Delay of Social Search on Small-world Graphs [★]

Hazer Inaltekin [a,*] Mung Chiang [b] H. Vincent Poor [b]

[a] *The University of Melbourne, Melbourne, VIC 3010, Australia*
[b] *Princeton University, Princeton, NJ 08544, United States*

Abstract

This paper introduces an analytical framework for two small-world network models, and studies the delay of targeted social search by considering messages traveling between source and target individuals in these networks. In particular, by considering graphs constructed on different network domains, such as rectangular, circular and spherical network domains, analytical solutions for the average social search delay and the delay distribution are 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 are first verified by agent-based simulations, and then compared with empirical observations in small-world experiments. Our formulas indicate 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. 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 qualitatively consistent with experimental observations made by Travers and Milgram in 1969 as well as by others. Moreover, analytical distributions predicted by our models for the delay of social search are also compared with corresponding empirical distributions, and good statistical matches between them are observed. Other somewhat surprising conclusions of the paper are that hubs have limited effect in reducing the delay of social search and variation in node degree distribution adversely affects this delay.

Key words: Small-world phenomenon, Small-world networks, Social networks, Social search, Delay of social search, Scale-free networks

1. Introduction

1.1. *Motivation*

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 (Klovdahl, 1985; Bearman et al., 2004), 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 structural properties of social networks, such as geodesics, to shed light on this diverse set of social processes.

Following the categorization introduced by Goel et al. (2009), such a wide spectrum of small-world research can be divided into the two related research frameworks of topological and algorithmic small-world research. The key questions addressed in the topological small-world research framework are related to global graph-theoretic properties, such as the average path length of the shortest path connecting any two randomly selected individuals (Watts and Strogatz, 1998), or the emergence of connected unique giant components of individuals (Callaway et al., 2000; Newman, 2003a; Leskovec and Horvitz, 2008). Discovery of such topological properties is particularly important for understanding of diffusion and contagion processes running on social networks, such as the spread of infectious diseases (Klovdahl, 1985; Bearman et al., 2004), and for assessing robustness and fragility of general random networks. On the other hand, algorithmic small-world research deals with the questions of how, and if, ordinary individuals are able to navigate the short chains existing in social networks by using their local information only. This line of research is particularly important for understanding of targeted search processes running on social networks such as a search for a new job (Granovetter, 1973, 1983) or for a service provider (Lee, 1969). The main focus of this paper fall within the algorithmic small-world research framework, particularly on the distribution and the average length of the search chain in targeted search processes running on social networks.[1]

Milgram and his colleagues, motivated by the topological small-world research initiated by Pool and Kochen around 1958 and published in 1978 (Pool and Kochen, 1978), 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). A similar experiment was repeated by Dodds et al. (2003) at the global scale

---

[1] Equivalently, if we follow the terminology introduced by Schenettler (2009a), we can also state that the main focus and contributions of this paper are on the structure and process related dimensions of small-world research, crosscutting algorithmic small-world research framework in the structural dimension and targeted search processes in the process dimension.

by means of e-mail messages, and they obtained results similar to those of Milgram's experiments by taking the chain attrition rates into account. In spite of some drawbacks of these empirical studies and some more recent findings indicating that the average length of referral chains is higher if the missing data in experimental studies is accounted for correctly (Goel et al., 2009), it is commonly agreed that people around the world are connected to each other through a small number of acquaintances (Schenettler, 2009b).

As a result of considerable research spanning more than 40 years, particularly during the last decade (Chaintreau et al., 2008; Clauset and Moore, 2003; Duchon et al., 2006; Nowell et al., 2005; Watts et al., 2002), significant progress has been made towards the understanding of dynamics of small-world social networks. However, apart from some bounds (Kleinberg, 2000a,b; Franceschetti and Meester, 2006), analytic expressions quantifying the dynamics of the targeted search in such networks are still unknown. In this paper, we consider the problem of targeted search in social networks analytically, seeking a mathematical confirmation of the above-noted experimental phenomena. We obtain analytical solutions for the average delay required for social search, as well as the probability distribution of the delay of social search. We will refer to this quantity as *the delay of social search*, or, alternatively, as *the length of social search chains connecting two individuals*. We next summarize our main contributions.

1.2. *Summary of Key Findings*

The main contributions of the paper are the introduction of an analytical framework for two small-world network models, and the derivation of the analytical solutions for the delay of social search and its probability distribution as a function of the source-target separation.

From a process point of view, our results indicate 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). Kleinberg drew similar conclusions based on the bounds in his novel work (Kleinberg, 2000a,b). Our results put all of these findings in a quantified framework by allowing us to visualize social delay versus social separation curves for a wide spectrum of network topologies and parameters.

From an algorithmic point of view, the results reveal the dynamics of the social search, and pinpoints the range of source-target separations for which long-range connections positively contribute to such search. They also quantify the effect of the number of long-range connections on the delay required for social search. In particular, they show that the delay required benefits from the network homogeneity parametrized by the variance of the node degree distribution for the number of long-range connections per node. Unlike 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 dimensions, such as geographic location, occupation, education, etc., rather than the number of friends of an individual.

In Model I, we introduce a generalization of the original small-world network model

(Watts and Strogatz, 1998) by means of random geometric graphs (Penrose, 2003) constructed on different continuum network domains. We call the new model the *octopus model*. The octopus model captures the order-disorder properties of the Watts and Strogatz (1998) model. It parametrizes the generation of complex networks as a function of the distribution of the number of long-range connections per individual in the social network. The last feature enables us to produce a wide spectrum of complex networks including scale-free small-world networks by changing the distribution of the number of long-range connections. (See Section 3 for details.)

In Model II, we present a generalization of Model I. We call it the *generalized octopus model*. Long-range connections exist at all scales of social distance in the generalized octopus model, which leads to a delay of social search growing logarithmically with network size. Networks in the second network model are also constructed on continuum network domains, and parametrized by the distribution of the number of long-range connections. (See Section 5 for details.)

For both models, we provide iterative analytical solutions for the (average) delay of social search and its probability distribution. (See Sections 3 and 5 for details.) These solutions are given as a function of source-target separation in a typical small-world experiment and the distribution of the number of long-range connections per individual. The iterative solution for the delay of social search for the octopus model is further reduced to a closed form expression. (See Eq. (1) and Table 2.) All solutions can be computed on a standard personal computer very quickly, with a computational gain of six orders of magnitude when compared with simulations. We verify the delay equations with agent-based simulations (see Section 4 for details), and compare the analytical distributions with those obtained in empirical studies. The standard chi-square test produced statistically significant results for the similarity between analytical and empirical distributions. (See Section 6 for details.)

1.3. *Scope and Limitations*

Our small-world models are specifically proposed for social network modeling, and for the analysis of targeted search processes in social networks. However, 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). Small-world networks have also found some applications in the design of communication networks and protocols (Manku et al., 2003; Nguyen and Martel, 2006). Therefore, the analysis presented in this paper may have broader implications to networks in which search agents exist and similar targeted search processes emerge.

However, it will usually require a substantive amount of research to interpret our model parameters for each particular research problem of interest, and to understand process dynamics in each field. For example, search agents may not exist in all applications, and diffusion processes may be more appropriate to model process dynamics in cases such as spread of viruses and diseases. As Schenettler (2009a) puts it "process model results were often far more partial and limited, for reasons related to the fact that scientific elucidation of 'processes' is widely context specific."

We will illustrate one particular interpretation of our models in the context of wireless

networks. A typical communication device in a wireless technological network usually has an upper bound on the power to be used for signal transmission. This limitation results in an electromagnetic signal footprint that can only be heard within some local neighborhoods of wireless communication devices (Gupta and Kumar, 2000). On the other hand, it has been recently discovered that some occasional random long-range connections can naturally emerge in wireless technological networks, thanks to some physical properties of wireless channels such as fading and log-normal shadowing (Gowaikar and Hassibi, 2006; Hekmat and Mieghem, 2004). As a result, wireless technological networks exhibit small-world properties, and the targeted social search problem in social networks becomes the directed geographical routing problem in wireless technological networks, in which the data packets hop from one device to another one until they are delivered to their final destinations. The distance is now the physical distance, rather than the abstract social distance, measured by means of a Euclidean distance metric.

1.4. *Definitions and Paper Organization*

Before concluding this introductory section of the paper, we briefly provide some clarifying notes on two concepts frequently used throughout the paper, and then present the organization for the rest of the paper. Our first note is on the concept of *social distance*. There are two different ways of measuring social distance between individuals in social network analysis (Watts, 2003). The first way is to count the number of hops in the shortest path, or in the locally discovered short path, connecting individuals over the network. The second one is the distance between two individuals measured by their similarity along multiple social dimensions, such as geography, profession, race and education. When we use the term social distance in this paper, it is this latter notion of social distance to which we refer. In particular, we assume that social similarity is measured on the network domains, such as rectangular, spherical and circular network domains, used to symbolize the social space.

More generally, our analysis can be repeated for any given metric, and the corresponding metric space, possibly higher dimensional and much different topologically from the network domains above, to measure social distances between individuals. We do not pursue this direction here due to unnecessary abstract mathematical complexities obscuring the simplicity of our final expressions. To form one such metric, for example, we can take a whole range of social categories, and count the overlap in categories between individuals to obtain a measure of social distance between them. For some of these social dimensions such as religion, profession and education, such social categories are almost reasonably well defined. For some other social dimensions such as age and geography in which social categories are not well defined, we can use some intuitive notions of similarity between individuals, and partition the entire population into smaller groups, such as age groups and regional groups, to form social categories.

Categorization as in this example is only one way of measuring social distance between individuals, and many other alternative and possibly sociologically more plausible ways of measuring social distance can be proposed. In particular, determination of the social distance measured along multiple social dimensions is an active area of research spanning several decades (Barnett, 1989; Davidson, 1983; Watts, 1999; Watts et al., 2002).

Our second note is on the definition of the *delay of social search*. When we use the term

delay of social search, we mean the number of steps needed to connect two individuals by means of a targeted search process. We will refer to this delay either as an average quantity, averaged over many realizations of the targeted search process, or as a random quantity realized in a particular realization of the targeted search process. It will be clear from the context whether we refer to an average quantity or to a random quantity when we use the term delay of social search. Finally, within the context of this paper, "delay" has no relation with the timing of social interactions and its effects on diffusion processes (Moody, 2002).

The rest of the paper is organized as follows. In Section 2, we provide a comparison between this work and other key related works. In Section 3, we explain the details of Model I, and derive the analytical solutions for the delay of social search and its probability distribution in Model I. Section 4 contains our agent-based simulation results verifying the derived equations for the delay of social search. In Section 5, we explain the details of Model II, and derive the analytical solutions for the delay of social search and its probability distribution in Model II. Section 6 presents a discussion on the implications of the derived delay equations, provides a comparison between analytical solutions and empirical results, and illustrates some applications of the derived delay equations. Section 7 concludes the paper.

