results (Adamic et al., 2001; Barabási, 2002), having hubs with large numbers of long-range connections does not help significantly if the social search criteria for selecting intermediaries are based on social dimensions, such as geographic location, occupation, education, etc., rather than the number of friends of an individual.

In addition to social networks, examples of small-world networks also abound in biology, neuroscience, physics, etc., as well as in human-made networks, such as the Internet, power grids, businesses, etc., (Strogatz, 2001; Watts, 2004; Bassett and Bullmore, 2006; Malaquias et al., 2006; Cohen et al., 2007). Therefore, we believe that the analysis presented in this paper has broader implications, and can also aid scientists in many diverse fields who wish to understand intrinsic properties of complex networks. In particular, along with Kleinberg's results (Kleinberg, 2000a,b), this analysis may be helpful to engineers in understanding how messages are routed in complex communication networks, to biologists and sociologists in understanding how infectious diseases spread in society, and to neuroscientists in discovering how signals are transmitted in the brain and other nervous systems of living organisms.

### 1.1. *Small-world Networks: Sociological Perspectives*

Small-world networks are network models consisting of two types of connections: *short-range connections* and *long-range connections*. These networks have two different but closely related interpretations in network analysis. The first interpretation is that they describe networks whose properties are intermediate to order and disorder (Watts and Strogatz, 1998; Watts, 2003). The second interpretation is related to the *strength of weak ties* concept introduced by Granovetter (1973, 1983), where short-range and long-range connections replace strong and weak ties, respectively. The analysis presented in this paper holds under both interpretations. Therefore, both are briefly explained below.

We begin with the order-disorder interpretation of small-world networks. Social networks are formed as a result of two conflicting forces: structure and randomness. The forces in favor of randomness arise as a result of personal choices such as moving to another country, going to college or graduate school in another city, joining hiking, climbing or camping clubs, etc. From an individual's point of view, these forces are not random, and are derived from a complicated mixture of one's personal history, psychology and interests; however, as far as an outside observer of a large population of individuals is concerned, these forces can be treated as being random (Watts, 2003). The importance of these forces is to form bridges between different social circles. Surprisingly very short paths are established between socially distant individuals, and social networks among people become small by means of these bridges (Watts and Strogatz, 1998). Each personal choice (e.g., moving to another country or joining social clubs) places an individual into a new social orbit. After making a personal choice, structural forces come into play, and drive an individual to form friendship circles within these social communities led by personal choices. These structural forces can be considered as constraints imposed on individuals by the structure of the social environment surrounding them, such as their workplace, the neighborhood they live in and their tendency to become friends with others sharing common interests. In small-world networks, structural forces are modeled by assuming a local structured friendship circle for each individual (Watts and Strogatz, 1998).

The second interpretation becomes salient when we consider that connections among people in a social network can be divided into two types: *strong ties* and *weak ties*. Interpersonal ties tend to become stronger as the time and effort invested in a relationship, the frequency of interactions, intimacy, reciprocity, and the degree to which friends, values and interests are shared in common increase (Granovetter, 1973; Homans, 1950; Newcomb, 1961; Rapoport, 1953a,b, 1954). Relationally strong ties are important for enhancing the trust and influence between people in a social network. However, they are structurally weak for the search for critical resources in social networks as well as for the diffusion of information and innovations. This appears to be primarily because strong ties tend to be transitive and form closed friendship triads as claimed by Granovetter (1973) and as supported by some empirical evidence (Davis and Leinhardt, 1971). That is, the probability of having an interpersonal tie between two individuals $A$ and $B$ becomes high if both are connected to another individual $C$ through a strong tie. As a result of these structural biases in interpersonal ties, individuals connected to one another by means of strong ties become socially close, and are usually exposed to the same set of resources and information, which inhibits the search and diffusion processes in social networks (Granovetter, 1973, 1983).

In contrast to strong ties, weak ties are causal relationships in the form of long bridges connecting socially distant individuals in social networks. These individuals are located in different social groups, they have access to different resources and information distributed over the social network, and they have few friends in common. These features of weak ties facilitate the spread of a search process or a diffusion process initiated by a seed individual in a close-knit social community to socially distant locations in the network. As Granovetter (1973) claims, "... whatever is to be diffused can reach a larger number of people, and traverse greater social distance (i.e., path length) when passed through weak ties rather than strong." This is the structural strength of relationally weak ties.

In this paper, we follow the convention introduced by Watts and Strogatz (1998), and categorize interpersonal ties as long-range connections and short-range connections depending on the social distance spanned by each of these ties. Our analytical results quantitatively characterize the importance of long-range connections for locating a target in social networks when source and target individuals are socially separated from each other. Our analytical results also reveal the importance of local friendship circles formed by short-range connections for locating a target in social networks.

### 1.2. *Decentralized Search in Small-world Networks*

The decentralized search process that we will focus on is inspired by Milgram's letter referral experiments (Milgram, 1967; Travers and Milgram, 1969; Korte and Milgram, 1970). After its first appearance, the letter referral technique and its modifications (i.e., using the telephone and e-mail instead of the post) attracted substantial attention and was repeated in various social contexts to discover social network structure (Erickson and Kringas, 1975; Guiot, 1976; Lin et al., 1977; Weimann, 1983; Dodds et al., 2003). A typical realization of such an experiment is as follows. Two socially separated individuals are selected randomly [1] . One of them is assigned as the message originator (i.e., source),

[1] *Randomly* in this context means that there is a sample selection process that involves stochastic dynamics while recruiting message originators (i.e., source individuals) and message recipients (i.e.,

and the other is assigned as the message recipient (i.e., target) to whom the message will be delivered. Depending on the social context, source and target individuals are located in the same country but in different states (Milgram, 1967; Travers and Milgram, 1969), or in different countries (Dodds et al., 2003), or in the same institution but have different professional ranks (Shotland, 1976; Stevenson et al., 1997), or are intentionally chosen to belong to different racial, religious or cultural backgrounds (Korte and Milgram, 1970; Weimann, 1983). The source is provided with some basic information about the target such as address and occupation, and she is allowed to send the message only to others whom she knows on a first-name basis. Therefore, the source is not allowed to send the message to the target directly unless she knows the target on a first-name basis. Intermediate message holders repeat the same step until the message reaches the target.

To repeat the steps of the small-world experiment explained above analytically, we first place source and target nodes at arbitrary positions inside a network domain that represents the social space[2]. The initial social separation between source and target nodes becomes an important factor determining the length of acquaintance chains connecting them. Therefore, we are able to tap into complex small-world search dynamics, and identify lengths of chains connecting socially close-by individuals and socially distant individuals by tuning initial displacement parameter. After the initial placement, the source node chooses one of its direct contacts, either short-range or long-range, that is closest to the target node as a next hop message holder. Then, each message holder repeats the same last step by forwarding the message to her contact closest to the target node until the message is delivered to the target node. The first implicit assumption in these steps is that the source node knows where the target node is located in the social space, which is equivalent to providing the message originator in small-world experiments with some basic information about the message recipient identifying her position in the social space. The second implicit assumption in these steps is that each message holder is able to select her best contact closest to the target individual. To the extent that the initial information supplied for the message originator in empirical small-world studies identifies the target individual's social location and each message holder is able to choose her best contact closest to the target individual, the decentralized search mechanism analyzed in this paper mimics the empirical small-world experiments more precisely, and our results better estimate the lengths of acquaintance chains connecting individuals for all scales of social separation between them.

target individuals). For example, Milgram and his students (Milgram, 1967; Travers and Milgram, 1969; Korte and Milgram, 1970) used mail lists and newspaper advertisements for sample recruitment, whereas Dodds et al. (2003) used the World Wide Web. Such a recruitment process may introduce sample selection bias to empirical small-world studies. In this paper, we do not run into the sample selection problem since we have complete flexibility in where the source and target individuals are located in the network.

[2] We will use the term *node* as a technical term to refer to a network element modeling an individual in a social network.

## 2. Model Details and the Derivation of the Delay Formula

### 2.1. *Model Details*

To model the social space over which individuals are distributed, we use three different network domains: *rectangular*, *spherical* and *circular*. We obtain the delay formula for each different network domain, which provides a robustness check of our results against the effect of network geometry on the delay. Our circular network model is similar to the small-world network model introduced by Watts and Strogatz (1998), which has been analyzed in many studies in different contexts in great details (e.g., Centola and Macy (2007); Newman et al. (2000)). Different meanings can also be attributed to our network models in other areas of science in which small-world networks emerge. For example, the spherical network model can find direct applications when modelling connections among nerve cells in the human brain due to geometrical similarities (Bassett and Bullmore, 2006; Malaquias et al., 2006). Similar analysis can also be extended to more abstract topological measure spaces (Rudin, 1987), however we will not pursue this direction in this paper since it introduces unnecessary notational and mathematical complexity.

The side lengths of the rectangular network domain and the radius of spherical and circular network domains are $R$ distance units. The radius of local friendship circles of individuals is $r$ distance units. It should be noted that we use the term "distance" here to mean social distance, that is, the topology here is social rather than physical. We assume that individuals are randomly distributed (from uniform distribution) over the network domain, and any two individual $A$ and $B$ are short-range contacts of each other if their social separation is smaller than $r$. The location of individual $i$ in the social space will be represented by $X_i$. Each individual also selects random number, $N$, of long-range contacts randomly over others that do not lie in her local friendship circle. $N$ can be drawn from any given discrete probability distribution $Q$, where $Q(n) = P(N = n)$, such as power law, Poisson, geometric or uniform. We are able to determine the delay of social search for any given distribution of $N$.

In order to identify the lengths of acquaintance chains connecting individuals for all scales of social separation, we place source and target nodes at arbitrary positions inside the network domain. The source and target nodes' locations will be denoted by $X_s$ and $X_t$, respectively. Three typical realizations of our network models are shown in Figs. 1, 2 and 3 below. The green regions in these figures represent the local friendship circle of a generic individual $A$. For the spherical network domain, all nodes are placed on the surface of the sphere, and the local friendship circle of $A$ is the spherical cap centered at $X_A$.

The generic name we give to all networks constructed in this manner is the *octopus model* for social network modeling in particular and complex network modeling in general, where the body of the "octopus" is formed by the local knit friendship circle of an individual, and the causal and relatively weak ties form the legs of the "octopus". It is a generalization of random geometric graphs (Penrose, 2003) to social network modeling by means of random long-range connections.

The octopus model captures the order-disorder properties of the Watts and Strogatz (1998) model due to the threshold rule for forming local contacts. The substrate network in the octopus model is formed on continuum network domains, whereas it usually

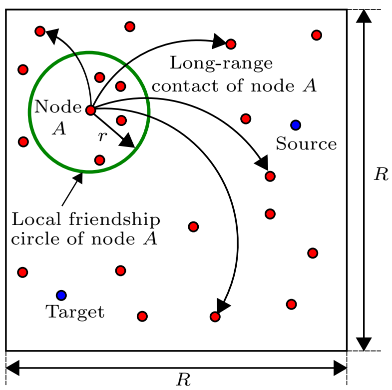


Fig. 1. A typical realization of the social network for the rectangular network domain.

