Small-worlds: Beyond Social Networking

Andrew R. Curtis
University of Wyoming, arcurtis@gmail.com

Small-worlds: Beyond social networking

Andrew R. Curtis[*]

Abstract

Small-world phenomena were initially studied in the 1960s through a series of social network experiments, and are, as evidenced by the game "The six degrees of Kevin Bacon," even part of our pop-culture. Recently, mathematicians and physicists have shown that most small-world phenomena are expected consequences of the mathematical properties of certain networks– known as *small-world networks*. In this paper, we survey some recent mathematical developments dealing with small-world networks, as well as present a new small-world network model and discuss new ideas for decentralized searching. The goal is to give the reader a sense of the importance of small-world networks, and some of the useful applications dealing with these networks.

1 Orgins

In order to create an appropriate mathematical model of small-world phenomena, it is important to understand the social network origination of this branch of mathematics.

We've all had experiences (e.g. sitting next to someone on an airplane who went to high school with someone from your Uncle's hometown) that made us say "it really is a small-world." Such a "coincidence" is an example of a small-world phenomenon. Sociologists were the first to study *small-world phenomena* when they tried to explain everyday observation that any two individuals in a social network are likely to be connected through a short chain of acquaintances. The study consisted of a series of experiments by Stanley Milgram and colleagues in the 1960s [10, 17, 19]. A typical run of the experiment delivered a letter to a *source* person in Nebraska, with a *target* in Massachusetts. The source was told relatively little information about the target (including his name and address), with instructions to send the letter to a person he knew on a first name basis whom he thought would know the target. The next person in the chain received the letter and the same instructions. Thus the letter hopped across the country in this manner, eventually reaching the target person. Many trials showed that an average of 6 steps were necessary for the letter to pass from the source to the target. This is now known as the "six degrees of separation" principle [12].

The existence of short paths between two persons is striking given the fact that there are over 260 million people in the United States alone, and that a person's close acquaintances number in the hundreds. If friendships were random (i.e. it's just as likely that you know the governor of Massachusetts, as you know your father), then one might expect short paths to exist. Indeed a mathematical theory, known as random graph theory, shows that the *diameter*, or the expected distance between a source and target, in a random network is relatively small [7]. If friendships were highly structured (i.e you are only friends to those who live within 10 miles of you), then we would expect it to take much more than 6 steps for a letter from Nebraska to reach Massachusetts.

Social networks are neither random nor highly structured. Two friends are likely to have a third, common friend, and social networks take geographic information into account. Two persons living thousands of miles away are not likely, but occasionally are, friends. Thus, social networks are a mix between random and highly structured networks. So why do short paths exist in social networks?

Even more startling is that the participants (working independently) in Milgram's experiments *found* short paths. The mere existence of short paths connecting persons through acquaintances pales in comparison to the ability for these short paths to naturally manifest themselves through individual choices. This suggests that the network stores latent structural cues, guiding members to a short path. This is startling due the lack of knowledge individuals have about the network. People know their own friends, and possibly have knowledge of their friends' friends. Yet no one has complete knowledge of the chain of individuals between themselves and an arbitrary target.

2 Modeling the Phenomenon

The natural question arising from Milgram's experiments is how do we mathematically model small-world networks? Several authors have investigated this question. We present their models here, as well as our own geographic model. However before examining small-world models, we discuss random and structured networks. The first small-world model examined is the original model of Watts and Strogatz. Next, we will examine networks with power-law distribution of node degrees. And finally, we will present our own small-world network model.

2.1 Random and Structured Networks

Before discussing models of small-world networks, we feel it is necessary to cover random and structured networks. A *network* is a collection of *vertices* or *nodes* connected by *edges*. In mathematics, a network is typically referred to as a *graph*. A graph $G = (V, E)$ consists of a (typically, finite) set of vertices V and a subset E of pairs of vertices. Small networks are sometimes best understood visually, see Figure 1 for an example.
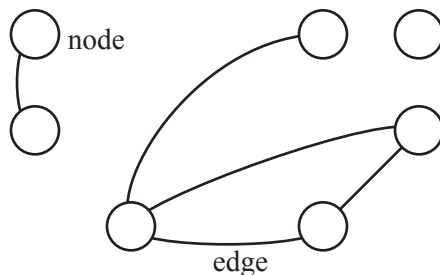

Figure 1: Example network with seven nodes and five edges.

Random networks are created by fixing a real number p with $0 \leq p \leq 1$, then creating a set of vertices V, and finally forming connections between any two vertices with probability p. Therefore the probability that u is connected to v is the same for each $v \in V$. The expected diameter of a random networks is low; that is, the longest path from any vertex to any other vertex is relatively small [7]. But random graphs have a very low clustering coefficient. This means that if an edge exists between two vertices u and v, then the probability that an edge exists from u to a neighbor of v is low ($(1 - p)^{p(n-1)}$ where n is the number of vertices).