2. Comparison with Key Related Works

Our results in this paper should be compared and contrasted with the results reported in closely related works by Kleinberg (2000a,b) and Watts et al. (2002). While seeking a mathematical explanation for local discovery of short paths, Kleinberg (2000a,b) used a rectangular grid with random long-range connections as a proxy for the social space. Unlike the original small-world network model (Watts and Strogatz, 1998), the probability of a long-range connection between two individuals in Kleinberg's model decreases with the social distance between them as measured on the rectangular grid. He defines a social space constructed in this way to be an algorithmically small social network if the delay of social search grows no faster than poly-logarithmically with the size of the network, and shows that searchable social networks exist when the decay rate of the probabilities associated with long-range connections exactly match the dimension of the grid. This requirement of "exact match" is quite brittle. Furthermore, Travers and Milgram (1969) only found that chain lengths connecting two individuals are small (Watts et al., 2002). There is no empirical evidence showing that lengths of small-world search chains grow logarithmically with the network size, even though such evidence exists for the lengths of shortest paths (Newman, 2003a; Schenettler, 2009a; Watts and Strogatz, 1998).

In this paper, we provide an alternative explanation for why we observe short chain lengths in empirical small-world experiments, by showing that the delay of social search, even with long-range connections formed over the social space uniformly at random as in the original small-world model (Watts and Strogatz, 1998), saturates to a constant as the social distance between source and target individuals increases. This constant depends on our model parameters, and we further show that it can be made small enough, as observed in empirical studies, for reasonable choices of the number of long-range connections per individuals. Our analytical expressions characterizing the chain length distribution produce compatible results when compared with those of Travers and Milgram (1969).

In Model II, we consider the targeted search process as progressing on a nested sequence of small-worlds with uniform random long-range connections.[2] This model exhibits logarithmically growing social search delays. These results imply that the original small-world model (Watts and Strogatz, 1998), in which long-range connections are formed uniformly at random without depending on any particular notion of social distance, is in fact adequate to explain logarithmically growing social search delays.

Starting from a sociologically more realistic premise than the low dimensional network model proposed by Kleinberg (2000a,b), Watts et al. (2002) showed that the space of searchable networks is, in fact, larger than what is predicted by Kleinberg's model. For their analysis, they use an ultra-metric that measures the similarity between individuals along multiple social dimensions such as geography, race, profession, religion and education. They claim that it is this multi-dimensioned nature of social identity that makes social networks searchable (Watts, 2003).

Even though the analytic approaches are different, our analysis to derive the expected delay of social search is not much different than the analysis of Watts et al. (2002) conceptually. By drawing hypothetical social circles centered around individuals, we partition the whole social space into smaller social groups whose sizes are determined by our model parameters. Going one step up in the hierarchy in the hierarchical model of Watts et al. (2002) corresponds to going to an outer circle in our model. Similar to Watts et al. (2002), we can also create multiple independent social spaces, and allow individuals to hop not only in one social space but also from one social space to another one while assessing the similarity between their contacts and the target individual.

One may also argue that two individuals are more likely to be friends with each other if they are closer in multiple dimensions simultaneously. For example, two individuals working in the same company and living in the same neighborhood are more likely to interact with each other than any two random individuals working in the same company. Therefore, multiple social dimensions do not work in isolation to determine the social distance between individuals as in Watts et al. (2002), but in fact work together to reinforce or weaken the social closeness between individuals. Context-dependent determination of the social distance between individuals is beyond the scope of this paper. As a result, we work with a more simplified definition of social distance when compared with the analytical approach taken in Watts et al. (2002).

---

[2] For example, if we consider geographic location as the sole next step message holder selection criterion for illustrative purposes, the task of the source individual in a typical targeted search process is to search her small-world at the global scale to advance the message to one of her acquaintances on the same continent as the target if source and target individuals are on different continents. Therefore, the targeted search process at the first step can be deemed to be a search problem at the global scale if source and target individuals are on different continents. Then, it becomes a search problem at the continental scale if the next step message holder and the target are in the same continent but in different countries, and then becomes a search problem at the country scale if the next step message holder and the target are in the same country but in different cities, and so on. The targeted search process continues in this manner on a nested sequence of small-worlds spanning various geographical distances until the message is delivered to the target individual.

3. Model I: Social Search and the Derivation of the Delay Formula

We will use two different but closely related network models, both of which lead to statistically convincing results when compared with those obtained in empirical studies, to explain why we observe short chains of social search in a typical small-world experiment. Our second model is an extension of the first model in which we consider a nested sequence of small-worlds modeled by the first model. Even though there does not exist a commonly accepted notion of smallness of the delay of social search from a sociological point of view (Newman, 2003a; Schenettler, 2009a), this notion, mainly motivated by Kleinberg's findings (Kleinberg, 2000a,b), is usually taken to be any delay that grows logarithmically or poly-logarithmically with the size of the network. Social delay formulas obtained by using our second model also capture such a smallness notion. This section and the next one will focus on the details of the first model, derivation of the delay formula and the distribution of this delay, and agent based simulations validating the derived formulas.

3.1. Details of Model I

To model the connections among people, we use small-world networks 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 statistically replace strong and weak ties, respectively. More precisely, our strong ties tend to be overlapping with a very high probability, and unable to open up channels into more distant social circles, and therefore can be considered as short for searching resources in social networks. However, some individuals in some rare cases may also have strong ties that are not overlapping, and in these cases, strong ties play a similar role with regard to information transmission as do weak ties. For example, the search for an abortionist described by Lee (1969) can be considered as an example for a search process that stretches over long social distances but only through strong (trustful) ties. Even though 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, the analysis presented below holds under both interpretations by considering the above statistical distinction between strong/short-range and weak/long-range connections.

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 detail (e.g., Centola and Macy (2007); Newman et al. (2000)). Different meanings can also be attributed to these 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 metric measure spaces (Rudin, 1987), however we will not pursue this direction in this paper since it introduces non-essential 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 rather than physical distance.

We assume that individuals are randomly distributed (from uniform distribution) over the network domain, and any two individuals A and B are short-range contacts of each other if their social separation is smaller than r. A further generalization of our models in which the social distance is continuous and not classified into short-range and long-range connections is possible if the r parameter of our models is drawn independently from a continuous distribution for each individual. With this modification to the models, a continuous range of social distances in which different individuals attribute different meanings to "who is close" and "who is far away" will emerge.

The location of individual i in the social space will be represented by X_i. Each individual also selects a random number, N, of long-range contacts, uniformly at random 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. [3]

In order to identify the lengths of acquaintance chains connecting individuals for all scales of social separation, we place source and target nodes [4] 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 Fig. 1 below. The green regions 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. This model will enable us to derive analytical solutions for the delay of social search for any given finite size network, as opposed to the common practice of deriving bounds when the network size grows large.

The generic name we give to all networks constructed in this manner is the *octopus model*, in which the body of the "octopus" is formed by the local knit friendship circle of an individual, and the casual and relatively weak ties form the legs of the "octopus". It is

---

[3] In Model I, formation of long-range connections uniformly at random over the network domain is mainly motivated by the recent findings (Goel et al., 2009) indicating that the average length of acquaintance chains connecting individuals in experimental small-world studies becomes larger (around 50 individuals) than the common belief of 6 if the missing data is accounted for correctly. We obtain surprisingly close estimates for our model parameters in Appendix E when two different corrected experimental data sets of Goel et al. (2009) accounting for missing chains are used for parameter estimation. In a future work, we are planning to extend the same analysis to other models in which the long-range contact formation probability is distance dependent, and are expecting to arrive at the same qualitative conclusions.

[4] We will use the term *node* as a technical term to refer to a network element modeling an individual in a social network.
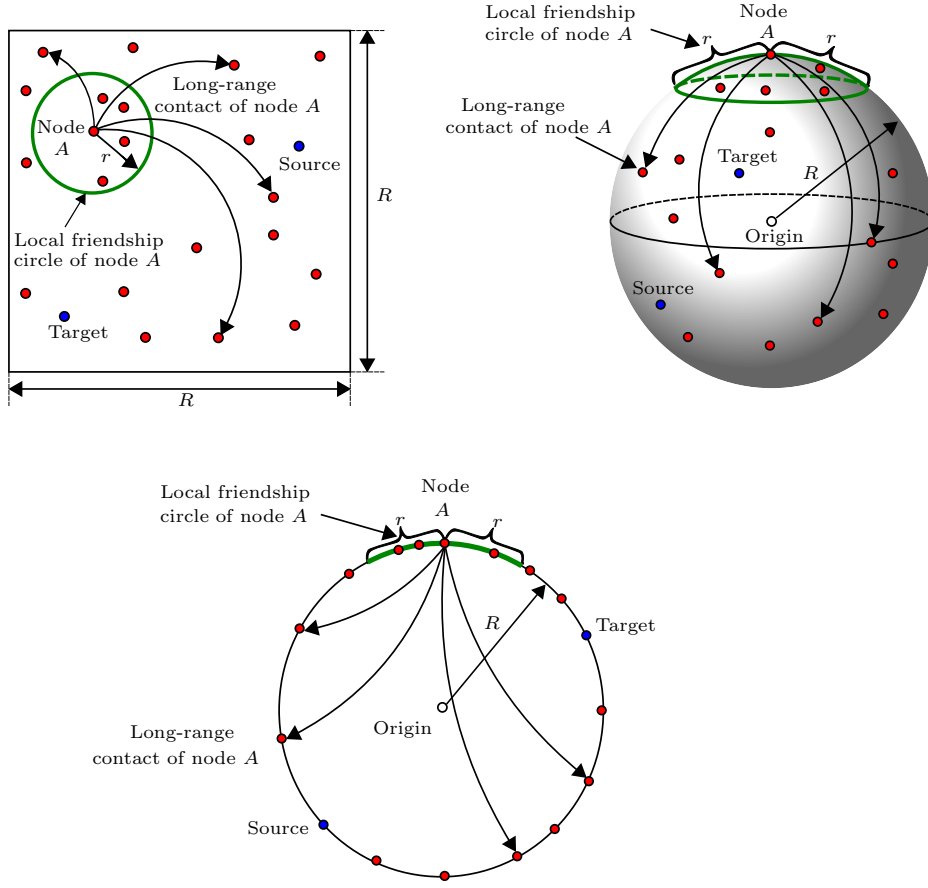
Fig. 1. Typical realizations of the social network for rectangular, spherical and circular network domains.

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 consisted of discrete lattices (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 the 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 parametrization 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
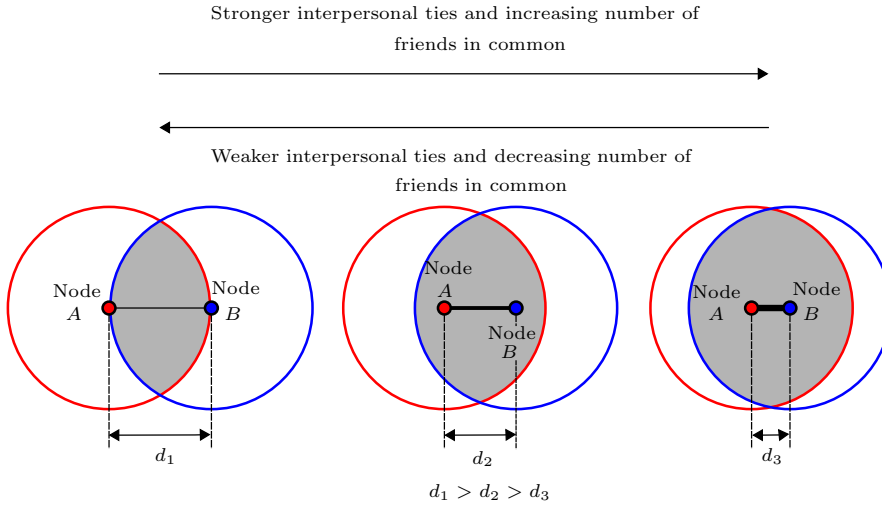
Fig. 2. Strong and weak ties.