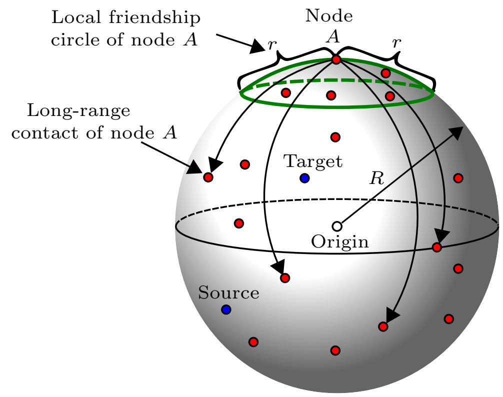


Fig. 2. A typical realization of the social network for the spherical network domain.

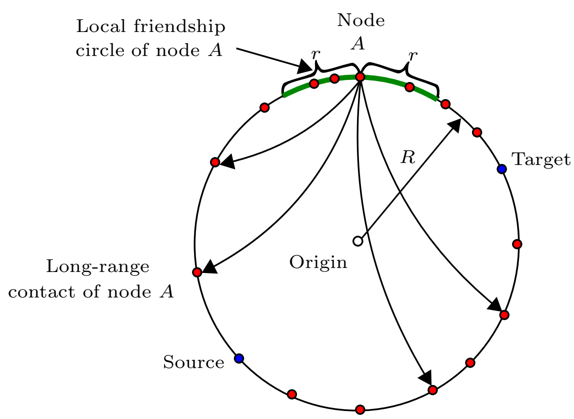


Fig. 3. A typical realization of the social network for the circular network domain.

consisted of discrete latices (Watts and Strogatz, 1998; Kleinberg, 2000a,b) in previous studies. The octopus model also parametrizes the generation of complex networks in terms of the distribution of $N$, whereas number of long-range connections is fixed, and all nodes have the same number of friends in the original small-world network model put forward by Watts and Strogatz (1998). The power of such a parameterization of complex networks is that we can now generate a wide variety of networks by changing the distribution of the number of long-range connections. For example, this model can also mimic scale-free features of the Barabási (2002) model when $N$ is drawn from a power law distribution.

One can also give further interpretations to this model by using Granovetter's insights (Granovetter, 1973, 1983). Granovetter (1973) claims that the tie between two individ-

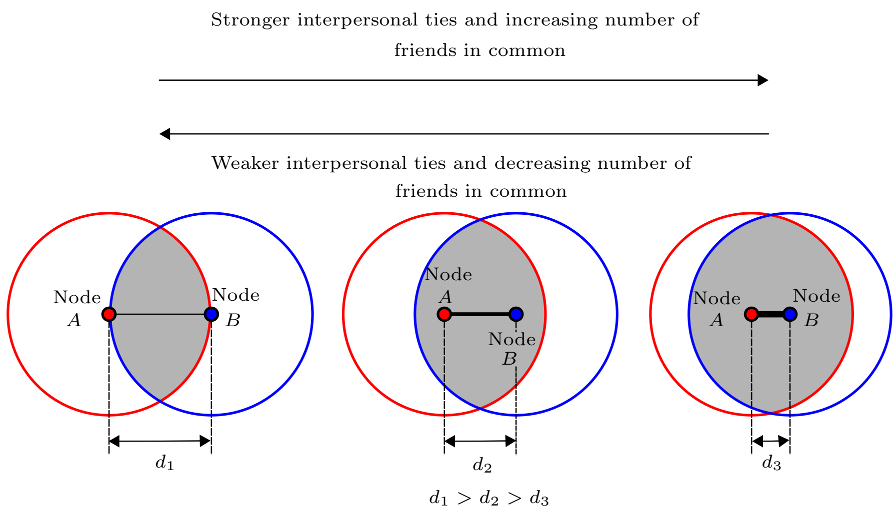


Fig. 4. Strong and weak ties.

uals tends to be stronger if they share more friends in common. In our model, as two individuals $A$ and $B$ become socially closer to each other, their friendship circles start to overlap more and more. The amount of overlap between their friendship circles is represented by the gray area in Fig. 4, which is proportional to the average number of friends shared in common by $A$ and $B$. Therefore, our model predicts stronger interpersonal ties as the social separation between two individuals becomes smaller. Whenever a message is passed through such a close-by contact, it traverses a smaller social distance when compared to forwarding it through a long-range connection; a fact that was also predicted by Granovetter's model (Granovetter, 1973). These interpretations are qualitatively shown in Fig. 4.

### 2.2. *Derivation of the Delay Formula*

We now present the derivation of the delay formula for social search. This part of the paper will be rather technical when compared to other parts. The details of the calculations can be skipped at the first reading as the exposition of the next sections is based on our main findings summarized in Tables 1 and 2. To avoid repetitions, we show steps of calculations only for rectangular network domains. Local friendship circles for rectangular network domains are easy to represent visually, which further facilitates the exposition and understanding of the derivation of the delay formula. The same steps can be repeated to derive the delay formula for spherical and circular network domains.

Our method of obtaining the analytical expression for the delay of social search is fairly general, and can also be employed in other discrete small-world models. It depends on a Markov-chain-based first-step analysis. To make sure that the social network is almost surely connected, we focus on the dense network limit in which there are infinitely many nodes lying inside the network domain. This assumption is very frequently employed in the random graph literature (e.g., Bollobás, 2001; Penrose, 2003) for technical purposes to guarantee network connectivity. If a connected discrete lattice is used as a substrate

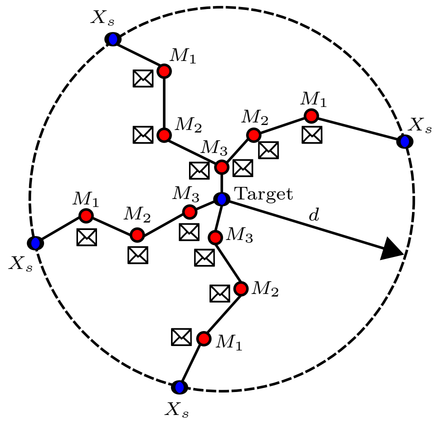


Fig. 5. Pictorial illustration of the spherical symmetry property of $T(d)$.

network as in the Watts and Strogatz (1998) model or in the Kleinberg (2000a,b) model, one can further remove this technical assumption. We also have performed agent-based simulations to identify the effects of the dense network assumption on the delay formula, and obtained fairly good qualitative and quantitative matches between simulation and analytical results for networks when the average number of local contacts per node is in between Dunbar's number 150 (Dunbar, 1993; Hill and Dunbar, 2003) and Killworth's estimate 290 (McCarty et al., 2001; Bernard et al., 2001) for the average size of personal networks. We report the details of our agent-based simulations in Section 3.

We denote the delay of social search in the dense network limit by $T$, the location of the message at the $k^{\text{th}}$ hop by $M_k$, and the distance between any two points $x$ and $y$ inside the network domain by $\rho(x, y)$. Note that the distance metric $\rho$ can be different from the usual Euclidean metric for different network geometries. For example, the distance between any two points on the surface of a sphere is measured in terms of the smaller arc length of the great circle connecting them. We always set $M_0$ to $X_s$ since the social search starts from the source node. The social search stops when the message reaches the target node. Therefore, $T$ can be analytically expressed as

$$T = E\left[\min\{k \geq 0 : M_k = X_t\}\right].$$

$T$ satisfies two important properties. First, it is a *spherically symmetric* function, i.e., it depends only on the social separation, $d = \rho(X_s, X_t)$, between source and target nodes. That is, the average length of the acquaintance chains connecting source and target nodes is the same for all source node positions $X_s$ satisfying $X_s \in \{x \in \mathcal{D} : \rho(X_t, x) = d\}$, where $\mathcal{D}$ represents the network domain. This property follows from the symmetry in the problem, and is illustrated in Fig. 5. Therefore, we represent $T$ as a function of the social separation between source and target nodes for the rest of the paper.

Second, $T(d)$ becomes a *step function* with jumps at integer multiples of $r$. It is easy to see that this is correct for $0 < d < r$ and $r \leq d < 2r$. In particular, $T(d) = 1$ for $0 < d < r$ since source and target nodes are local contacts of each other. $T(d) = 2$ for $r \leq d < 2r$ since source and target nodes are guaranteed to have at least one common friend lying inside the intersection of their friendship circles (i.e., the gray area in Fig. 6) for this

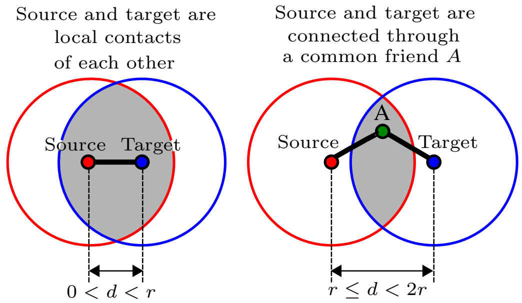


Fig. 6. Length of acquaintance chains connecting source and target nodes when $0 < d < r$ and $r \leq d < 2r$.

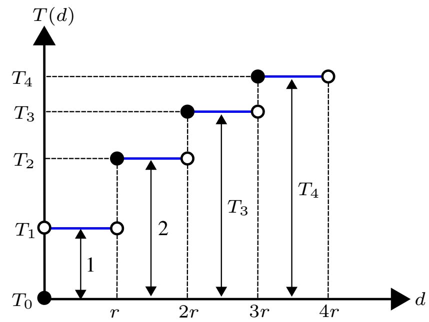


Fig. 7. Step function property of $T(d)$ with jumps at integer multiples of $r$.

range of social separation between them in the dense network limit. By induction, the step function property of $T(d)$ can be shown to hold for all values of $d$. These properties are shown in Figs. 6 and 7. We set $T(0)$ to 0 since having 0 social separation between source and target nodes means that they are the same individual; therefore, there is no need for message forwarding. We define the heights of the step function as

$$
\begin{aligned}
T_0 &= T(0) = 0, \\
T_1 &= T(d) = 1 \text{ for } 0 < d < r, \\
T_2 &= T(d) = 2 \text{ for } r \leq d < 2r, \\
T_k &= T(d), \text{ for } (k-1)r \leq d < kr \text{ and } k \geq 3.
\end{aligned}
$$

We first illustrate the calculations for $T_3$, and then generalize them to any $k \geq 4$. When the social separation between source and target nodes is in $[2r, 3r)$, the acquaintance chains connecting them can be categorized into two types depending on where the message is located after the first hop. In the first type, the message after the first hop is located at an intermediate message holder lying inside the local friendship circle of the target node. Such an acquaintance chain emerges if and only if the source node happens to have a long-range contact who is a local contact of the target node. The delay of social search becomes 2 if there is such an acquitance chain connecting source and target nodes: the message is first forwarded through the long-range contact, and then delivered to the target node through one of its local contacts. The top solid path in Fig. 8 is an example message trajectory in which it takes two hops for the message to reach the target.