Structured networks are created as the name suggests, by forming connections between nodes in a structured manner. For instance, one way to create a structured network is to place a collection of n vertices on a line. Then, connect each vertex to the vertex closest on either side. This will form a path of n vertices. Structured networks which only join close vertices together, will have high clustering coefficient, but relatively large diameter. In our line example, the diameter of the network is $n-1$, since $n-1$ steps are required to send a message from the vertices on either end of the line to each other. Another simple method to build a structured network is to place n^2 nodes on an $n \times n$ lattice, then connect each node to the surrounding four nodes.

2.2 Basic Model

The small-world phenomenon was first modeled by Watts and Strogatz in their 1998 paper "Collective dynamics of small-world networks" [21]. Watts and Strogatz give a simple, yet novel method to model several naturally arising and man-made small-world networks. Their model places a set V of n network nodes in a circle, then connects the nearest k nodes together. For $k=1$, this yields a cycle, each node is able to directly communicate with its closest two neighbors, one on either side. Then one adds a few connections between nodes in V at random. These edges serve to add randomness to a structured network. This model creates networks with low diameter like random networks, yet the nodes remain "clustered." Each node in V has "local contacts," eg. a neighbor and long-range "weak links," eg. a friend who has moved to another country. An example of a network created with Watts and Strogaz's model is drawn in Figure 2.

In addition to this model, Watts and Strogatz showed three naturally arising networks exhibit the small-world phenomenon; that is, these networks have a low diameter and a high degree of clustering. The networks examined were the neural network of the worm *Caenorhabditis elegans*, the power grid of the United States, and the collaboration graph of film actors.

2.3 Power-Law Networks

Not all small-world networks are the same, however. Many networks with the small-world property also have a *power-law* property, the distributions of the degrees of each node follow a power-law. This means that a chart of how many connections each node in the network has looks like graphing $y = \log_\gamma x$ on the (x, y) plane, for some positive number γ. More precisely, the probability $P(k)$ a node will have connections to k other nodes follows $P(k) \sim k^{-\gamma}$. A model for generating power-law networks is given by Barabási and Albert [5]. The model of Barabási and Albert creates a small, random network then adds more nodes, forming connections between a new node and existing nodes preferentially to nodes with higher degrees. A small power-law network is drawn in Figure 3.

This "rich-get-richer" network design paradigm is a common theme in many networks. For instance, the link structure of the World Wide Web (WWW) is a power-law network. Let's say a person creates a personal web page for himself. He decides to put a few links to some of his favorite pages on his personal web page. Most likely, this person will add links to popular web pages. If a web site is already popular, it is more likely to receive another link. This uneven distribution of links makes the structure of links between pages in the WWW a power-law network.

The model of Barabási and Albert begins by creating a small random network $G = (V, E)$ of m_0 nodes. Then additional nodes are added to the network, forming connections preferentially to nodes with high connectivity. The probability Π that a new vertex u will be connected to an existing vertex v is

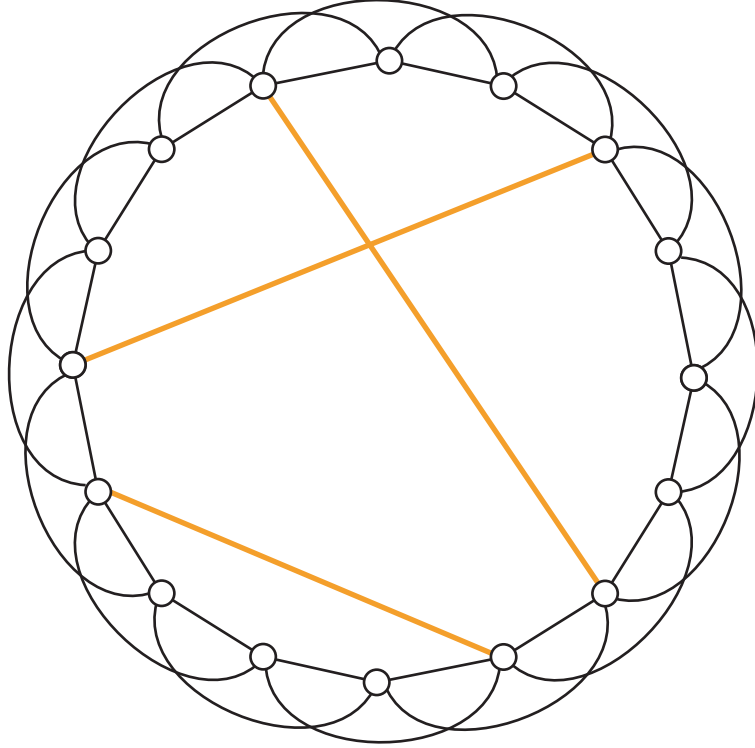$$\Pi(u, v) = \frac{deg(v)}{\sum_{w \in V} deg(w)}$$