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) claims that the tie between two individuals 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. 2, 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. 2.

3.2. *Social Search*

The social 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 [5]. One of them is assigned as the message originator (i.e., source),

---

[5] *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., target individuals). For example, Milgram and his colleagues (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

11

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 groups (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. 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 an 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 step 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. The second implicit assumption in these steps is that each message holder is able to select her best contact closest to the target individual.

3.3. *Derivation of Delay Formula in Model I*

We now present the derivation of the delay formula for social search. This part of the paper will be rather technical compared to most other parts. The details of the calculations can be skipped at the first reading as the exposition of Section 6 discussing the implications of the derived delay formulas is based on our main findings summarized in Tables 1 and 2. We show steps of calculations only for rectangular network domains. Local friendship circles for rectangular network domains are easy to visualize, 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 ensure 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

---

selection bias into 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.
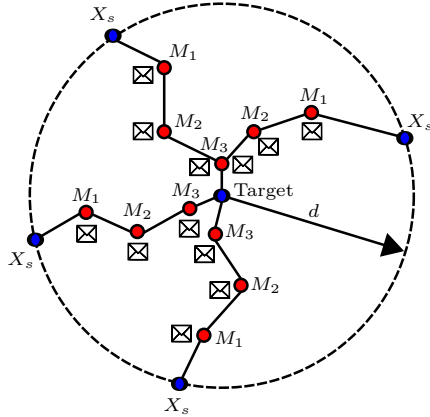
Fig. 3. Pictorial illustration of the spherical symmetry property of $T(d)$.

to guarantee network connectivity. If a connected discrete lattice is used as a substrate network as in the Watts and Strogatz (1998) model or in the Kleinberg (2000a,b) model, one can further remove this technical assumption. We have also performed agent-based simulations to identify the effects of the dense network assumption on the delay formula, and have obtained reasonably good qualitative and quantitative matches between simulation and analytical results for networks 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. We report the details of our agent based simulations in Section 4.

We denote the delay of social search in the dense network limit by T, the location of the message at the kth step 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 ρ 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 expressed analytically as

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

where $E\left[\cdot\right]$ denotes expectation.

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 social search 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. 3. 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 $r \leq d < 2r$
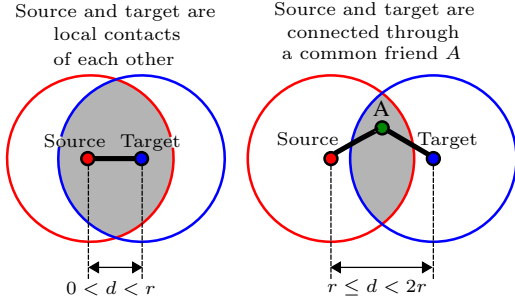
Fig. 4. Lengths of acquaintance chains connecting source and target nodes when $0 < d < r$ and $r \le d < 2r$.
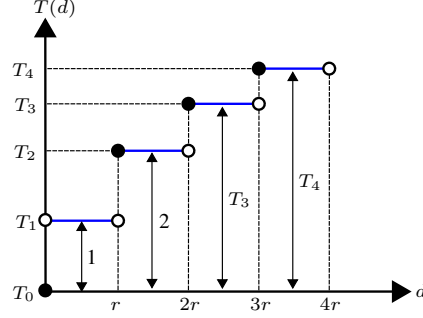

Fig. 5. Step function property of $T(d)$ with jumps at integer multiples of r.

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. 4) for this 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. 4 and 5. 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

$$T_0 = T(0) = 0,$$
$$T_1 = T(d) = 1 \text{ for } 0 < d < r,$$
$$T_2 = T(d) = 2 \text{ for } r \le d < 2r,$$
$$T_k = T(d) \text{ for } (k-1)r \le d < kr \text{ and } k \ge 3.$$

We first illustrate the calculations for T_3, and then generalize them to any $k \ge 4$. 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 step. In the first type, the message after the first step 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 acquaintance 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. 6 is an example of message trajectory in which it takes two steps for the message to reach the target.

In the second type, the message after the first step 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 shares a common friend with the target node. For such chains, the source node has two options when selecting the next step message holder: the message is forwarded through either a long-range contact or a short-range contact at the first step.
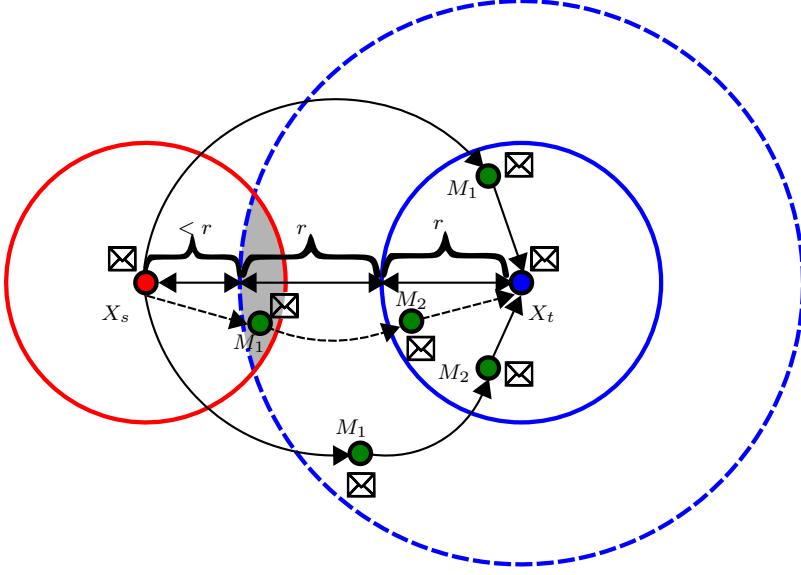
14

Fig. 6. Example trajectories of the message and the acquaintance chains connecting source and target nodes when $2r \leq \rho(X_s, X_t) < 3r$.

The bottom solid path in Fig. 6 shows an example of message trajectory in which the message is first forwarded through one of the source node's long-range contacts. On the other hand, the bottom dotted path in Fig. 6 shows an example of message trajectory in which the message is first forwarded through one of the source node's short-range contacts. In either case, the message at the first step is delivered to an intermediate message holder whose social separation from the target node is between r and $2r$. Then, it takes two steps 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 step. These observations are pictorially shown in Fig. 6, and are formally expressed as

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

where $\mathbf{1}_{\{X \in \mathcal{E}\}}$ 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 step 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,

15

$P\left(0 < \rho(M_1, X_t) < r\right)$ is calculated as

$$P\left(0 < \rho(M_1, X_t) < r\right) = \sum_{n=1}^{\infty} Q(n) P\left(0 < \rho(M_1, X_t) < r | N = n\right)$$

$$= \sum_{n=1}^{\infty} \sum_{i=1}^{n} Q(n) \binom{n}{i} \alpha^i (1 - \alpha)^{n-i}$$

$$= \sum_{n=1}^{\infty} Q(n)(1 - \alpha)^n \left(\left(\frac{1}{1 - \alpha}\right)^n - 1\right)$$

$$= 1 - \varphi(1 - \alpha),$$

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:

$$T_{k+1} = 1 + E\left[T\left(\rho\left(M_1, X_t\right)\right)\right]$$

$$= 1 + E\left[\sum_{m=1}^{k} T_m \mathbf{1}_{\{(m-1)r \le \rho(M_1, X_t) < mr\}}\right]$$

$$= 1 + T_k P\left((k-1)r \le \rho(M_1, X_t) < kr\right) + \sum_{m=1}^{k-1} T_m P\left((m-1)r \le \rho(M_1, X_t) < mr\right).$$

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

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

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

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

Using rectangular network domains to model the social space has pros and cons. A disadvantage of using rectangular network domains is the edge effects that occur in this type of network. When a node comes closer to the edges of the network, it starts to have fewer 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 far from the edge of the network. [6] 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.

We also extended 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 illustrates the robustness of our delay formula against different network geometries. These results are summarized in Tables 1 and 2, the latter of which also provides explicit delay formulas for social networks with various node degree distributions for the number of long-range connections.

Beyond the average delay of social search (Milgram, 1967; Travers and Milgram, 1969; Dodds et al., 2003), our techniques can also calculate the entire probability distribution of the delay of social search. For $k \geq 2$, we let $p_k(j)$ denote the probability that the delay of social search is equal to j when the source-target social separation d is in $[(k-1)r, kr)$, i.e.,

$$p_k(j) = P(T(d) = j) \text{ for } d \in [(k-1)r, kr).$$

We define $p_1(j)$ for $0 < d < r$ similarly. Intuitively, $p_k(j)$ represents the fraction of individuals with social separation d in $[(k-1)r, kr)$ that can be connected by an acquaintance chain of length j by means of a targeted search process. As explained above, $T(d) = 1$ for $0 < d < r$ and $T(d) = 2$ for $r \leq d < 2r$, and therefore, $p_1(j)$ and $p_2(j)$ only contain a probability mass of unity at $j = 1$ and $j = 2$, respectively. When $k \geq 3$, repeating steps similar to those used to obtain the average delay formula in Eq. (1), the following recursive solution for each $j \in \{2, 3, \cdots, k\}$ can be obtained:

$$p_k(j) = p_{k-1}(j-1)\varphi(\beta_{k-1}) + \sum_{i=1}^{k-2} p_i(j-1)\left(\varphi(\beta_i) - \varphi(\beta_{i+1})\right), \tag{2}$$

where $\varphi(t)$, $t > 0$, is the probability generating function of the distribution of the number of long-range connections per node, and β_i's for different network domains are given in Table 1.

---

[6] 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 1
General formula for the delay of social search (Model I).

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

$T_k = T(d)$ for $(k-1)r \le d < kr$, 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.

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

The details of the derivation of Eq. (2) are presented in Appendix B. Due to the dependence of $p_k(j)$ to both k and j indices simultaneously, further simplification of Eq. (2) to a more compact solution free from the recursion as in the average delay formula for the social search seems to be more difficult; however the recursive solution provided in Eq. (2) can be computed in a very computationally efficient way using a standard personal computer.

4. Model I: Agent Based Simulation

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 bottom-up computational tool to explain many diverse 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 originating 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 uniformly at random 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

18

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

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

$\gamma > 1$ 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$(\gamma, \beta_i) = \sum_{n=1}^{\infty} \frac{\beta_i^n}{n^\gamma}$. $\lambda > 0$ is the mean value of the Poisson distribution.

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

M is a non-negative integer determining the upper bound for binomial, uniform and Bernoulli distributions.