In the second type, the message after the first hop is located at an intermediate message holder whose social separation from the target node is smaller than $2r$ but greater than $r$. Such an acquaintance chain emerges if and only if the source node does not share any common friend with the target node and has at least one contact who is not a friend of the target node but sharing a common friend with the target node. For such chains, the source node has two options when selecting the next hop message holder: the message is forwarded through either a long-range contact or a short-range contact at the first step. The bottom solid path in Fig. 8 shows an example message trajectory in which the message is first forwarded through one of source node's long-range contacts. On the other hand, the bottom dotted path in Fig. 8 shows an example message trajectory in

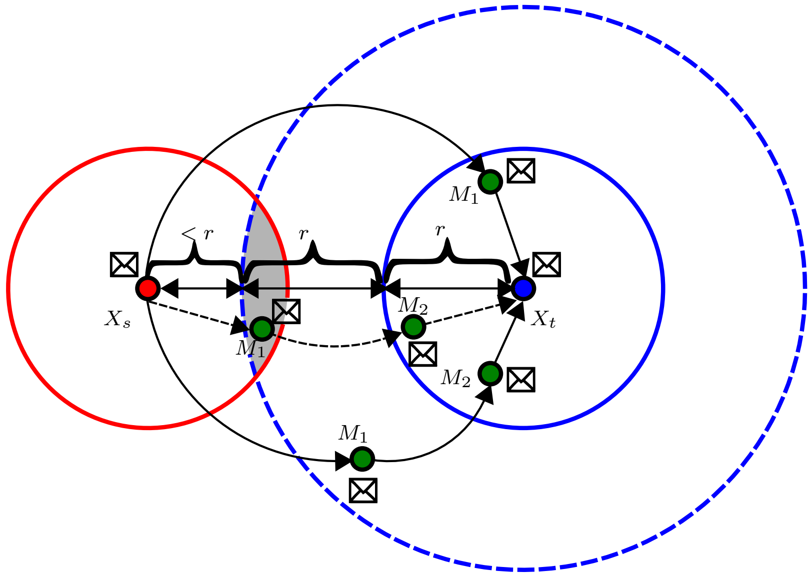


Fig. 8. Example trajectories of the message and the acquaintance chains connecting source and target nodes when $2r \leq \rho(X_s, X_t) < 3r$.

which the message is first forwarded through one of source node's short-range contacts. In either case, the message at the first hop is delivered to an intermediate message holder whose social separation from the target node is in between $r$ and $2r$. Then, it takes two hops from this point on to deliver the message to the target node since $T(d) = 2$ for $r \leq d < 2r$ as explained above. Therefore, the delay of social search becomes 3 for such chains.

As a result, the delay of social search critically depends on the location of the message $M_1$ after the first hop. These observations are pictorially shown in Fig. 8, and are formally expressed as

$$\begin{aligned} T_3 &= 1 + E\left[T\left(\rho\left(M_1, X_t\right)\right)\right] \\ &= 1 + E\left[T\left(\rho\left(M_1, X_t\right)\right)\mathbb{1}_{\{0<\rho(M_1,X_t)<r\}}\right] + E\left[T\left(\rho\left(M_1, X_t\right)\right)\mathbb{1}_{\{r\leq\rho(M_1,X_t)<2r\}}\right] \\ &= 1 + T_1 P\left(0 < \rho(M_1, X_t) < r\right) + T_2 P\left(r \leq \rho(M_1, X_t) < 2r\right) \\ &= 1 + P\left(0 < \rho(M_1, X_t) < r\right) + 2P\left(r \leq \rho(M_1, X_t) < 2r\right), \end{aligned}$$

where $\mathbb{1}_{\{X\in\mathcal{E}\}}$ is an indicator function that equals 1 when $X$ belongs to the set $\mathcal{E}$ and equals 0 otherwise.

If the source node has a long-range contact lying inside the local friendship circle of the target node, then the message after the first hop is located at an intermediate message holder whose social separation from the target node is not greater than $r$ (i.e., $0 < \rho(M_1, X_t) < r$). The probability of a given long-range contact of the source node lies in the local friendship circle of the target node is equal to $\alpha = \frac{\pi r^2}{R^2 - \pi r^2}$. Thus, $P\left(0 < \rho(M_1, X_t) < r\right)$ is calculated as

$$
\begin{aligned}
P\left(0<\rho(M_1,X_t)<r\right) &= \sum_{n=1}^{\infty} Q(n) P\left(0<\rho(M_1,X_t)<r \mid N=n\right) \\
&= \sum_{n=1}^{\infty}\sum_{i=1}^{n} Q(n)\binom{n}{i}\alpha^i(1-\alpha)^{n-i} \\
&= \sum_{n=1}^{\infty} Q(n)(1-\alpha)^n\left(\left(\frac{1}{1-\alpha}\right)^n-1\right) \\
&= 1-\varphi(1-\alpha),
\end{aligned}
$$

where $\varphi(t)$, $t>0$, is the probability generating function of the distribution $Q$, defined as $\varphi(t)=E\left[t^N\right]$. If the source node does not have any long-range contacts lying in the local friendship circle of the target node, the message is guaranteed to be delivered to an intermediate message holder whose social separation from the target node is not greater than $2r$ in the dense network limit. As a result, $P\left(r \le \rho(M_1,X_t)<2r\right)=\varphi(1-\alpha)$. Therefore,

$$
T_3 = 2+\varphi(1-\alpha) \text{ for } 2r \le d < 3r.
$$

For a general $k \ge 4$, we proceed as follows:

$$
\begin{aligned}
T_{k+1} &= 1+E\left[T\left(\rho\left(M_1,X_t\right)\right)\right] \\
&= 1+E\left[\sum_{m=1}^{k} T_m \mathbb{1}_{\{(m-1)r\le\rho(M_1,X_t)<mr\}}\right] \\
&= 1+T_k P\left((k-1)r \le \rho(M_1,X_t)<kr\right)+\sum_{m=1}^{k-1} T_m P\left((m-1)r\le\rho(M_1,X_t)<mr\right).
\end{aligned}
$$

The probabilities $P\left((m-1)r\le\rho(M_1,X_t)<mr\right)$ for $1\le m\le k-1$ and $P\left((k-1)r\le\rho(M_1,X_t)<kr\right)$ are calculated in Appendix A as

$$
\begin{aligned}
&P\left((m-1)r\le\rho(M_1,X_t)<mr\right)=\varphi\left(\beta_m\right)-\varphi(\beta_{m+1}) \text{ and} \\
&\ P\left((k-1)r\le\rho(M_1,X_t)<kr\right)=\varphi(\beta_k),
\end{aligned}
$$

where $\beta_m = 1-\frac{\pi r^2(m-1)^2}{R^2-\pi r^2}$. As a result, $T_{k+1}$ is given by the following iterative equation:

$$
T_{k+1} = 1+T_k\varphi(\beta_k)+\sum_{m=1}^{k-1} T_m\left(\varphi(\beta_m)-\varphi(\beta_{m+1})\right).
$$

We can further simplify the iterative equation for $T_k$ as follows. We define $u_k = T_{k+1}-T_k$. Then, $u_k=\varphi(\beta_k)u_{k-1}$, which further leads to $u_k=\prod_{i=1}^{k}\varphi(\beta_i)$. Finally, the solution for $T_k$ for rectangular network domains is obtained as

$$
T_{k+1} = 1+\sum_{j=1}^{k}\prod_{i=1}^{j}\varphi(\beta_i). \tag{1}
$$

There are some advantages and disadvantages of using rectangular network domains to model the social space. A disadvantage of using rectangular network domains is the

Table 1
General formula for the delay of social search.

| General Delay Formula ($T_k$)[a] | Parameters ($\beta_i$) | | |
|---|---|---|---|
| | Rectangular | Spherical | Circular |
| $T_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \varphi(\beta_i)$[b] | $\beta_i = 1 - \frac{\pi r^2 (i-1)^2}{R^2 - \pi r^2}$ | $\beta_i = \frac{\cos\left(\frac{r}{R}\right) + \cos\left((i-1)\frac{r}{R}\right)}{1+\cos\left(\frac{r}{R}\right)}$ | $\beta_i = \frac{\pi - i\frac{r}{R}}{\pi - \frac{r}{R}}$ |

[a] $T_k = T(d)$ for $(k-1)r \leq d < kr$, where $r$ is the radius of local friendship circles of individuals, and $T(d)$ is the delay of social search when the social separation between source and target individuals is $d$.

[b] $\varphi(t)$ is the probability generating function of the number of long-range connections, $N$, per node. It is defined as $\varphi(t) = E[t^N]$.

edge effects that occur in this type of networks. When a node comes closer to the edges of the network, it starts to have less number of local contacts. As a result, it becomes harder and harder to find the target node as it becomes closer and closer to the edges of the network. Our calculations presented above are correct when the target node is sufficiently close to the center of the network[3]. On the other hand, an advantage of having edge effects in rectangular network domains is to be able to model the effect of socially segregated individuals on the delay of social search by placing target nodes close to the edges of the network. Our analysis above can also be extended to model socially segregated individuals at the expense of more tedious calculations by deriving equations for the area of the intersection between partially intersecting disks and rectangles. However, we will not pursue this direction since our aim in this paper is to obtain simple, intuitive and insightful expressions to understand fundamental social search dynamics rather than obtaining complicated formulas for rare situations. To this end, we extend similar calculations to spherical and circular network domains where no edge effect occurs.

When similar analysis is repeated for spherical and circular network domains, the same functional form for the delay of social search is obtained. In particular, the delays of social search for spherical and circular network domains are also given by Eq. (1), with $\beta_i = \frac{\cos\left(\frac{r}{R}\right) + \cos\left((i-1)\frac{r}{R}\right)}{1+\cos\left(\frac{r}{R}\right)}$ for spherical domains and $\beta_i = \frac{\pi - i\frac{r}{R}}{\pi - \frac{r}{R}}$ for circular domains. This finding is a justification of the universality of our method for obtaining the delay formula as well as serving as a robustness check for our delay formula against different network geometries.

These results are summarized in Tables 1 and 2, the latter of which also providing explicit delay formulas for social networks with various node degree distributions for the number of long-range connections.

## 3. Agent Based Simulations for the Delay of Social Search

In this section, we present agent based simulations to verify the analytical expressions derived above and to understand the effect of the dense network assumption on the delay formula for social search. Agent based modeling (Macy and Willer, 2002; Epstein, 1999) is a powerful bottom-up computational tool to explain many diverse complex

[3] The technical condition for this is to have the disk centered at the target node with radius $\rho(X_t, X_s)+r$ contained inside the network domain.

Table 2
The delay of social search with various node degree distributions for the number of long-range connections.

| Network Type | Delay Formula | Node Degree Distribution for the Number of Long-range Connections |
|---|---|---|
| Scale-free Networks | $T_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \frac{\text{Polylog}(\gamma, \beta_i)}{\zeta(\gamma)}$ [a] | $P(N = n) = \frac{n^{-\gamma}}{\zeta(\gamma)}$, $N \in \{1, 2, 3, \cdots\}$ |
| Poisson Networks | $T_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \exp(-\lambda(1 - \beta_i))$ [b] | $P(N = n) = \frac{\lambda^n \exp^{-\lambda}}{n!}$, $N \in \{0, 1, 2, \cdots\}$ |
| Geometric Networks | $T_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \frac{p}{1 - (1 - p)\beta_i}$ [c] | $P(N = n) = p(1 - p)^n$, $N \in \{0, 1, 2, \cdots\}$ |
| Binomial Networks | $T_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} (1 - p + p\beta_i)^M$ [c,d] | $P(N = n) = \binom{M}{n} p^n (1 - p)^{M-n}$, $N \in \{0, 1, \cdots, M\}$ |
| Uniform Networks | $T_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \frac{1 - \beta_i^{M+1}}{(M + 1)(1 - \beta_i)}$ [d] | $P(N = n) = \frac{1}{M + 1}$, $N \in \{0, 1, \cdots, M\}$ |
| Bernoulli Networks | $T_{k+1} = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \left(1 - p + p\beta_i^M\right)$ [c,d] | $P(N = M) = 1 - P(N = 0) = p$, $N \in \{0, M\}$ |

[a] $\gamma > 1$ is the power law decay exponent. $\zeta(\gamma)$ is the Reimann zeta function defined as $\zeta(\gamma) = \sum_{n=1}^{\infty} \frac{1}{n^\gamma}$. Polylog$(\gamma, \beta_i)$ is the polylogarithm function defined as Polylog$(\gamma, \beta_i) = \sum_{n=1}^{\infty} \frac{\beta_i^n}{n^\gamma}$.

[b] $\lambda > 0$ is the mean value of the Poisson distribution.

[c] $0 < p < 1$ is the shaping parameter for geometric, binomial and Bernoulli distributions. We also allow it to take values 0 and 1 for Bernoulli networks.

[d] $M$ is a natural number determining the upper bound for binomial, uniform and Bernoulli distributions.

and enigmatic social phenomena including social segregation (Schelling, 1971), cultural polarization (Axelrod, 1997), emergence of norms (Centola et al., 2005), emergence of cooperation (Nowak and Sigmund, 1992), and growth and collapse of populations (Axtell et al., 2002). In contrast to the common trend in agent based modeling, the purpose of our simulations is not to predict and explain an enigmatic social phenomenon but rather to verify the derived formulas predicting and explaining such a phenomenon. To this end, all agents in the network in our simulations work together to achieve the collective macro goal of delivering a message originated from a source node to a target node.

Our simulation set-up is as follows. An agent in the network can be one of the following three types: *source node*, *target node* and *relay node*. For each realization of the social search process, there are one source node, one target node and a variable number of relay nodes (in large quantities) in the network. Relay nodes are randomly distributed over the network domain, and source and target nodes are placed at arbitrary positions inside the network domain. The adjacency matrix consisting of social ties among agents is formed according to the threshold rule for forming local contacts (i.e., any two agents are local contacts of each other if their social separation is smaller than $r$) and to the random long-range contact formation rule (i.e., each agent has a random number of long-range connections formed randomly over all agents lying outside of its local friendship circle).

Based on its local information at the micro level, each agent tries to minimize the delay of social search by selecting its contact closest to the target node as a next-hop message holder. This assumes myopic agents having information horizons limited to their immediate contacts. On the other hand, if agents have access to greater macro level information about their contacts, such as knowing the locations of friends of their friends and utilize such information, this local rule for choosing next-hop message holders leads to sub-optimum results even if it is optimum at the micro level.

Since the primary aim of this forwarding rule is to mimic the behavior of individuals in empirical small-world experiments and to obtain representative results for the lengths of acquaintance chains connecting individuals, this observation poses an important question: how much network information is used by individuals when choosing their next-hop neighbors in empirical small-world experiments? We contend that this information is limited to only immediate contacts, and therefore the above micro rule is a reasonable approximation of behavior observed in empirical small-world experiments. The reason for this contention is related to cognitive information processing capabilities of human beings that play critical roles in determining the average size (Dunbar's number) for personal networks (Dunbar, 1993). The number of friends of friends of an individual is typically very large [4] , and therefore it is hard to process all social information pertaining to that many people even if a message holder in a typical small-world experiment was made aware of the whole or some parts of such information in the past. The validity of this forwarding rule is further reinforced by the global small-world experiment conducted by Dodds et al. (2003), who reported that people usually nominate one of their friends as a next-hop message holder because they think that she is socially close to the target node (i.e., in terms of geography, occupation or education) but not because of some other reason, such as knowing that one of the contacts of the selected friend is close to the target. (See Table 2 of Dodds et al. (2003).)

There are various simulation parameters over which we have control: network geometry, node degree distribution for long-range connections, network size, size of the local friendship circles of nodes, number of nodes and source-target separation. Below, we report our simulation results for fixed network geometry (i.e., rectangular network domains) and fixed node degree distribution for long-range connections (i.e., one long-range contact per node). One can also arrive at similar quantitative and qualitative conclusions when these parameters are varied. We vary the network size relative to the radius of the local friendship circle of nodes by considering $R$-by-$R$ rectangular network domains with $R = 50r, 100r, 250r$ and $500r$. For each different size of the network domain, the number of nodes in the network is varied so that nodes have approximately 15, 30, 80, 160 and 320 local friends on the average. Source-target separation is varied between minimum and maximum possible values for rectangular network domains. For any given source-target

[4] A simple estimate for this number can be obtained as follows (Bernard et al., 2001; Newman, 2003). Assume each individual has $c$ friends on the average. Then, the number of friends of friends of an individual is equal to $c(c-1)/\lambda^2$, where $\lambda$ is the *lead-in-factor* to account for triad closure. Estimates for $c$ vary; two most commonly accepted ones being 150 (Dunbar, 1993) and 290 (McCarty et al., 2001). Taking $c = 150$ and $\lambda = 1.6$ as in Bernard et al. (2001), the number of friends of friends of an individual can be estimated to be 8730.

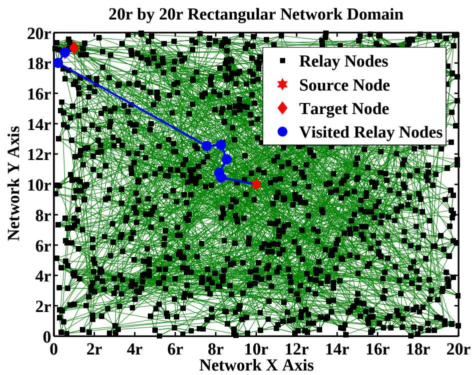


Fig. 9. A typical realization of the social search process in which the delay becomes equal to 8.

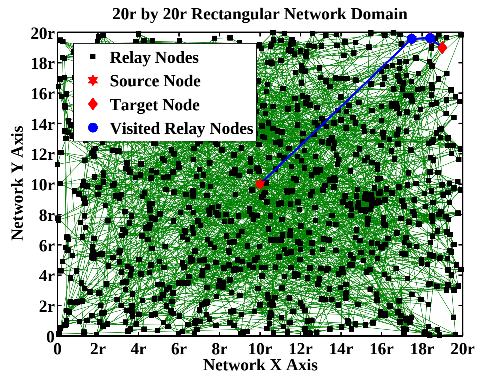


Fig. 10. A typical realization of the social search process in which the delay becomes equal to 3.

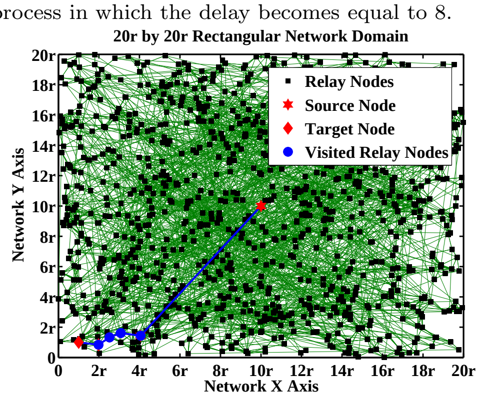


Fig. 11. A typical realization of the social search process in which the delay becomes equal to 5.

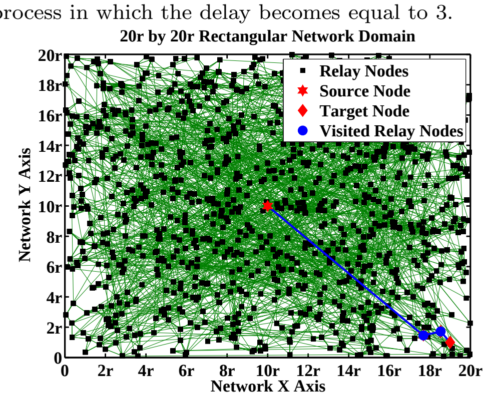


Fig. 12. A typical realization of the social search process in which the delay becomes equal to 3.

separation, the number of chains initiated is set to at least 1600 to obtain statistically significant results for the delay of social search[5].

We built our agent-based simulator by using the C programing language, and all of our simulations used 200 CPUs running in parallel within the *TIGREES* (Terascale Infrastructure for Groundbreaking Research in Engineering and Science) supercomputing facility at Princeton University. Even with this much computational power, some of the simulations for large network domains can took up to one week to complete. On the other hand, our analytical formulas take only several seconds to compute on an ordinary personal computer, which is a tremendous gain in terms of time and effort required to obtain predictive and explanatory results in social search processes. This point further reinforces the utility of our analytical results in the study of social search processes. Our simulation setting and parameters are summarized in Table 3.

In Figures 9 - 12, we show four typical realizations of the social search process initiated by a source node located at the center of a $20r$-by-$20r$ rectangular network domain containing 1000 relay nodes. The number of long-range connections per node is set to

[5] For small network sizes $R = 50r, 100r$ and $250r$, the number of chains initiated is set to 16000. For $R = 500r$, the number of chains initiated is set to 1600 due to limitations on simulation run times.

Table 3. Summary of simulation setting and parameters.

| Network Geometry | Agent Types | | | Macro Level Goal | Micro Level Message Forwarding Rule | Simulation Environment |
|---|---|---|---|---|---|---|
| $R$-by-$R$ rectangle | One source node | One target node | Variable number of relay nodes[a] | Deliver a message from the source node to the target node | Select the contact closest to the target node as a next hop message holder | C simulator run on clusters |

| Network Size | Average Number of Local Contacts per Node | | | | | Number of Long-range Contacts per Node | Source-target Separation | Number of Chains Initiated |
|---|---|---|---|---|---|---|---|---|
| $R = 50r$ | 15 | 30 | 80 | 160 | 320 | 1 | $0 - 25r$ | $16 \cdot 10^3$ |
| $R = 100r$ | 15 | 30 | 80 | 160 | 320 | 1 | $0 - 50r$ | $16 \cdot 10^3$ |
| $R = 250r$ | 15 | 30 | 80 | 160 | 320 | 1 | $0 - 125r$ | $16 \cdot 10^3$ |
| $R = 500r$ | 15 | 30 | 80 | 160 | 320 | 1 | $0 - 250r$ | $16 \cdot 10^2$ |

[a] The number of relay nodes is varied for each network size such that all nodes on the average have approximately 15, 30, 80, 160 and 320 local contacts.

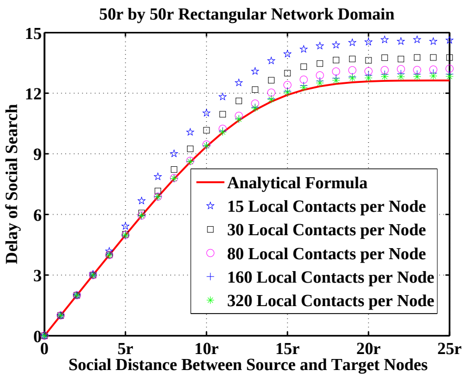


Fig. 13. Comparison of the analytical formula with agent-based simulation results for a $50r$-by-$50r$ rectangular network domain containing various numbers of relay nodes. The number of long-range connections per node is set to 1.

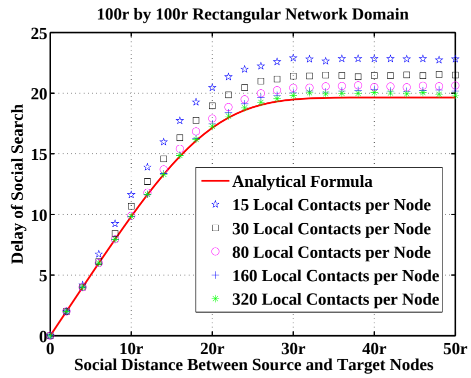


Fig. 14. Comparison of the analytical formula with agent-based simulation results for a $100r$-by-$100r$ rectangular network domain containing various numbers of relay nodes. The number of long-range connections per node is set to 1.