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 make use of 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 [7], and therefore

---

[7] A simple estimate for this number can be obtained as follows (Bernard et al., 2001; Newman, 2003b).

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. 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 a fixed network geometry (rectangular network domains) and a fixed node degree distribution for long-range connections (one long-range contact per node).

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 separation, the number of chains initiated is set to at least 1600 to obtain statistically significant results for the delay of social search. [8]

We built our agent based simulator by using the C programing language, and all of our simulations used 200 CPUs running in parallel. Even with this much computational power, some of the simulations for large network domains 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 (i.e., numerical computation is more than 2×10^6 times faster than simulations) in terms of the computation time required to obtain predictive and explanatory results in social search processes.

Our simulation setting and parameters are summarized in Table 3. In Figure 7, 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 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, albeit possible edge effects due to target node

---

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 λ 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.

[8] 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.

20

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 step 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.
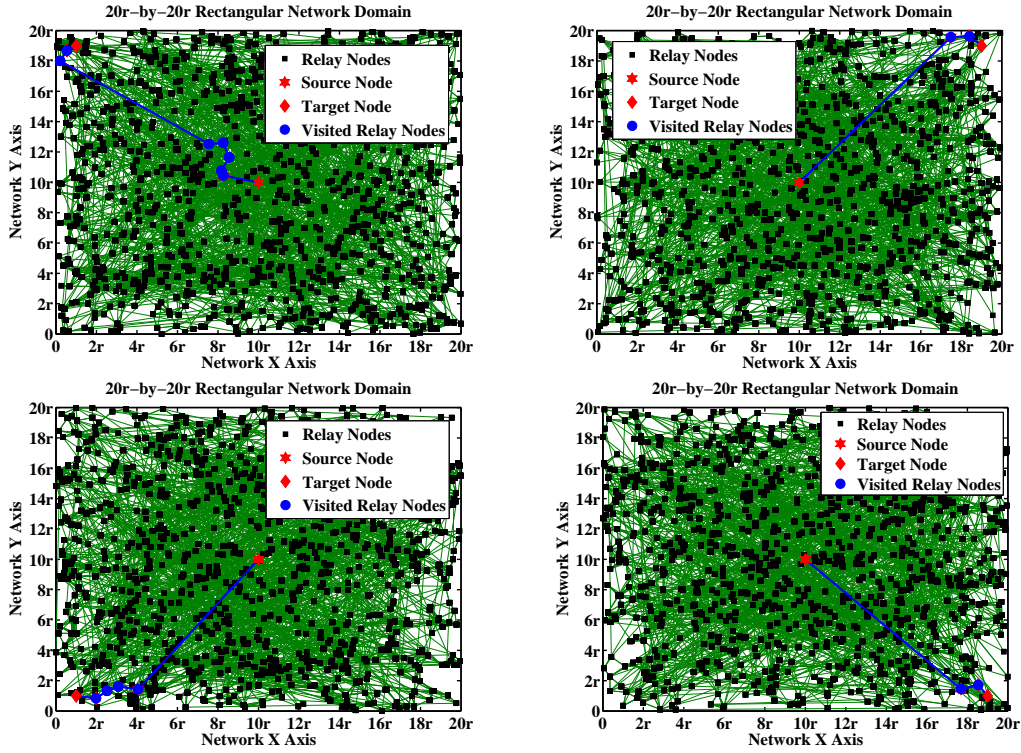
Fig. 7. Typical realizations of the social search process in which the delay becomes equal to 3 (top right and bottom right figures), 5 (bottom left figure) and 8 (top left figure).

locations. In particular, the message in some cases readily enters a close vicinity of the target node by means of a long-range contact. (See the typical realizations of the search process exhibited by top right, bottom left and bottom right figures.) Once the message is in this small vicinity of the target node, it takes several more steps 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 steps.

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 has 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 the top left part of Fig. 7. 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 Figure 8, where the unit of the delay of social search is the number of steps required to connect source and target nodes, we compare our analytical formula against the results of our agent based simulations for a $250r$-by-$250r$ rectangular network domain. The agent based simulations were also performed for other rectangular network domains with side lengths varied between $50r$ and $500r$ as summarized in Table 3. We only report our
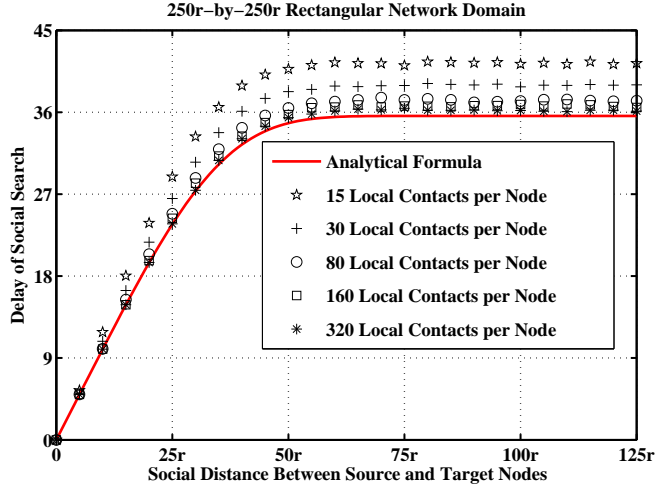
22

Fig. 8. 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. The unit of delay is the number of steps to complete the search process. The target node is placed at the center of the network domain to avoid edge effects.

results for $250r$-by-$250r$ rectangular network domains because the convergence behavior of social search as observed in simulations to analytical results looked essentially the same for all cases considered. 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 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 deviation between analytical and simulation results becomes 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 estimated by the analytical formula deviates from simulation results only by 2.7%. (See the curve corresponding to 160 local contacts per node in Fig. 8.) Furthermore, the gain in accuracy 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 Fig. 8.) 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.

This result has some further ramifications. In our analytical model, we assume that two individuals are local contacts of each other if the social separation between them is smaller than our model parameter r. Based on our simulation results, we can relax this analytically helpful but sociologically inaccurate assumption. It will be enough to assume that an individual has only around 300 friends in her local social neighborhood even if this neighborhood may contain more than 300 people. This relaxation can only create a problem at the very last step of the social search chain, where our analytical model

23

predicts that the target individual is connected to any message holder lying in her local neighborhood. On the other hand, with the above relaxation on the model, the target individual may not be connected to a message holder even if the two lie in local social neighborhoods of each other. However, at this level of the targeted social search process, the whole population is already thinned enough that the message will be very likely to take at most several more steps than the delay predicted by our model before the final delivery. [9]

5. Model II: Nested Sequence of Small-worlds

Mainly motivated by Kleinberg's results (Kleinberg, 2000a,b), an algorithmically small social network is considered in the computer science community to be a network in which the delay of social search grows no faster than logarithmically or poly-logarithmically with the number of individuals in the network. In this part of the paper, we will show that our basic model, Model I, can be extended to also account for this commonly used notion of smallness.

Kleinberg used a rectangular grid on top of which random long-range connections are added by means of distance dependent probability distributions to model social networks. He showed that social networks become algorithmically searchable if and only if the probability of having a long-range connection between two individuals decay as a power-law of their social distance, as measured on the grid, with exponent matching the dimension of the grid. As Watts (2003) puts it, "the network achieves a kind of optimal balance between the navigational convenience of the lattice and the distance-cutting power of long-range shortcuts" with this particular choice of exponent for distance dependent long-range contact formation probabilities.

We will now propose an extension of our basic model to provide an alternative explanation for the small-world phenomenon from an algorithmic point of view. The generic name that we give to these extended network models is the *generalized octopus model* for social network modeling.

To this end, we view the social space as a nested sequence of small-worlds $\mathcal{SW}_1 \supset \mathcal{SW}_2 \supset, \cdots, \supset \mathcal{SW}_K$ in which the aim of each message holder, by using her long-range and short-range connections, is to advance a message from a greater small-world spanning longer social distances to a smaller one spanning shorter social distances to deliver the message to a target located in $\mathcal{SW}_K$. Long-range connections in each small-world are formed uniformly at random within this world as in the original small-world model (Watts and Strogatz, 1998). Therefore, unlike Kleinberg's model, formation of long-range connections within a small-world does *not* depend on an individual's particular perception of distance, yet we can still find long-range connections at all scales of social distance as the social search progresses from one small-world to another one. From this point of view, our results indicate that the original Watts-Strogatz small-world model is indeed enough to explain the logarithmically growing delay of social search.

The Chinese peasant farmer example provided by Watts (2003) puts the above discussion into a more concrete context. Suppose we, as individuals living in the United States, are presented with a small-world challenge of finding a Chinese peasant farmer living in

---

[9] For example, if the target is a stock broker in Boston, the message ought to be with individuals living in Boston and working in related professions at the very last steps.

a given rural area in China. At the initial steps of the search chain, we are confronted with the problem of advancing the message from our small-worlds at the global scale to a small-world of the next step message holder living, for example, in China. Once the message is delivered to an individual living in China, the scale of the social search, all of a sudden, reduces from the global scale to the scale of mainland China, and the small-world problem confronted by the message holder now is to search her small-world in China to advance the message to the next phase of the social search process in which the next step message holder and the target lie in a smaller small-world. The social search process progresses through a nested sequence of small-worlds in this manner until the message is delivered to the Chinese peasant farmer in question.

To make these arguments mathematically precise, we only require a message holder to have N, a random number with distribution $Q(n)$, long-range connections formed uniformly at random over all individuals lying in the disk centered around the target individual with radius kr if the social distance between the message holder and the target individual is in between $(k-1)r$ and kr. A message holder uses these N long-range connections while searching for the target individual lying in the small-world spanning the social space between them. Other than this new long-range contact formation rule, all the interpretations and assumptions given for the original octopus model in previous sections hold for this generalized octopus model.

Following similar steps to those used above while deriving the delay of social search for Model I, we obtain the following recursive formula:

$$T_{k+1} = 1 + T_k \varphi(\beta_{k,k+1}) + \sum_{i=1}^{k-1} T_i(\varphi(\beta_{i,k+1}) - \varphi(\beta_{i+1,k+1})), \tag{3}$$

for the delay of social search in Model II, where T_k is defined as in Section 3, $\varphi(t)$ is the probability generating function of the number of long-range connections per node, and $\beta_{i,k}$ is given by $1 - \left(\frac{i-1}{k}\right)^2$ for rectangular network domains, by $\frac{\cos\left(\frac{(i-1)r}{R}\right) - \cos\left(\frac{kr}{R}\right)}{1 - \cos\left(\frac{kr}{R}\right)}$ for spherical network domains and by $1 - \frac{i-1}{k}$ for circular network domains.[10] The details of the derivation of the delay formula are presented in Appendix C.