1. These four figures help us to visualize a typical social search process, and thereby further improve our understanding of the dynamics of social search. In these figures, the target node is placed at different corners of the network domain: $(r, 19r)$, $(19r, 19r)$, $(r, r)$ and $(19r, r)$. We see that the micro level message forwarding rule performs well for finding short paths connecting source and target nodes in all four realizations of the social search process. In particular, the message in some cases readily enters a close vicinity of the target node by means of a long-range contact. (See Figs. 10, 11 and 12.) Once the message is in this small vicinity of the target node, it takes several more hops to reach the target node by means of local contacts. For such chains in these figures, the delay of the social search varies between 3 and 5 hops. As these figures show, nodes tend to use their long-range contacts at the initial steps of the social search. However, in some rare cases, it also happens that neither the source node nor the initial message holders have a long-range contact who can get the message to the close vicinity of the target node. In these cases, the message gets stuck around the source node, and can make only small forward progress toward the target node in the initial steps. For such chains, it takes longer to reach the target node. An example realization of such a chain is shown in Fig. 9. In order to understand the net effect, the delay of social search has to be averaged over many realizations of the social search process, which will be the focus of our analysis next.

In Figures 13 - 16, we compare our analytical formula against the results of our agent-based simulations. For a given source-target separation, many different realizations of the social search process are considered, and the delay is averaged over all realizations for different network sizes, various numbers of nodes contained in the network and all source-target separations. (See Table 3 for the summary of simulation parameters.) The results are promising. As the average number of local contacts per node increases, the deviations between analytical and simulation results become negligible. In particular, when the average number of local contacts per node is between Dunbar's number 150 (Dunbar, 1993; Hill and Dunbar, 2003) and Killworth's estimate 290 (McCarty et al., 2001; Bernard et al., 2001) for the average size of personal networks, the social delay

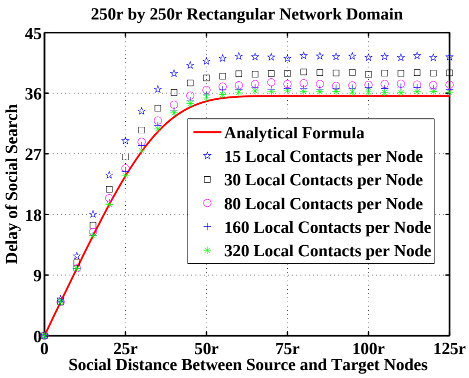


Fig. 15. Comparison of the analytical formula with agent-based simulation results for a $250r$-by-$250r$ rectangular network domain containing various numbers of relay nodes. The number of long-range connections per node is set to 1.

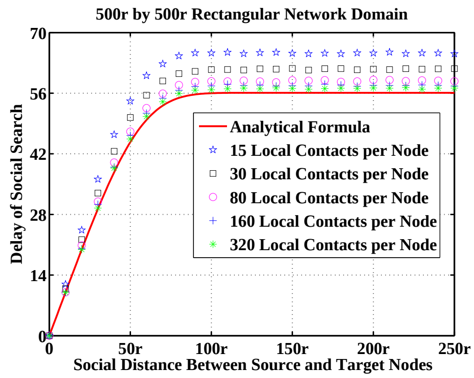


Fig. 16. Comparison of the analytical formula with agent-based simulation results for a $500r$-by-$500r$ rectangular network domain containing various numbers of relay nodes. The number of long-range connections per node is set to 1.

estimated by the analytical formula deviates only from simulation results 2.7%. (See the curves corresponding to 160 local contacts per node in Figs. 13, 14, 15 and 16.) Furthermore, the gains obtained by increasing the number of nodes become very marginal after the number of local contacts per node is larger than 80. (Compare the curves corresponding to 80, 160 and 320 local contacts per node in Figs. 13, 14, 15 and 16.) In summary, our analytical formula derived for dense networks very well approximates the social delay curve when the average number of local contacts per node is around the commonly accepted average sizes for personal networks.

## 4. Discussion of Results

In this section, we discuss the implications of our analytical formulas in detail. Our discussion will, in particular, focus on three important facets of social search processes: (i) roles of short-range and long-range contacts in social search, (ii) the effect of hubs on the delay of social search, and (iii) the effect of social inequality on the delay of social search.

### 4.1. *Roles of short-range and long-range contacts in social search*

In Figures 17 - 22, we plot the delay of social search predicted by our analytical formulas as a function of source-target separation for three different network domains with different sizes and for different numbers of long-range connections per node. In all of these figures, we observe two regimes in the delay as a function of source-target separation. The first regime is the linear growth regime for small source-target separations. When source and target nodes are socially close to each other, long-range contacts cannot help to locate the target node since they overshoot the target node with high probability. As a result, nodes tend to use their local contacts to reach the target node as they come closer to it. This results in the linear increase of the delay of social search for small source-target separations.

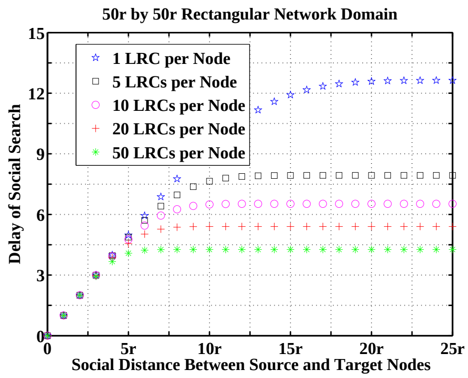


Fig. 17. Delay of social search as a function of source-target separation with various numbers of long-range connections (LRCs) per node for a $50r$-by-$50r$ rectangular network domain.

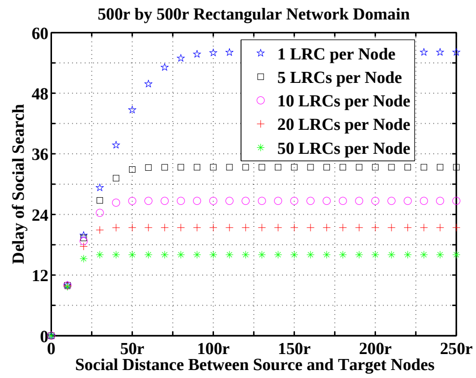


Fig. 18. Delay of social search as a function of source-target separation with various numbers of long-range connections (LRCs) per node for a $500r$-by-$500r$ rectangular network domain.

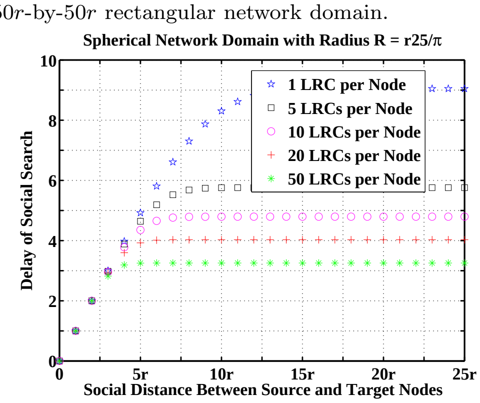


Fig. 19. Delay of social search as a function of source-target separation with various numbers of long-range connections (LRCs) per node for a spherical network domain with radius $R = r\frac{25}{\pi}$.

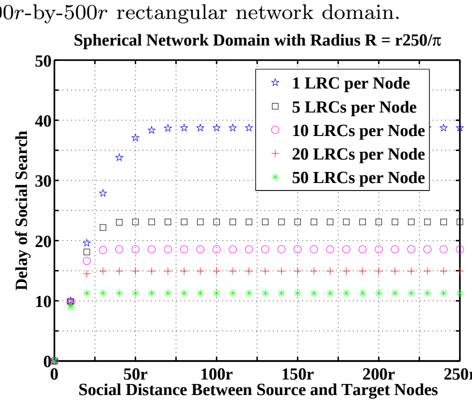


Fig. 20. Delay of social search as a function of source-target separation with various numbers of long-range connections (LRCs) per node for a spherical network domain with radius $R = r\frac{250}{\pi}$.

On the other hand, as the social separation between source and target nodes become larger, the delay saturates to a constant number. Saturation of the delay to a constant arises because long-range connections can better serve as bridges connecting socially distant regions of the network when the source-target separation is large. In this large social separation regime, the message can readily enter to the close vicinity of the target node at the initial steps of the social search by means of long-range contacts. (See also Figs. 9 - 12.) However, once the message is in this small vicinity of the target node, the linear growth regime emerges again, and the message hops through local contacts before it is delivered to the target node.

As an analogy, one can also visualize the roles of long-range and short-range contacts in social search by considering them as two knobs for tuning in a desired radio channel. The first knob (long-range contacts) provide a coarse level granularity to tune in the radio channel. Once the coarse level granularity is accomplished, the perfect tuning is achieved by the second knob (short-range contacts).

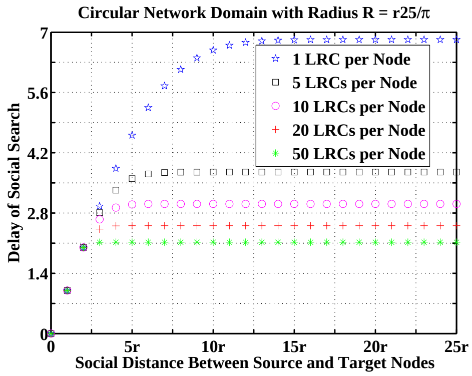


Fig. 21. Delay of social search as a function of source-target separation with various numbers of long-range connections (LRCs) per node for a circular network domain with radius $R = r\frac{25}{\pi}$.

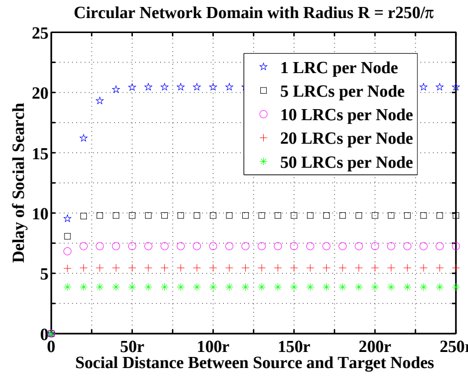


Fig. 22. Delay of social search as a function of source-target separation with various numbers of long-range connections (LRCs) per node for a circular network domain with radius $R = r\frac{250}{\pi}$.

Our observations are mostly consistent with empirical qualitative findings in previous small-world experiments. For example, Travers and Milgram (1969) observed that "Chains which converge on the target principally by using geographic information reach his hometown or surrounding areas readily, but once there often circulate before entering the target's circle of acquaintances." (Dodds et al., 2003) reports that "On the one hand, all targets may in fact be reachable from initial senders in only a few steps, *with surprisingly little variants across targets in different countries and professions*. On the other hand, ..."

However, our theory predicts longer acquaintance chains connecting source-target nodes when compared to empirical quantitative findings in previous small-world experiments. There are two reasons for this. The first and the less important one is that our analytical formulas take a wide range of inputs, such as the number of long-range connections per node and the network size, all of which have salient effects on the delay. It is possible to make the delay of social search around the *magic number* 6 by adjusting them. Therefore, if 6 is the correct number for the lengths of acquaintance chains connecting individuals, our analytical expressions can help to find out the conditions under which it can be achieved in a typical small-world experiment.