To obtain the distribution of the delay of social search in Model II, we let $p_k(j)$, for $k \geq 2$, denote the probability that the delay of social search is equal to j when the source-target social separation d is in $[(k-1)r, kr)$, i.e.,

$$p_k(j) = P(T(d) = j) \text{ for } d \in [(k-1)r, kr).$$

We define $p_1(j)$ for $0 < d < r$ similarly. As explained in Section 3, $T(d) = 1$ for $0 < d < r$ and $T(d) = 2$ for $r \leq d < 2r$, and therefore, $p_1(j)$ and $p_2(j)$ only contain a probability mass of unity at $j = 1$ and $j = 2$, respectively. When $k \geq 3$, repeating the same steps used to derive the distribution of the delay of social search in Model I, the following recursive solution for each $j \in \{2, 3, \cdots, k\}$ can be obtained:

---

[10] Unlike Model I, we do not require a randomly formed long-range contact to be always outside of the local neighborhood of a node in Model II to make the analytical formulas simpler. Otherwise, we will need to consider situations in which a long-range contact of a node lies in the intersection of the local neighborhood of this node and the relevant disks centered around the target node while calculating $\beta_{i,k}$'s.
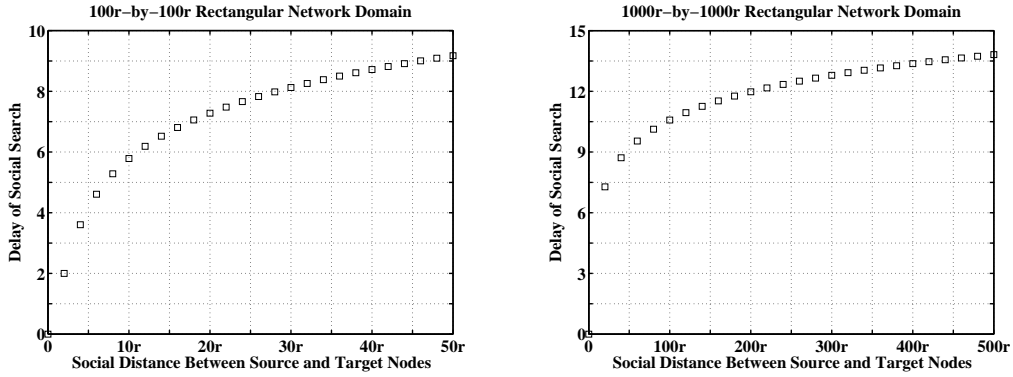
Fig. 9. Delay of social search in Model II for different network sizes with N set to 1. Side lengths of a rectangular network domain are increased 10 times, but the maximum delay increases only 1.5 times.

$$p_k(j) = p_{k-1}(j-1)\varphi(\beta_{k-1,k}) + \sum_{i=1}^{k-2} p_i(j-1)\left(\varphi(\beta_{i,k}) - \varphi(\beta_{i+1,k})\right) \tag{4}$$

where $\beta_{i,k}$'s are as given above.

The details of the derivation of the delay distribution are presented in Appendix D. Even though these recursions cannot be reduced to a simpler form easily as in the delay formula in the original octopus model, they can be solved very efficiently on a standard personal computer today.

Fig. 9 showing the growth rate of the social search delay for different network sizes should be intuition-confirming for researchers familiar with logarithmically growing social search delays. We set N to 1 to generate this figure. We increase the side length of the rectangular network domain 10 times from $100r$ to $1000r$, but the maximum delay of social search grows only 1.5 times from 9.17 steps to 13.82 steps. Note also that the delay of social search under the generalized octopus model seems to exhibit self-similarity: two graphs look similar to each other except for a change of scale in the social distance and delay dimensions.

6. Discussion of Results

In this section, we discuss the implications of our analytical formulas, focusing 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 node degree heterogeneity on the delay of social search.

6.1. *Roles of Short-range and Long-range Contacts in Social Search*

In Fig. 10, we plot the delay of social search predicted by our analytical formulas as a function of source-target separation for spherical network domains with different sizes and for different numbers of long-range connections per node. For other network domains, the same conclusions continue to hold. If we start our discussion with the delay curves depicted for Model I, top two figures in Fig. 10, we observe two regimes in the delay as
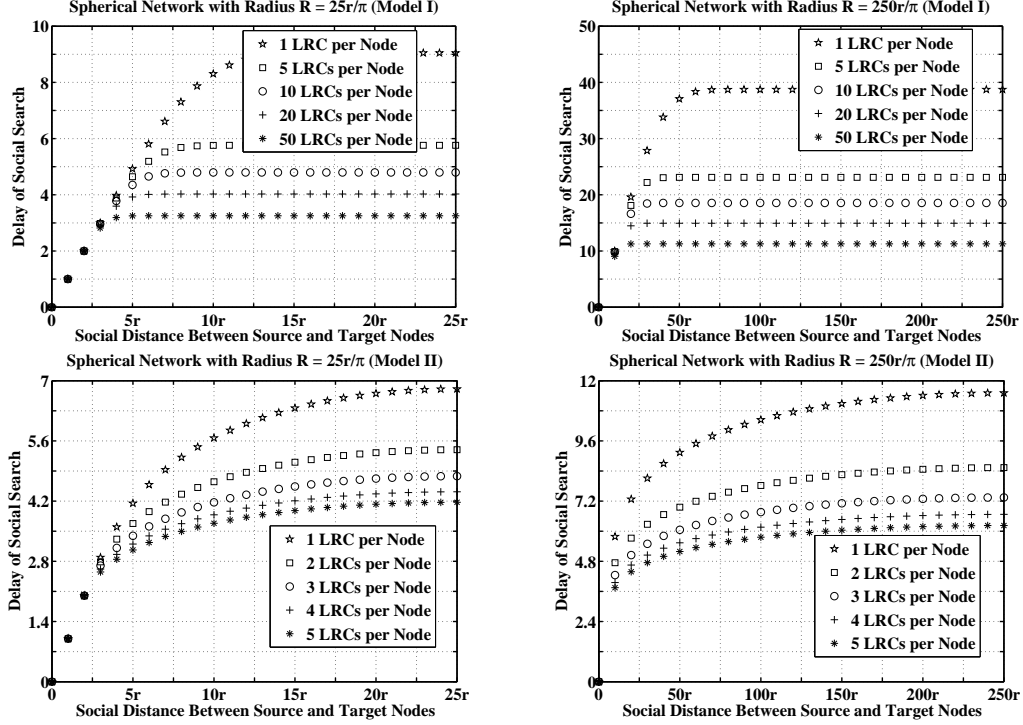
26

Fig. 10. Delay of social search in both models as a function of source-target separation with various numbers of long-range connections (LRCs) per node for a spherical network domain with radius $R = \frac{25r}{\pi}$ (left figures) and $R = \frac{250r}{\pi}$ (right figures). The unit of delay is the number of steps to complete the search process.

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 in Model I.

For Model II, on the other hand, we observe a logarithmic growth for the delay of social search as a function of source-target separation. By considering Model II as a nested sequence of small-worlds, each of which is modeled by Model I and spanning decreasing social distances, this result can be interpreted as a blending of the delay curves of Model I for different network sizes in a way that leads to a logarithmic growth of the delay in Model II. It is also important to note that the logarithm function $\log(x)$ grows almost linearly for small values of its argument x, and therefore there is still a linear growth region in the delay curves of Model II for small values of the social distance between source and target individuals, but this linear growth region is much more restricted when compared with the linear growth region of the delay curves of Model I.

As the social separation between source and target nodes become larger, the delay saturates to a constant number in Model I. 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 Fig. 7.) However, once the message is in this small vicinity of the target node, the linear growth regime emerges again in Model I, and the message hops through local contacts before it is delivered to the target node.

In Model II, on the other hand, the delay of social search continues to increase gradually with increasing source-target separation, but the rate of this increase becomes smaller and smaller as the source becomes socially more and more distant from the target. This can be easily seen by recalling that the derivative of the logarithm function $\log(x)$ is equal to $\frac{1}{x}$. Therefore, for all practical purposes, even though it is more prominent in Model I, the delay of social search in Model II can also be considered to be the same for all large social separations between source and target individuals. The delay of social search for maximally separated source and target individuals in Model II is always smaller than that in Model I due to the availability of the long-range contacts at all scales of social distance in Model II.

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

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, ..."

Our theory is also consistent with small-world experiments statistically. In Fig. 11, we compare the distributions of the delay of social search predicted by Model I in Eq. (2) and by Model II in Eq. (4) with the one obtained by using data from Travers and Milgram (1969). For a rectangular network domain with $R = 32r$, $d = 15r$ and $N = 5$, Model I yields an average delay around 6.14, whereas it is around 6.55 in Milgram's experiment (Travers and Milgram, 1969). Some further empirical evidence in favor of Model I is also provided in Appendix E. For a rectangular network domain with $R = 28r$, $d = 13r$ and $N = 1$, Model II yields an average delay around 6.36. Moreover, the delay distributions statistically resemble each other: the standard chi-square test (with six bins) for discrete distributions produces a P-value around 0.53 for Model I and 0.63 for Model II, which compares well with the results reported by Watts et al. (2002).[11] A larger P-value

---

[11] Sample selection bias, low participation rates and chain attrition rates (Kleinfeld, 2002; Schenettler, 2009a,b) lead to underestimated values for the delay of social search in empirical studies. On the other hand, the fact that individuals do not always remember and choose their best contacts to reach the target (McCarty et al., 2001; Killworth et al., 2006; Bell et al., 2007) causes the delay numbers to be overestimated in empirical studies, and therefore the net effect of these conflicting forces are undetermined. Hence, we do not consider the chain attrition rates in the analysis above. If it is taken to be 0.25
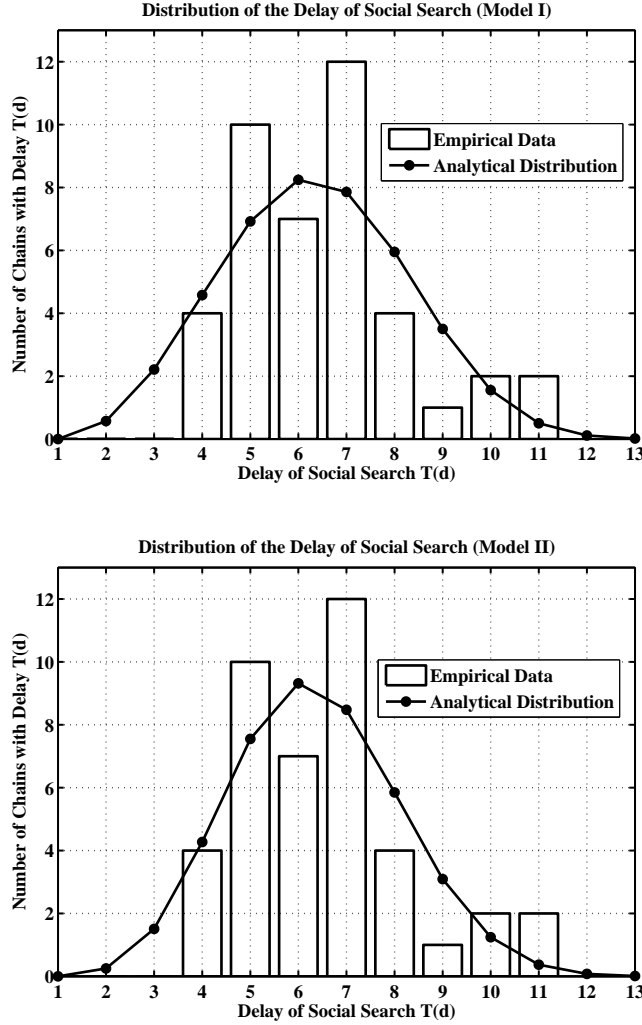
Distribution of the Delay of Social Search (Model I)

Distribution of the Delay of Social Search (Model II)

Fig. 11. Distributions of the delay of social search obtained by using the proposed models and Travers and Milgram (1969) data for the 42 completed chains that originated in Nebraska. 24 completed chains originated in Boston were excluded to keep the source-target separation in the models constant. Model I parameters are $R = 32r$, $d = 15r$ and $N = 5$ for a rectangular network domain. Model II parameters are $R = 28r$, $d = 13r$ and $N = 1$ for a rectangular network domain.

means a better fit between the analytical distribution and the empirical distribution, hence Model II produces analytical delay distributions closer to empirical ones.

Dodds et al. (2003) reports that geographical distance is one of the frequently used social dimensions along which individuals measure social distance in a typical small-world experiment. That is, geographical distance is one type of social categorization frequently

---

as in Watts et al. (2002), we can obtain a P-value of 0.22 for Model I and 0.56 for Model II with R and d set to $45r$ and $22r$, respectively, in both models, different from the R (and d) values set to $32r$ ($15r$) and $28r$ ($13r$) above due to attrition.

used for measuring social similarity. Therefore, if we assume that individuals primarily use geographical information to forward messages, which is a reasonable assumption for a typical small-world experiment in the light of the results reported by Dodds et al. (2003), the choice of the network domain means that we approximate the United States as a rectangular network domain, and the choice of source-target separation, by ignoring the possible edge effects, means that source and target nodes are located at the center and at the edge of the network domain, respectively.

In reality, people usually use multiple social dimensions to measure social similarity among themselves, and therefore associating only geographical meaning to our model parameters is not the most precise way of interpreting the choice of model parameters. It should also be noted that our results are purely analytical, and the recursive solution for the distributions of the social search delay in Eq. (2) and Eq. (4) can be computed in seconds using an ordinary computer. On the other hand, the model proposed by Watts et al. (2002) requires running simulations to obtain the distribution of the social search delay, with possibly long simulation run times for large networks.

In summary, both of our models produce qualitatively and quantitatively convincing results when compared to empirical small-world studies. However, Model II is more favorable than Model I due to its ability to produce analytical delay distributions fitting better to the empirical distributions and logarithmically growing delay values. Therefore, we focus our attention on Model II for the rest of the paper, and provide corresponding figures for Model I only in appendices as supplementary data. We believe that combining our models with distance dependent long-range formation probabilities (Kleinberg, 2000a,b) will further improve the matches observed between analytical and empirical results.

6.2. *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. Adamic et al. (2001) showed that hubs can be important for minimizing the delay of social search. However, their model assumes that nodes use the number of friends of their friends as the next step message holder selection criterion. This is in contrast to what Dodds et al. (2003) found in their experiments: "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
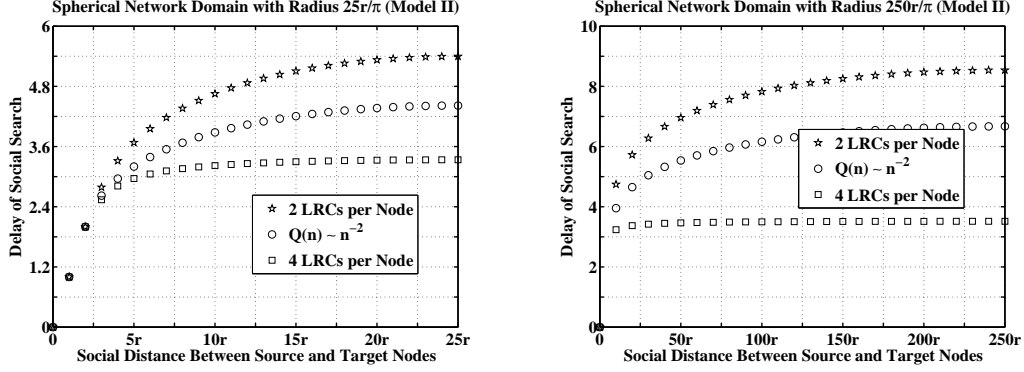
Fig. 12. Delay of social search for scale-free networks and the networks with fixed numbers of long-range connections (LRCs) per node for spherical network domains with radius $R = \frac{25r}{\pi}$ (left figure) and $R = \frac{250r}{\pi}$ (right figure). The unit of delay is the number of steps to complete the search process.

individual.

To this end, we generated two types of networks for spherical network domains with various sizes under both models. In the first type, the number of long-range connections per node, N, 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). [12] In our models, 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.

In the second type, N is set to deterministic numbers. Since the expected number of long-range connections per node is infinite in the first network type, one can 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 in Fig. 12 are again quite surprising, and illustrate the flaw in this intuition.

The delay of social search for power-law networks under both models becomes smaller than that for networks with $N = 2$ but greater than that for networks with $N = 4$. These results, together with findings of Dodds et al. (2003), show that hubs neither have high relevance to the social search nor can their existence significantly reduce the length of social search 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

---

[12] In previous studies, it has been observed that even though node degree distributions in real social networks may have heavy tails, they still exhibit quite severe cutoffs (Bernard et al., 1991; Kossinets and Watts, 2006; Rapoport, 1963). As a result, these distributions are never scale-free at all scales. The fundamental reason behind this behavior is that maintaining social relationships is costly, and therefore people cannot have an infinite number friends (Amaral et al., 2000). Our models are flexible enough that scale-free distributions for the number of long-range connections can be easily replaced with any right skewed node degree distribution by calculating its probability generating function, and then the same analysis in this section can be repeated for those right skewed distributions.

31

as geography, occupation and education for selecting intermediaries in a typical social search process.

Finally, it should be noted that these conclusions depend critically on the empirically supported assumption that individuals in a typical small-world experiment relatively rarely nominate an acquaintance as a next step message holder because she has many friends (Dodds et al., 2003). Our results show that this behavior, although perhaps irrational, suppresses the effects of hubs in targeted search processes. On the other hand, their effects are magnified if individuals nominate the next step message holders solely based on the numbers of friends of their friends (Adamic et al., 2001). In the middle ground between these two extreme ends, we can envision a large spectrum of human decision-making processes for choosing next step message holders in targeted search processes. As a result, even though our results and the results of Adamic et al. (2001) are representative of the impacts of the hubs at the two extreme ends, further research is needed to precisely determine the effects of hubs if individuals mix between social dimensions like geography, profession and education, and the numbers of friends of their friends while choosing next step message holders.

6.3. *The Effect of Node Degree Heterogeneity 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 less precise question: What is the effect of network heterogeneity on the delay of social search? We quantify the "degree" of heterogeneity existing in a society by the variance of the distribution of the connectedness, and analyze this point by studying simple but intuitive network models in which the numbers of long-range connections per node exhibit large variances. We will show that the distribution of the number of long-range connections over individuals is as important as the average number of long-range connections per individual in terms of the determination of the delay of social search.

Following Burt (1992), we can associate three kinds of capital to individuals in a social network. The first one is the financial capital such as cash in hand, reserves in the bank and lines of credit. The second one is the human capital exhibited by individuals' natural qualities (e.g., charm, health and intelligence) combined with other acquired skills. The third one is the social capital: who is connected to whom. The first two types of capital are related to individuals' local properties, whereas the social capital is related to the social network structure. One measure of the social capital possessed by an individual is the number of friends (with non-overlapping social circles) of this individual. The more non-ovelapping friends you have, the more channels opening up to a diverse set of opportunities located in different parts of the network you have. Or, the bigger the size of the personal network you have, the more likely you are to hear a useful bit of information aired in 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." From this perspective, our discussion in this part of the paper can also be related to the concept of social capital as measured by the number of long-range
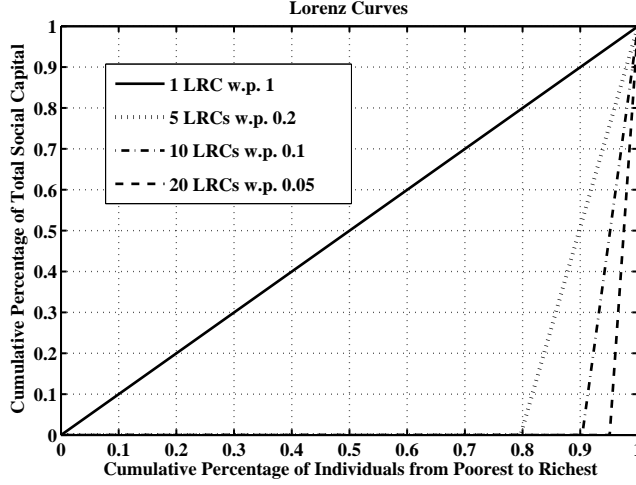
Fig. 13. Lorenz curves for different values of n. For each n, the number of long-range connections (LRCs) is equal to n with probability (w.p.) $\frac{1}{n}$, or to 0 w.p. $1 - \frac{1}{n}$.

connections per individual, and, more specifically, to the effects of variance in social capital on the delay of social search.

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 n with probability $\frac{1}{n}$, or 0 with probability $1 - \frac{1}{n}$. We analyze the delay of social search for different values of n: $n = 1, 5, 10$ and 20. The $n = 1$ case can be considered as corresponding to the situation in which the total existing social capital in the network is evenly distributed among all individuals. The large n case can be considered as corresponding to the situation in which the total existing social capital in the network is concentrated on a few hub-like individuals.

One way to visualize and to quantify the degree of heterogeneity among individuals for different values of n is to draw the corresponding Lorenz curves. Lorenz curves showing the variation in social capital for different values of n are shown in Fig. 13. As seen in this figure, the Lorenz curve drawn for $n = 1$ is the line of perfect homogeneity. As n increases, the Lorenz curves are shifted to the right, and the heterogeneity among individuals increases, approaching the perfect inhomogeneity case. In particular, Gini coefficients are equal to 0, 0.8, 0.9 and 0.95 for $n = 1, 5, 10$ and 20, respectively.

We plot the curves showing the effect of node degree heterogeneity on the delay of social search for spherical network domains in Fig. 14. The same conclusions continue to hold for other network domains. As these figures clearly show, the delay of social search is minimized when all the nodes have the same degree. As node degree becomes more varied, the delay starts to increase. The adverse effects of node degree heterogeneity on the growth behavior of the social search delay becomes more prominent for small network sizes.