The second and the more important reason is the missing data and very low chain completion rates in empirical small-world experiments (Kleinfeld, 2002). Until the very recent publication of Goel et al. (2009), there has not been an advanced statistical technique to account for the missing data in empirical small-world experiments. Similar to our analytical results, their findings also predict longer acquaintance chains connecting source and target nodes whose mean lengths can be as much as 49 (See Table 3 of Goel et al. (2009).) when the empirical data are corrected by considering chains not completed. We also performed a simple robustness check of our analytical results by estimating the $r$ parameter in our model. We obtain surprisingly very close estimates for $r$ when it is calculated by using two different independent estimates of Goel et al. (2009). Goel et al. (2009) estimate the mean delay of the social search in the United States to be 11.8 by using the data of Travers and Milgram (1969). They also estimate the mean delay of the social search worldwide to be 41.5 by using the data of Dodds et al. (2003). In both of these estimates, homogenous chain attrition rate is assumed to account for the missing data. We approximate the United States as a rectangular network domain, and the Earth as a spherical network domain. Since the area of the United States is equal to $9,826,630$ [km$^2$], we set the network side length to $R_{\text{US}} = 3134$ [km]. The radius of the earth is, on the other hand, equal to $R_{\text{Earth}} = 6371$ [km]. Assuming one long-range connection per node and that individuals mainly use geographical information to forward messages, the delay of social search saturates to 41.5 when $R_{\text{Earth}} = \frac{280r_1}{\pi}$, and it saturates to 11.8 when $R_{\text{US}} = 45r_2$. After simple calculations, $r_1 = 71.5$ [km] and $r_2 = 69.6$ [km]. Together with our agent-based simulation results, this means that having approximately 150 friends in a region with radius 70 [km] around individuals and one long-range connection per person justifies mean delay estimates in Goel et al. (2009). When similar calculations are repeated for 5 long-range connections per node, we obtained $R_{\text{Earth}} = \frac{620r_1}{\pi}$ and $R_{\text{US}} = 97r_2$, which means $r_1 = 32.3$ [km] and $r_2 = 32.3$ [km]. When there are 10 long-range connections per node, we have $R_{\text{Earth}} = \frac{875r_1}{\pi}$ and $R_{\text{US}} = 137r_2$, which means $r_1 = 22.9$ [km] and $r_2 = 22.9$ [km]. For 20 long-range connections per node, we have $R_{\text{Earth}} = \frac{1240r_1}{\pi}$ and $R_{\text{US}} = 193r_2$, which means $r_1 = 16.14$ [km] and $r_2 = 16.24$ [km]. And finally for 50 long-range connections per node, we have $R_{\text{Earth}} = \frac{1960r_1}{\pi}$ and $R_{\text{US}} = 305r_2$, which means $r_1 = 10.21$ [km] and $r_2 = 10.28$ [km]. Even though this topic requires further study and more empirical data, these results provide a useful, and somewhat surprising, starting point.

### 4.2. *The effect of hubs on the delay of social search*

Second, we analyze the effects of hubs on the delay of social search. As shown by Barabási (2002), many real-life networks including social networks (e.g., movie actors) exhibit an 80/20 rule: many nodes in these networks have small numbers of links, but there are small numbers of nodes having many links. These highly connected nodes are called hubs of the network, and they play critical roles in achieving network connectivity. Adamic et al. (2001) showed that hubs are also very critical for minimizing the delay of social search. However, their model assumes that nodes use the number of friends of their friends as the next hop message holder selection criterion. On the other hand, Dodds et al. (2003) found in their experiments that "Participants relatively rarely nominated an acquaintance primarily because he or she had many friends (Table 2, "Friends"), and individuals in successful chains were far less likely than those in incomplete chains to send messages to hubs (1.6 versus 8.2%) (table S6)." As a result, highly connected hubs seem to have limited relevance to the global small-world experiment. However, one can still imagine that the existence of highly connected individuals in social networks should help reducing the delay since once a message reaches one of them, even if unintentionally, it is very likely that among many of her contacts, one of the friends of the hub is close to the target node. Therefore, the importance of the role played by hub like individuals in reducing the delay of social search still remains an open research problem if the social search criteria for selecting intermediaries are social dimensions that are different than the number of friends of an individual.

To this end, we generated two types of networks for three different network domains with various sizes. In the first type, the number of long-range connections per node is drawn from a power law distribution so that there are small numbers of hubs with large numbers of long-range ties connecting socially distant regions in the network. The first network type generated in this manner is a variation of the scale-free networks of Barabási (2002). In our model, nodes have both local short-range contacts and random long-range connections, whereas, the concept of local friends is missing in Barabási (2002). We set the decay exponent of the power law to 2 in the first network type; however, the same conclusions can also be made when a similar analysis is repeated for other power law decay exponents. In the second type, the number of long-range connections per node is set to fixed numbers (e.g., 2 and 4). Since the expected number of long-range connections per node is infinite in the first network type, one can naturally expect that the delay of social search for the first network type must be smaller than the delay of social search for the second network type, for any number of long-range connections per node. However, our results shown in Figs. 23 - 28 are again quite surprising, and contrary to this flawed intuition. The delay of social search for power-law networks becomes smaller than that for networks with two long-range connections per node but greater than that for networks with four long-range connections per node. These results, together with findings of Dodds et al. (2003), show that hubs neither have relevance to the social search nor can their existence significantly reduce the length of acquaintance chains connecting individuals. The main reason behind this finding is that it becomes very unlikely to hit a hub when individuals use other social dimensions, such as geography, occupation and education for selecting intermediaries in a typical social search process.

### 4.3. *The effect of social inequality on the delay of social search*

Finally, our previous analysis on the effect of hubs on social search leads us to an equally important and somewhat fuzzier question: what is the effect of social inequality on the delay of social search? We analyze this point by studying simple but intuitive network models in which the number of long-range connections per node exhibit large variances.

In a social network, it is expected that some individuals will invest their time and energy heavily in making friends and maintaining their existing social relationships with others. In turn, these individuals have greater social capital than an ordinary individual has, which is among the potential reasons for the emergence of social inequality in social networks. These socially rich people become structurally very powerful in reaching to many diverse resources located in different parts of the network. As Burt (1992) puts it "You have friends, colleagues, and more general contacts through whom you receive opportunities to use your financial and human capital. I refer to opportunities in a broad sense, but I certainly mean to include the obvious examples of job promotions, participation in significant projects, influential access to important decisions, and so on."

To study this issue of inequality in social capital, we generate networks by keeping the mean number of long-range connections per node constant, but increasing the variance of the distribution from which the number of long-range connections per node is drawn. In particular, we let the number of long-range connections per node be equal to either $k$ with probability $\frac{1}{k}$, or 0 with probability $1 - \frac{1}{k}$. We analyze the delay of social search for different values of $k$: $k = 1, 5, 10$ and 20. The $k = 1$ case corresponds to the situation in which the total existing social capital in the network is distributed equally among all individuals. The large $k$ case corresponds to the situation in which the total existing

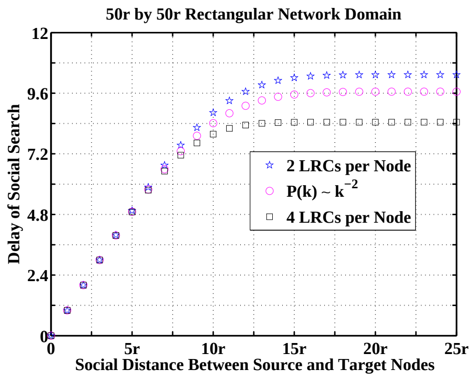


Fig. 23. Delay of social search for scale-free networks and the networks with fixed numbers of long-range connections (LRCs) per node for a $50r$-by-$50r$ rectangular network domain.

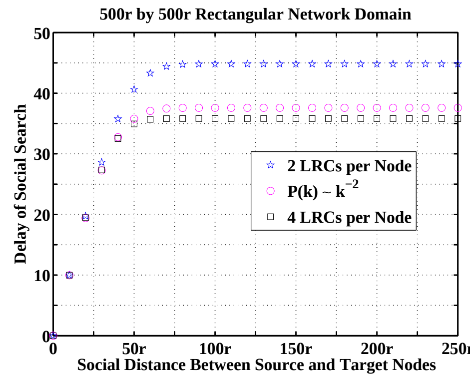


Fig. 24. Delay of social search for scale-free networks and the networks with fixed numbers of long-range connections (LRCs) per node for a $500r$-by-$500r$ rectangular network domain.

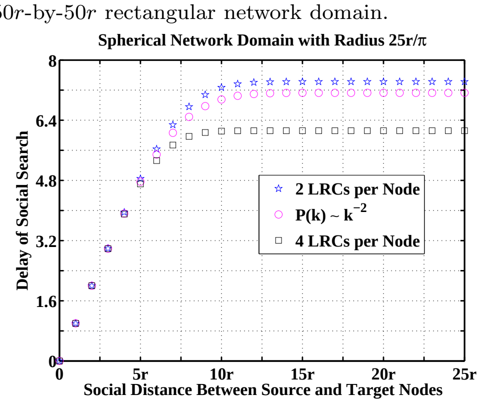


Fig. 25. Delay of social search for scale-free networks and the networks with fixed numbers of long-range connections (LRCs) per node for a spherical network domain with radius $R = r\frac{25}{\pi}$.

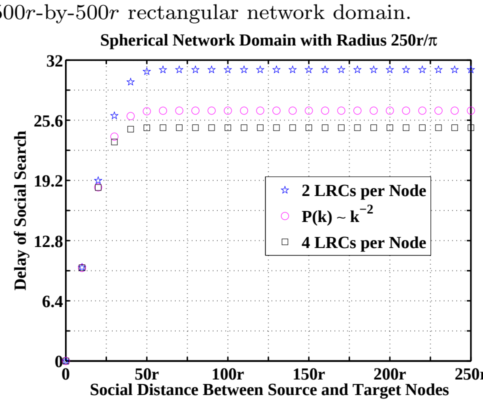


Fig. 26. Delay of social search for scale-free networks and the networks with fixed numbers of long-range connections (LRCs) per node for a spherical network domain with radius $R = r\frac{250}{\pi}$.

social capital in the network is concentrated on a few hub like individuals.

Although our aim in this part of the paper is not to go into the details of the concepts of social capital and social inequality, it is useful to introduce Lorenz curves briefly. One way to visualize and to quantify the inequality among individuals for different values of $k$ is to draw corresponding Lorenz curves. Lorenz curves showing social inequality for different values of $k$ are shown in Fig. 29. As seen in this figure, the Lorenz curve drawn for $k = 1$ is the line of perfect equality. As $k$ increases, the Lorenz curves are shifted to the right, and the inequality among individuals increases, approaching the perfect inequality case. In particular, Gini coefficients are equal to 0, 0.8, 0.9 and 0.95 for $k = 1, 5, 10$ and 20, respectively.

We plot the curves showing the effect of social inequality on the delay of social search in Figs. 30 - 35. As all of these figures clearly show, the delay of social search is minimized when there is perfect equality among individuals in a social network. As the social inequality increases, the delay starts to increase. The adverse effects of social inequality

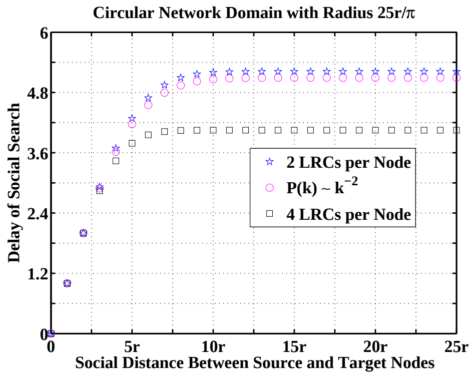


Fig. 27. Delay of social search for scale-free networks and the networks with fixed numbers of long-range connections (LRCs) per node for a circular network domain with radius $R = r\frac{25}{\pi}$.