The reason for this phenomenon is similar to that behind the effect of hubs on the delay of social search. When the total existing social capital (measured by the number of
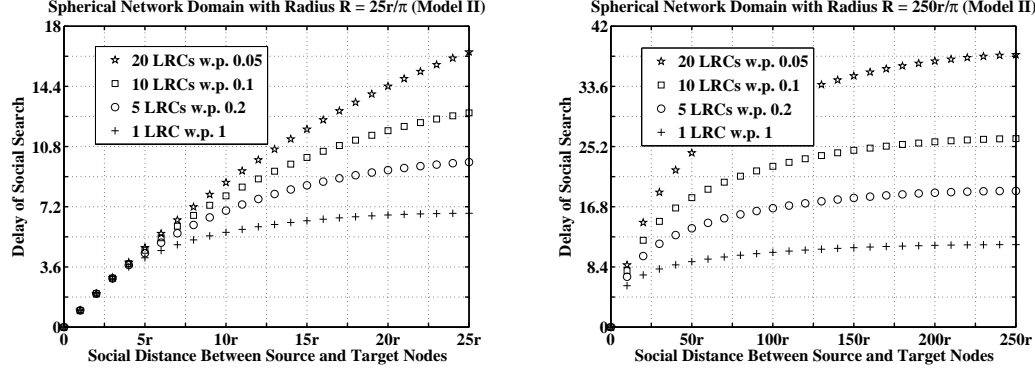
Fig. 14. Effect of node degree heterogeneity on the delay of social search for spherical network domains with radius $R = r\frac{25r}{\pi}$ (left figure) and $R = \frac{250r}{\pi}$ (right figure). Each node has either n long-range connections (LRCs) with probability (w.p.) $\frac{1}{n}$, or 0 LRC w.p. $1 - \frac{1}{n}$. The unit of delay is the number of steps to complete the search process.

long-range connections per individual) 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 step 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 small for the entire end-to-end path. As a result, we see an almost linear increase in the delay as a function of source-target separation for small networks with large variance in node degrees.

On the other hand, as the network size increases, it becomes more probable to have at least one individual with large social capital as a relay. Once the message reaches one such individual, it is likely to be close to the target node at the next step through one of her large number of friends. As a result, large network sizes relieve the negative effect of network heterogeneity to some extent, and we start to observe a sub-linear growth for the delay of social search again as a function of source-target separation.

7. Conclusions

We have studied the lengths of social search chains needed to connect any two individuals in a social network by means of a targeted search process (i.e., the delay of social search) as a function of the social separation between them. We have proposed a unified analytical framework involving two related models in which social networks are modeled as small-world random geometric graphs, and have obtained analytical solutions for the delay of social search as well as its distributions for both models. The first model is referred to as the *octopus model*, and the second model is referred to as the *generalized octopus model*.

In our network models, each individual has both local short-range connections and random long-range connections. 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 their social separation is small enough. Long-range contacts of an individual are formed uniformly at random over other individuals. These long-range connections can be regarded as bridges

34

connecting socially distant regions in a network. Our models allow the number of long-range connections per individual to be drawn from any given distribution including scale-free distributions. Therefore, they can be considered to capture important properties of both small-world networks of Watts and Strogatz (1998) 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 also 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.

We have also performed agent-based simulations on large computer clusters to illustrate our analytical expressions. Simulation results indicate that when the number of local contacts per individual 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 our analytical formulas deviates at most 2.7% from simulation results. 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 take only several seconds to compute on an ordinary personal computer, i.e., a more than 2×10^6 fold computational gain.

Our analytical formulas help us to quantify some hidden dynamics of social search. To start with, they indicate 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) qualitatively, as well as with the conclusions that can be drawn based on the bounds obtained by Kleinberg (2000a,b). We have also compared the analytical distributions of the delay of social search predicted by our models and the empirical distribution obtained by Travers and Milgram (1969), and have observed good statistical matches between analytical and empirical distributions.

Finally, we have studied the effect of hubs and node degree heterogeneity 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 rarely nominate their friends as next step message holders because their friends have many friends in a typical social search experiment (Dodds et al., 2003). Similarly, node degree heterogeneity adversely affects the delay of social search.

8. Acknowledgments

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 step 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

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

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

To obtain $P\left((k-1)r \leq \rho(M_1, X_t) < kr\right)$, we first realize that the message after the first step 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

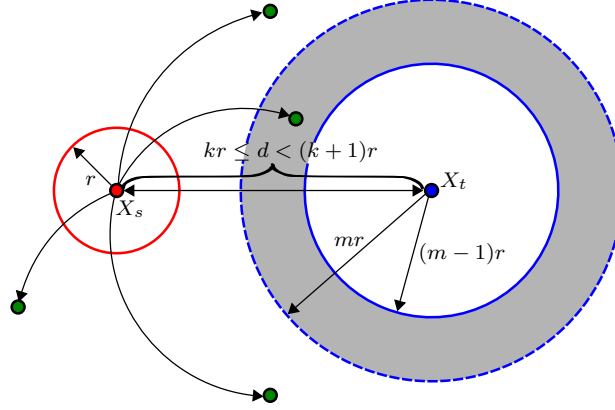$$P\left((k-1)r \leq \rho(M_1, X_t) < kr\right) = \varphi(\beta_k).$$

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

Appendix B.

We define the probability vector $\vec{p}_k$ for $k \geq 2$, which corresponds to the probability distribution of the delay of social search when the social search starts at a source node with social separation from the target node in $[(k-1)r, kr)$, as

$$\vec{p}_k = [p_k(0), p_k(1), p_k(2), \cdots, p_k(k), p_k(k+1), \cdots].$$

Some immediate results for $k \geq 2$ are

$$p_k(0) = 0$$
$$p_k(1) = 0$$

and

$$p_k(j) = 0 \text{ for } j \geq k+1.$$

For $k = 0$ and $k = 1$, we define $\vec{p}_0$ and $\vec{p}_1$ as

$$p_0(j) = \mathbf{1}_{\{j=0\}} \text{ or } \vec{p}_0 = [1, 0, 0, 0, \cdots, 0, \cdots]$$

and

$$p_1(j) = \mathbf{1}_{\{j=1\}} \text{ or } \vec{p}_1 = [0, 1, 0, 0, \cdots, 0, \cdots].$$

It also directly follows that $p_2(j) = \mathbf{1}_{\{j=2\}}$ or $\vec{p}_2 = [0, 0, 1, 0, \cdots, 0, \cdots]$. For $k \geq 3$, we write $p_k(j), j \in \{2, 3, \cdots, k\}$, as

$$p_k(j) = p_{k-1}(j-1)P((k-2)r \leq \rho(M_1, X_t) < (k-1)r)$$
$$+ \sum_{i=1}^{k-2} p_i(j-1)P((i-1)r \leq \rho(M_1, X_t) < ir).$$

As in Appendix A, we can calculate $P((i-1)r \leq \rho(M_1, X_t) < ir)$ as

$$P((i-1)r \leq \rho(M_1, X_t) < ir) = \varphi(\beta_i) - \varphi(\beta_{i+1})$$

and $P((k-2)r \leq \rho(M_1, X_t) < (k-1)r)$ as

$$P((k-2)r \leq \rho(M_1, X_t) < (k-1)r) = \varphi(\beta_{k-1}),$$

which completes the derivation of the distribution of the delay of social search in Model I.

Appendix C.

To calculate T_{k+1} in Model II, we use the same first-step analysis used to calculate T_{k+1} in Model I. That is, we first look at where the message can be located after the first step, and then consider how long it takes on the average to reach the target from this point onward. For $k \geq 2$, this approach leads to

$$T_{k+1} = 1 + T_k P((k-1)r \leq \rho(M_1, X_t) < kr) + \sum_{i=1}^{k-1} T_i P((i-1)r \leq \rho(M_1, X_t) < ir).$$

For a source individual with social distance to the target in the range $[kr, (k+1)r)$, N long-range contacts are formed uniformly at random over the individuals lying in the disk centered around the target with radius $(k+1)r$ in Model II, and therefore we can obtain the above probabilities as

$$P((i-1)r \leq \rho(M_1, X_t) < ir) = \varphi(\beta_{i,k+1}) - \varphi(\beta_{i+1,k+1})$$

and

$$P((k-1)r \leq \rho(M_1, X_t) < kr) = \varphi(\beta_{k,k+1}).$$

The calculations for obtaining the above probabilities are essentially the same as those in Appendix A. Hence, we conclude the proof without repeating them here.

Appendix D.

To derive the distribution of the delay of social search in Model II, we first define the probability vector $\vec{p}_k = [p_k(0), p_k(1), p_k(2), \cdots, p_k(k), p_k(k+1), \cdots]$ for $k \geq 2$ as in Appendix B. Then, it follows that $p_k(0) = 0$, $p_k(1) = 0$ and $p_k(j) = 0$ for $j \geq k+1$, and therefore $p_2(j) = 1_{\{j=2\}}$. For $k = 0$ and $k = 1$, we define $\vec{p}_0$ and $\vec{p}_1$ such that $p_0(j) = 1_{\{j=0\}}$ and $p_1(j) = 1_{\{j=1\}}$.

For $k \geq 3$, we write $p_k(j), j \in \{2, 3, \cdots, k\}$, as

$$p_k(j) = p_{k-1}(j-1)P((k-2)r \leq \rho(M_1, X_t) < (k-1)r)$$
$$+ \sum_{i=1}^{k-2} p_i(j-1)P((i-1)r \leq \rho(M_1, X_t) < ir).$$

Calculations for obtaining probabilities $P((k-2)r \leq \rho(M_1, X_t) < (k-1)r)$ and $P((i-1)r \leq \rho(M_1, X_t) < ir)$ are essentially the same as those in Appendix A, the end results

are shown in Appendix C, and therefore the proof is concluded without repeating them here.

Appendix E.

One common problem with the empirical small-world experiments is the missing data and very low chain completion rates (Kleinfeld, 2002). Therefore, on the one hand, if 6 is the correct number for the lengths of acquaintance chains connecting individuals, any sound theoretical model should predict the *magic number* 6 for a reasonable choice of model parameters, which is the case for our models as explained in Section 6, and on the other hand, any such model should also conform well with empirical results when the missing data is taken into consideration by ignoring the accuracy of individuals remembering and choosing their best contacts to reach the target (McCarty et al., 2001; Killworth et al., 2006; Bell et al., 2007). This is what we investigate next.

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. Their findings predict longer acquaintance chains connecting source and target nodes on the globe, 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. As shown in Fig. 10, our Model I can also predict similar numbers for the delay of social search for reasonable choices of model parameters. Below, we performed a simple robustness check of our analytical results by estimating the r parameter in Model I, and obtained surprisingly very close estimates for r when it was 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 $R_{\text{US}} = 3,134$ [km]. The radius of the Earth is, on the other hand, equal to $R_{\text{Earth}} = 6,371$ [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 $R_{\text{Earth}} = \frac{280 r_{\text{Earth}}}{\pi}$, and it saturates to 11.8 when $R_{\text{US}} = 45 r_{\text{US}}$. After simple calculations, $r_{\text{Earth}} = 71.5$ [km] and $r_{\text{US}} = 69.6$ [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 $R_{\text{Earth}} = \frac{620 r_{\text{Earth}}}{\pi}$ and $R_{\text{US}} = 97 r_{\text{US}}$, which means $r_{\text{Earth}} = 32.3$ [km] and $r_{\text{US}} = 32.3$ [km]. When there are 10 long-range connections per node, we have $R_{\text{Earth}} = \frac{875 r_{\text{Earth}}}{\pi}$ and $R_{\text{US}} = 137 r_{\text{US}}$, which means $r_{\text{Earth}} = 22.9$ [km] and $r_{\text{US}} = 22.9$ [km]. For 20 long-range connections per node, we have $R_{\text{Earth}} = \frac{1240 r_{\text{Earth}}}{\pi}$ and $R_{\text{US}} = 193 r_{\text{US}}$, which means $r_{\text{Earth}} = 16.1$ [km] and $r_{\text{US}} = 16.2$ [km]. And finally for 50 long-range connections per node, we have $R_{\text{Earth}} = \frac{1960 r_{\text{Earth}}}{\pi}$ and $R_{\text{US}} = 305 r_{\text{US}}$, which means $r_{\text{Earth}} = 10.2$ [km]

Table E.1

Summary of estimated values and confidence intervals (CI) for r obtained by using two different data sets (Travers and Milgram, 1969; Dodds et al., 2003), and Goel et al. (2009) estimates.