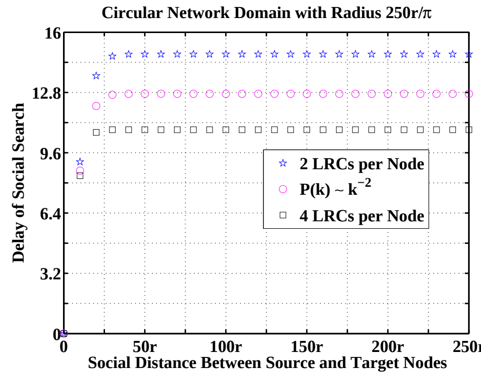


Fig. 28. Delay of social search for scale-free networks and the networks with fixed numbers of long-range connections (LRCs) per node for a circular network domain with radius $R = r\frac{250}{\pi}$.

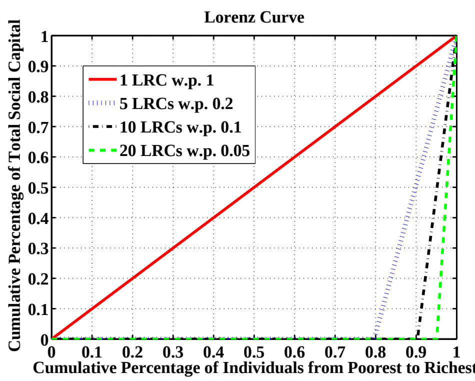


Fig. 29. Lorenz curves for different values of $k$. For each $k$, the number of long-range connections (LRCs) is equal to $k$ with probability (w.p.) $\frac{1}{k}$, or to 0 w.p. $1 - \frac{1}{k}$.

on the delay becomes more prominent for small network sizes. The fundamental reason for this phenomenon is similar to the one explained above while explaining the effect of hubs on the delay of social search. When the total existing social capital in a network is concentrated on fewer and fewer individuals, it becomes less and less likely for a message to hit one of these individuals at each hop during its journey from source to target node. Since it usually takes a small number of intermediaries to connect source and target nodes for small network sizes, the overall probability of hitting one such individual with large social capital remains still small even for the whole journey. As a result, we see an almost linear increase in the delay as a function of source-target separation for networks with large social inequality.

On the other hand, as the network size increases, it becomes more probable to have at least one individual with large social capital as an intermediate message holder. Once the message reaches one such individual, it comes very close to the target node at the next hop through one of her large number of friends. As a result, large network sizes relieve the negative effect of social inequality to some extent, and we observe similar delay curves

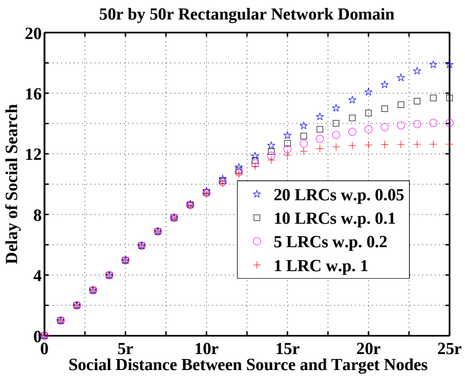


Fig. 30. Effect of social inequality on the delay of social search for a $50r$-by-$50r$ rectangular network domain. Each node has either $k$ long-range connections (LRCs) with probability (w.p.) $\frac{1}{k}$, or 0 LRC w.p. $1 - \frac{1}{k}$.

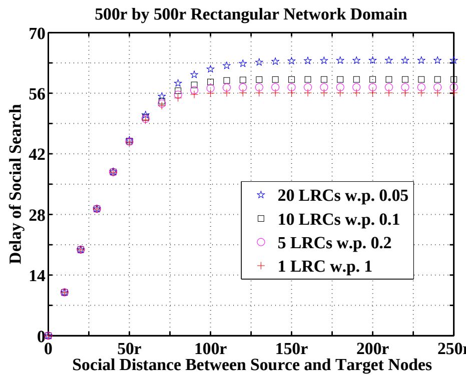


Fig. 31. Effect of social inequality on the delay of social search for a $500r$-by-$500r$ rectangular network domain. Each node has either $k$ long-range connections (LRCs) with probability (w.p.) $\frac{1}{k}$, or 0 LRC w.p. $1 - \frac{1}{k}$.

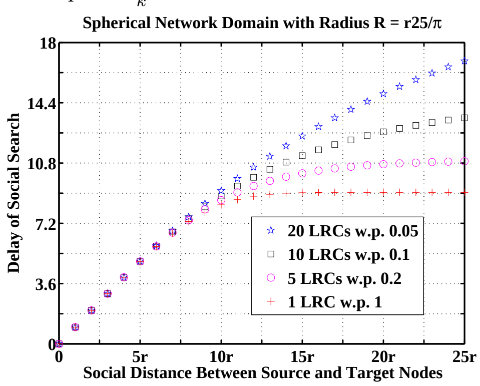


Fig. 32. Effect of social inequality on the delay of social search for a spherical network domain with radius $R = r\frac{25}{\pi}$. Each node has either $k$ long-range connections (LRCs) with probability (w.p.) $\frac{1}{k}$, or 0 LRC w.p. $1 - \frac{1}{k}$.

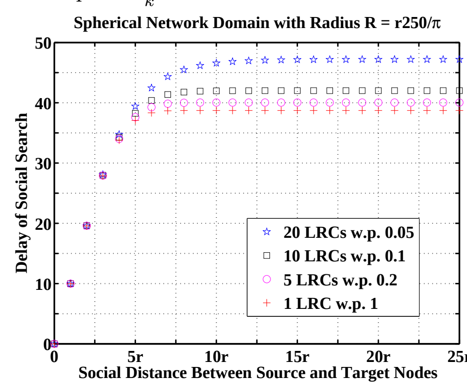


Fig. 33. Effect of social inequality on the delay of social search for a spherical network domain with radius $R = r\frac{250}{\pi}$. Each node has either $k$ long-range connections (LRCs) with probability (w.p.) $\frac{1}{k}$, or 0 LRC w.p. $1 - \frac{1}{k}$.

as a function of source-target separation for different $k$ values.

## 5. Conclusions

In this paper, we have studied the length of acquaintance chains needed to connect any two individuals in a social network (i.e., the delay of social search) as a function of the social separation between them. We have modeled social networks as small-world random geometric graphs, and have obtained closed form expressions for the delay of social search. In our network model, each individual has both local short-range contacts and random long-range contacts. Local contacts are formed according to a threshold rule in which individuals $A$ and $B$ are declared to be local contacts of each other if

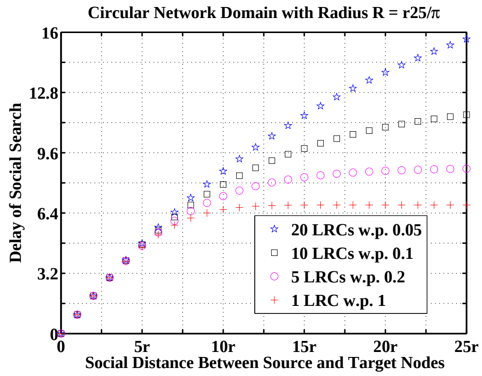


Fig. 34. Effect of social inequality on the delay of social search for a circular network domain with radius $R = r\frac{25}{\pi}$. Each node has either $k$ long-range connections (LRCs) with probability (w.p.) $\frac{1}{k}$, or 0 LRC w.p. $1 - \frac{1}{k}$.

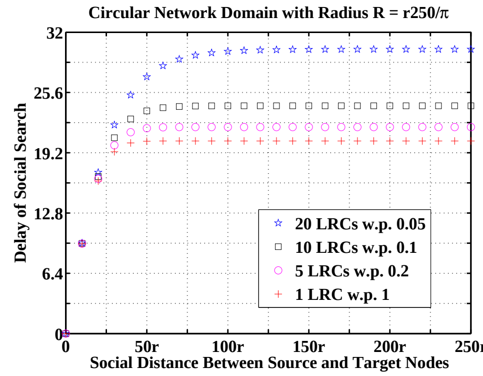


Fig. 35. Effect of social inequality on the delay of social search for a circular network domain with radius $R = r\frac{250}{\pi}$. Each node has either $k$ long-range connections (LRCs) with probability (w.p.) $\frac{1}{k}$, or 0 LRC w.p. $1 - \frac{1}{k}$.

their social separation is small enough. Long-range contacts of an individual are formed uniformly randomly over all other individuals who are not her local contact. These long-range connections can be regarded as bridges connecting socially distant regions in a network. Our model allows the number of long-range connections per node to be drawn from any given distribution including scale-free distributions. Therefore, our model can be considered to capture important properties of both small-world networks of Watts (2003) and scale-free networks of Barabási (2002). We have constructed our networks on three different network domains (rectangular, spherical and circular) to examine robustness of our results against the effect of network geometry on the delay. We have arrived at similar conclusions about the delay of social search for all three network domains. We also have compared our findings with empirical small-world experiments, and have observed both qualitative and quantitative matches between our results and those of previous small-world experiments.

Our main contribution is the derivation of an analytical closed form solution for the delay of social search in the dense network limit. We also have performed agent-based simulations on large computer clusters to verify our analytical expressions. Our simulation results indicate that when the number of local contacts per node is in between Dunbar's number 150 (Dunbar, 1993; Hill and Dunbar, 2003) and Killworth's estimate 290 (McCarty et al., 2001; Bernard et al., 2001) for the average size of personal networks, the social delay estimated by the analytical formula deviates only at most 2.7% from simulation results. However, the computational gains earned by analytical expressions are tremendous. Some simulations can take up to one week to complete on clusters consisting of 200 CPUs, whereas the analytical formulas takes only several seconds to compute on an ordinary personal computer.

Our analytical formulas help us to unravel some hidden dynamics of social search. To start with, we have observed that the delay of social search first increases linearly for small values of source-target social separation, and then it saturates to a constant value for large values of source-target social separation. This happens because nodes tend to use long-range contacts (short-range contacts) when the message is far away from (close to) the target node. These results are consistent with empirical findings of Travers and Milgram (1969) and Dodds et al. (2003).

Secondly, we have studied the effect of hubs and social inequality on the delay of social search. Contrary to our intuition and some previous results (Adamic et al., 2001), our results show that hubs have limited impact in reducing the delay of social search. The main reason for this result is that individuals relatively rarely nominate their friends as next hop neighbors because they have many friends in a typical social search experiment Dodds et al. (2003). Once this happens, gains achieved by hubs become very limited to make the *world small* in small-world experiments. Finally, our investigation regarding the effect of social inequality on the delay of social search indicates that social inequality adversely affects the delay of social search. As the social inequality increases, the delay increases. Therefore, this result implies that when the existing social capital in a network is distributed equally among all individuals, the delay of social search is minimized.

We believe insights that can be gained by using our analytical formulas in social search processes may have other applications in addition to those indicated by two applications related to the effects of hubs and social inequality on the delay of social search. We hope that this paper serves as a starting point towards this direction, and is useful to social scientists working on understanding hidden dynamics of social search processes.

## Appendix A.

We first derive the formula for $P\left((m-1)r \leq \rho(M_1, X_t) < mr\right)$ and $1 \leq m \leq k-1$. The message after the first hop will be located at an intermediate message holder whose social separation from the target node is smaller than $mr$ but greater than $(m-1)r$ if and only if none of source's long-range contacts is socially closer than $(m-1)r$ to the target node, and the source node has at least one long-range contact lying in between the disks centered at $X_t$ with radii $(m-1)r$ and $mr$ (i.e., the gray area in Fig. A.1).

A given long-range contact of the source node lies in between the disks centered at $X_t$ with radii $(m-1)r$ and $mr$ with probability

$$\alpha_m = \frac{\pi r^2(2m-1)}{R^2 - \pi r^2}.$$

A given long-range contact of the source node lies outside of the disk centered at $X_t$ with radius $mr$ with probability

$$\beta_{m+1} = 1 - \frac{\pi r^2 m^2}{R^2 - \pi r^2}.$$

We then have

$$\begin{aligned} P\left((m-1)r \leq \rho(M_1, X_t) < mr\right) &= \sum_{n=1}^{\infty}\sum_{i=1}^{n} Q(n)\binom{n}{i}\alpha_m^i \beta_{m+1}^{n-i} \\ &= \sum_{n=1}^{\infty} Q(n)\beta_{m+1}^n \left(\left(1 + \frac{\alpha_m}{\beta_{m+1}}\right)^n - 1\right) \\ &= \sum_{n=1}^{\infty} Q(n)\left(\beta_{m+1} + \alpha_m\right)^n - \sum_{n=1}^{\infty} Q(n)\beta_{m+1}^n \end{aligned}$$

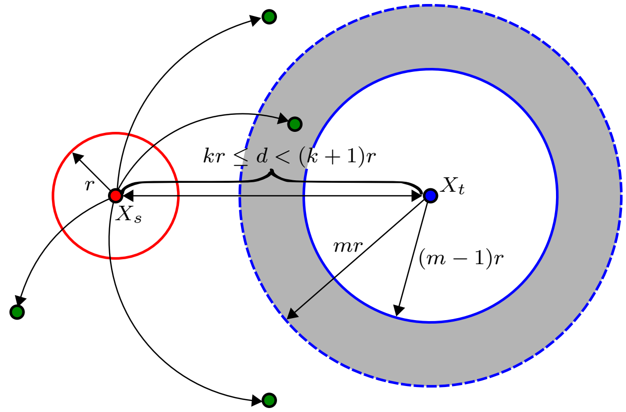


Fig. A.1. Graphical representation of the derivation of $P\left((m-1)r \le \rho(M_1, X_t) < mr\right)$.

$$= \varphi\left(\beta_{m+1} + \alpha_m\right) - \varphi\left(\beta_{m+1}\right).$$

Note also that

$$\begin{aligned}\beta_{m+1} + \alpha_m &= 1 - \frac{\pi r^2 m^2}{R^2 - \pi r^2} + \frac{\pi r^2(2m-1)}{R^2 - \pi r^2} \\ &= 1 - \frac{\pi r^2 (m-1)^2}{R^2 - \pi r^2} \\ &= \beta_m.\end{aligned}$$

Finally, we have $P\left((m-1)r \le \rho(M_1, X_t) < mr\right) = \varphi(\beta_m) - \varphi(\beta_{m+1})$.

To obtain $P\left((k-1)r \le \rho(M_1, X_t) < kr\right)$, we first realize that the message after the first hop is located at an intermediate message holder whose social separation from the target node is in between $(k-1)r$ and $kr$ if and only if all of the long-range contacts of the source node have social separation from the target node larger than $(k-1)r$. By repeating steps similar to those above, we obtain

$$P\left((k-1)r \le \rho(M_1, X_t) < kr\right) = \varphi(\beta_k).$$

References

Adamic, L. A., Lukose, R. M., Puniyani. A. R., Huberman, B. A., 2001. Search in power-law networks. Physical Review E 64, 046135.

Axelrod, R., 1997. The dissemination of culture: a model with local convergence and global polarization. The Journal of Conflict Resolution 41, 203-226.

Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., Parker, M., 2002. Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. Proc. Natl. Acad. Sci. 99, 7275-7279.

Barabási, A.-L., 2002. Linked: The New Science of Networks. Perseus, Cambridge, MA.

Bassett, D. S., Bullmore, E., 2006. Small-world brain networks. Neuroscientist 12, pp. 512-523.

Bearman, P. S., Moody, J., Stovel, K., 2004. Chains of affection: The structure of adolescent romantic and social networks. American Journal of Sociology 110, 44-01.

Bernard, H. R., Killworth, P. D., Johnsen, E. C., Shelley, G. A., McCarty, C., 2001. Estimating the ripple effect of a disaster. Connections 24, 18-22.

Bollobás, B., 2001. Random Graphs. Cambridge University Press, Cambridge.

Burt, R. S., 1992. Structural Holes. The Social Structure of Competition. Harvard University Press, Cambridge, MA.

Centola, D., Willer, R., Macy, M., 2005. The emperor's dilemma: A computational model of self-enforcing norms. American Journal of Sociology 110, 1009-1040.

Centola, D., Macy, M., 2007. Complex contagions and the weakness of long ties. American Journal of Sociology 113, 702-734.

Cohen, J. E., Briand, F., Newman, C. M., 1990. Community Food Webs: Data and Theory. Springer, Berlin.

Cohen, L., Frazzini, A., Malloy, C., 2007. The small world of investing: board connections and mutual fund returns. European Finance Association Meeting. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=967544.

Coleman, J., Katz, E., Menzel, H., 1957. The diffusion of an innovation among physicians. Sociometry 20, 253-270.

Davis, J. A., Leinhardt, S., 1971. The Structure of Positive Interpersonal Relations in Small Groups. Sociological Theories in Progress, vol. 2, Houghton-Mifflin, Boston.

Dodds, P. S., Muhamad, R., Watts, D.J., 2003. An experimental study of search in global social networks. Science 301, 827-829.

Donovan, P., 2007. How idle is idle talk? One hundred years of rumor research. Diogenes 54, 59-83.

Dunbar, R. I. M., 1993. Coevolution of neocortical size, group size and language in humans. Behavioral and Brain Sciences 16, 681-735.

Epstein, J. M., 1999. Agent-based computational models and generative social science. Complexity 4, 41-60.

Erickson, B. H., Kringas, P., 1975. Small world of politics, or seeking elites from bottom up. Canadian Research Sociology Quarterly 12, 585-593.

Franceschetti, M., Meester, R., 2006. Navigation in small-world networks, a scale-free continuum model. Journal of Applied Probability 43, pp. 1173-1180.

Goel, S., Muhamad, R., Watts, D. J., 2009. Social search in "small-world" experiments. 18th International World Wide Web Conference, 701-710.

Granovetter, M. S., 1973. The strength of weak ties. American Journal of Sociology 78, 1360-1380.

Granovetter, M. S., 1983. The strength of weak ties: A network theory revisited. Sociological Theory 1, 201-233.

Guiot, J. M., 1976. A modification of Milgram's small world method. European Journal of Social Psychology 6, 503-507.

Hill, R. A., Dunbar, R. I. M., 2003. Social network size in humans. Human Nature 14, 53-72.

Homans, G. A., 1950. The Human Group. Brace & World, New York.

Kleinberg, J. M., 2000a. Navigation in a small world. Nature 406, 845.

Kleinberg, J. M., 2000b. The small-world phenomenon: An algorithmic perspective. Proc. 32nd ACM Symposium on Theory of Computing.

Kleinfeld, J., 2002. The small world problem. Society 39, 61-66

Klovdahl, A. S., 1985. Social networks and the spread of infectious diseases: The AIDS example. Social Science and Medicine 21, 1203-1216.

Kochen, M., 1985. The structure of acquaintance nets and rates of social development. Social Networks 7, 323-339.

Korte, C., Milgram, S., 1970. Acquaintance linking between white and Negro populations: applications of the small world problem. Journal of Personality and Social Psychology 15, 101-118.

Lee, N. H., 1969. The Search for an Abortionist. University of Chicago Press, Chicago.

Lin, N., Dayton, P. W., Greenwald, P., 1977. The urban communication network and social stratification: A "small world" experiment. In: Ruben, B. D. (Ed.). Communication Yearbook I. Transaction, New Brunswick, NJ, pp. 107-119.

Macy, M. W., Willer, R., 2002. From factors to actors: computational sociology and agent-based modeling. Annual Review of Sociology 28, 143-166.

Malaquias, J. L., Rosa, A. C., Correia, C. M. B. A., 2006. A Small-world model of the human mind. Proceedings of the 2006 ACM Symposium on Applied Computing, pp. 15-22.

McCarty, C., Killworth, P. D., Bernard, H. R., Johnsen, E., and Shelley, G. A., 2001. Comparing two methods for estimating network size. Human Organization 60, 28-39.

Milgram, S., 1967. The small world problem. Psychology Today 1, 61-67.

Newcomb, T. M., 1961. The Acquaintance Process. Rinehart & Winston, New York.

Newman, M. E. J., Moore, C., Watts, D. J., 2000. Mean-field solution of the small-world network model, Phys. Rev. Lett. 84, 3201-3204.

Newman, M. E. J., 2001. The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. 98, 404-409.

Newman, M. E. J., 2003. Ego-centered networks and the ripple effect. Social Networks 25, 83-95.

Nowak, M. A., Sigmund, K., 1992. Tit for tat in heterogenous populations. Nature 355, 250-253.

Penrose, M., 2003. Random Geometric Graphs. Oxford University Press, New York.

Pool, I. D. S., Kochen, M., 1978. Contacts and influence. Social Networks, 1, 5-51.

Rapoport, A., 1953. Spread of information through a population with socio-structural bias I: Assumption of transivity. Bulletin of Mathematical Biophysics 15, pp. 523-533.

Rapoport, A., 1953. Spread of information through a population with socio-structural bias II: Various models with partial transitivity. Bulletin of Mathematical Biophysics 15, pp. 535-546.

Rapoport, A., 1954. Spread of information through a population with socio-structural bias III: Suggested experimental procedures. Bulletin of Mathematical Biophysics 16, pp. 75-81.

Rudin, W., 1987. Real and Complex Analysis. McGraw-Hill, New York.

Schelling, T. C., 1971. Dynamic models of segregation. Journal of Mathematical Sociology 1, 143-186.

Schenettler, S., A structured overview of 50 years of small-world research. Social Networks (2009).

Schenettler, S., A small world on feet of clay? A comparison of empirical small-world studies against best-practice criteria. Social Networks (2009).

Shotland, R. L., 1976. University Communication Networks: The Small World Method. Wiley, New York.

Stevenson, W. B., Davidson, B., Manev, I., Walsh, K., 1997. The small world of the university: a classroom exercise in the study of networks. Connections 20, 23-33.

Strogatz, S. H., 2001. Exploring complex networks. Nature 410, 268-276.

Travers, J., Milgram, S., 1969. An experimental study of the small world problem. Sociometry 32, 425-443.

Watts, D. J., Strogatz, S. H., 1998. Collective dynamics of small-world networks. Nature 393, 440-442.

Watts, D. J., 2003. Six Degrees: The Science of a Connected Age. Norton & Company, New York, USA.

Watts, D. J., 2004. The “new” science of networks. Annual Review of Sociology 30, 243-270.

Weimann, G., 1983. The not-so-small world - ethnicity and acquaintance networks in Israel. Social Networks 5, 289-302.

Williams, R. J., Martinez, N. D., 2000. Simple rules yield complex food webs. Nature 404, 180-183.