Number of Long-range Connections	Travers and Milgram (1969) Data and Goel et al. (2009) Estimates		Dodds et al. (2003) Data and Goel et al. (2009) Estimates	
	Mean Delay (95% CI)	r_{US} in [km] (95% CI)	Mean Delay (95% CI)	r_{Earth} in [km] (95% CI)
$N = 1$	11.8 (8.5 − 15)	69.6 (48.2 − 116)	41.5 (20 − 68)	71.5 (33.9 − 222)
$N = 5$	11.8 (8.5 − 15)	32.3 (21.9 − 56)	41.5 (20 − 68)	32.3 (15.2 − 100)
$N = 10$	11.8 (8.5 − 15)	22.9 (15.4 − 36.7)	41.5 (20 − 68)	22.9 (10.8 − 71.5)
$N = 20$	11.8 (8.5 − 15)	16.2 (11 − 28.2)	41.5 (20 − 68)	16.1 (7.6 − 50)
$N = 50$	11.8 (8.5 − 15)	10.3 (7 − 17.9)	41.5 (20 − 68)	10.2 (4.8 − 31.8)

and $r_{\text{US}} = 10.3$ [km]. It should also be noted that 95% confidence intervals for the mean delay estimates in Goel et al. (2009) are quite large, and therefore these surprisingly close estimates must be interpreted with caution by considering corresponding 95% confidence intervals on the r parameter in our model. These results are summarized in Table E.1. As this topic takes further study and more empirical data to improve confidence intervals, these results provide a useful, and somewhat surprising, starting point.

Appendix F.

This appendix provides corresponding data for the effect of hubs on the delay of social search for Model I in Fig. F.1. In this case, the results for Model I should also be compared and contrasted with other relevant recent findings (Fraigniaud and Giakkoupis, 2009) indicating that the effect of power-law degree distributions on the social search delay can be significant. Our observations above are correct for finite size networks. On the other hand, the number of nodes and the network size grow large in the work by Fraigniaud and Giakkoupis (2009), where they start to harvest the gains from power-law degree distributions in the asymptotically large network regime. In fact, delay curves under Model I in Fig. 12 also indicate that the social search delay of the power-law network approaches to that of the network with four long-range connections per node as the network size increases. By continuing to increase network sizes, we can arrive the regime where the delay of social search in the power-law network is smaller than that in the network with four long-range connections per node, which is consistent with the results reported by Fraigniaud and Giakkoupis (2009). However, for any finite sized network, the power-law network having an infinite average number of long-range connections per node will not perform any better, in terms of the delay of social search, than networks with finite average numbers of long-range connections per node.
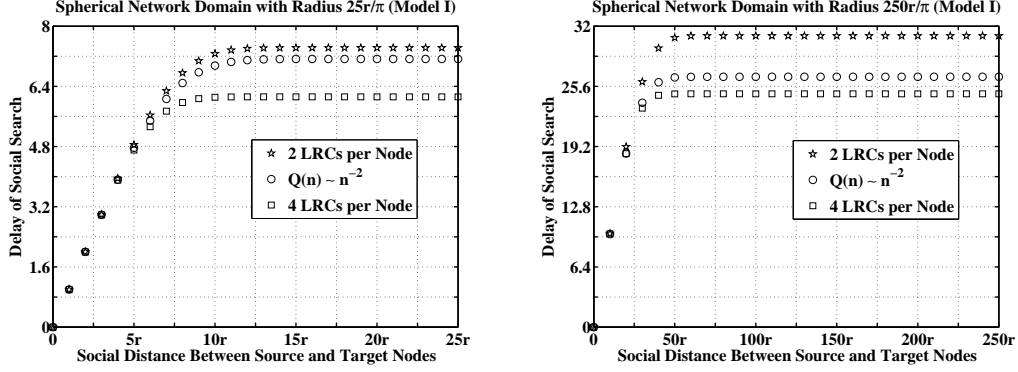
Fig. F.1. Delay of social search for scale-free networks and the networks with fixed numbers of long-range connections (LRCs) per node for spherical network domains with radius $R = \frac{25r}{\pi}$ (left figure) and $R = \frac{250r}{\pi}$ (right figure). The unit of delay is the number of steps to complete the search process.

Appendix G.

This appendix provides corresponding data for the effect of node degree heterogeneity on the delay of social search for Model I in Fig. G.1.
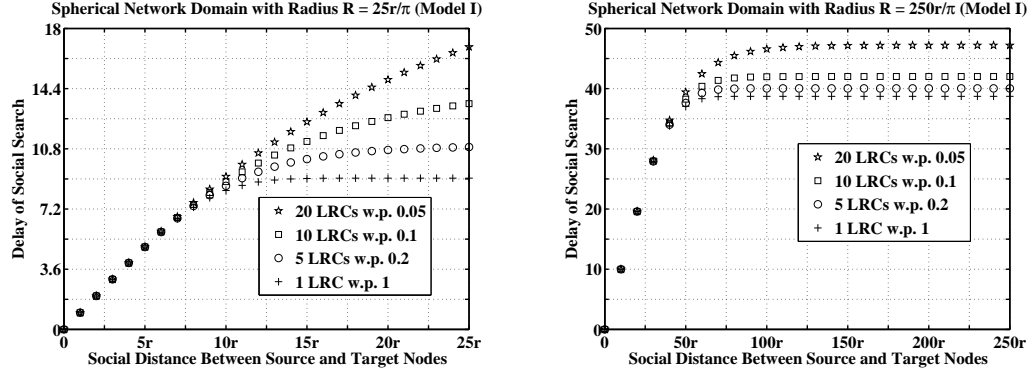

Fig. G.1. Effect of node degree heterogeneity on the delay of social search for spherical network domains with radius $R = r\frac{25r}{\pi}$ (left figure) and $R = \frac{250r}{\pi}$ (right figure). Each node has either n long-range connections (LRCs) with probability (w.p.) $\frac{1}{n}$, or 0 LRC w.p. $1 - \frac{1}{n}$. The unit of delay is the number of steps to complete the search process.

References

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

Amaral, L. A. N., Scala, A., Barthelemy, M., Stanley, H. E., 2000. Classes of small-world networks. Proceedings of the National Academy of Sciences 97, 11149-11152.

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. Proceedings of the National Academy of Sciences 99, 7275-7279.

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

Barnett, G. A., 1989. Approaches to non-Euclidean network analysis. In: Kochen, M. (Ed.), The Small World. Ablex, Norwood, NJ , USA, pp. 349-372.

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

Bell, D. C., Belli-McQueen, B., Haider A., 2007. Partner naming and forgetting: recall of network members. Social Networks 29, 279-299.

Bernard, H. R., Johnsen, E. C., Killworth, P. D., Robinson, S., 1991. Estimating the size of an average personal network and of an event population: Some empirical results. Social Science Research 20, 109-121.

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, UK.

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

Callaway, D. S., Newman, M. E. J., Strogatz, S. H., Watts, D. J., 2000. Network robustness and fragility: Percolation on random graphs. Physical Review Letters 85, 5468-5471.

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.

Chaintreau, A., Fraigniaud, P., Lebhar, E., 2008. Networks become navigable as nodes move and forget. Proceedings of the 35th International Colloquium on Automata, Languages and Programming 5125, 133-144.

Clauset, A., Moore, C., 2003. How do networks become navigable? Available online: http://arxiv.org/abs/cond-mat/0309415.

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.

Davidson, M. L., 1983. Multidimensional Scaling. Wiley Series in Probability and Mathematical Statistics, New York, NY, USA.

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.

Duchon, P., Hanusse, N., Lebhar, E., Schabanel, N., 2006. Could any graph be turned

into a small-world? Theoretical Computer Science 355, 96-103.

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.

Fraigniaud, P., Giakkoupis, G., 2009. The effect of power-law degrees on the navigability of small worlds. Proceedings of the 28th ACM Symposium on Principles of Distributed Computing, Calgary, Canada, 240-249.

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

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

Gowaikar, R., Hassibi, B., 2006. Communication over a wireless network with random connections. IEEE Transactions on Information Theory 52, 2857-2871.

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.

Gupta, P., Kumar, P. R., 2000. The capacity of wireless networks. IEEE Transactions on Information Theory 46, 388-404.

Hekmat, R., Mieghem, P. V., 2004. Study of connectivity in wireless ad hoc networks with an improved radio model. 2nd Workshop on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, 142-151.

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

Killworth, P. D., McCarty, C., Bernard, H. R., House, M., 2006. Social networks: the accuracy of small world chains in sociological networks. Social Networks 28, 85-96.

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.

Kossinets, G., Watts, D. J., 2006. Empirical analysis of an evolving social network. Science 311, 88-90.

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

Leskovec, J., Horvitz, E., 2008. Planetary-scale views on a large instant-messaging net-

work. Proc. 17th International World Wide Web Conference.

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, USA, 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, 15-22.

Manku, G. S., Bawa, M., Raghavan, P., 2003. Symphony: Distributed hashing in a small world. In Proceedings of the 4th USENIX Symposium on Internet Technologies and Systems,127-140.

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.

Moody, J., 2002. The importance of relationship timing for diffusion. Social Forces 81, 25-56.

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. Proceedings of the National Academia of Sciences 98, 404-409.

Newman, M. E. J., 2003a. The structure and function of complex networks. SIAM Review 45, 167-256.

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

Nguyen, V., Martel, C. U., 2006. Designing low cost networks with short routes and low congestion. Proceedings of the 25th IEEE International Conference on Computer Communications, Barcelona, Spain, 1-12.

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

Nowell, D. L., Novak, J., Kumar, R., Raghavan, P., Tomkins, A., 2005. Geographic routing in social networks. Proceedings of the National Academia of Sciences 102, 11623-11628.

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. 1963. Mathematical models of social interaction. In: Luce, R. D., Bush, R. R., Galanter, E. (Eds.), Handbook of Mathematical Psychology. John Wiley and Sons, New York, NY, USA, pp. 493-579.

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

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

Schenettler, S., 2009a. A structured overview of 50 years of small-world research. Social Networks 31, 165-178.

Schenettler, S., 2009b. A small world on feet of clay? A comparison of empirical small-world studies against best-practice criteria. Social Networks 31, 179-189.

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., 1999. Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton University Press, Princeton, NJ, USA.

Watts, D. J., Dodds, P. S., Newman, M. E. J., 2002. Identity and search in social networks. Science 296, 1302-1304.

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