3

Figure 2: Watts and Strogatz example where $k = 2$. The random edges have been colored orange.
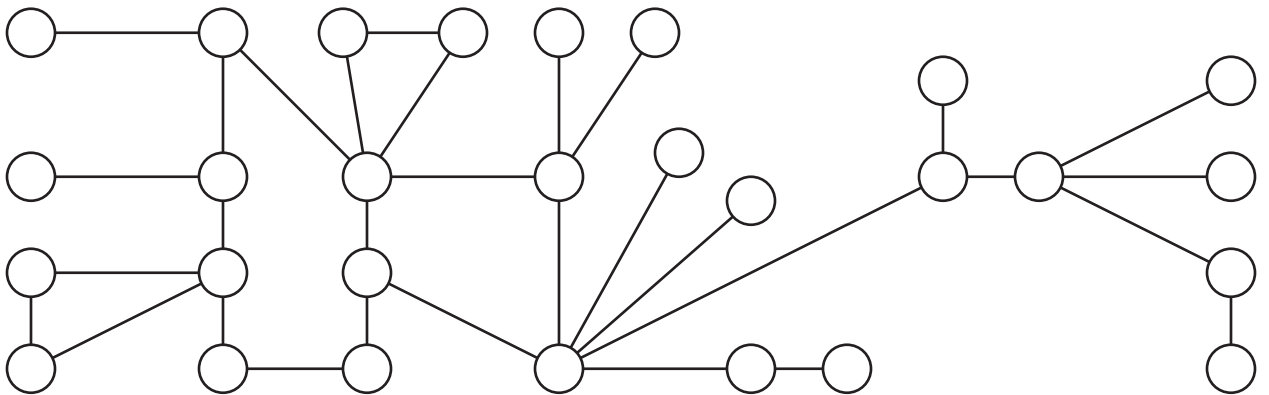

Figure 3: Power-law example. Note the uneven degree distributions, some nodes have many connections, while most have few.

so after t nodes are added, $|V| = m_0 + t$. This network will scale from a small, random network to a power-law network with $\gamma = 2.9 \pm 0.1$.

We mentioned the WWW forms a power-law network, and much work has been done to model the link structure of the WWW as a graph. This is done by allowing each online document to be a vertex and each link from document u to document v to form a directed edge (u, v). Many authors have modeled the degree distribution of nodes in the WWW [5, 6, 11, 13, 14]. In addition, authors have estimated the diameter of the WWW [3]. The WWW is a particularly interesting example since anyone can participate in its creation and its structure is constantly changing. Despite the billions of pages in the WWW, the diameter has been estimated to be 19 by Albert, Jeong and Barabási [3]. This means that if one was to start at a random page, they should be able to reach any other page on the WWW by clicking at most 19 links. This is possible simply because of the power-law structure of the WWW.

2.4 Geometric Networks

We now present our geometric model for constructing small-world networks. This model incorporates features of other models, but remains flexible. So flexible that we offer methods to modify our model to create networks with a power-law degree distribution.

Our model uses geometric information to form connections between nodes. To begin, we place n points randomly on an $n \times n$ plane. Then we form connections between nodes u and v that are "close" with probability

$$P_{close}(u, v) = \frac{1}{k_{close}}$$

with k_{close} a small constant. We then add the notion of a "far" contact. We add a connection between u and v with probability

$$P_{far}(u, v) = \frac{1}{k_{far}}$$

where k_{far} is a constant dependent on n and $1 \leq k_{close} < k_{far} \leq n$. The notion of close and far depend on n and how many connections per node one would like. We use a discrete relationship for close and far. A node v is close to u if the distance $d(u, v)$ is less than a positive real number m. This distance is simply the Euclidean distance between the points representing u and v on the plane. This relationship is described in Figure 4. Any node v close to u is then connected with probability $P_{close}(u, v)$, and each node not close to v is connected with probability $P_{far}(u, v)$. So for higher clustering, lower k_{close}; and for more random connections, lower k_{far}. This model naturally extends to higher dimensions as well.

The resulting network has a low diameter due to the random, long-distance connections between "far" nodes, while maintaining a high clustering coefficient because of the "close" connections. The geometric implications of this network are also easily understood. A person is much more likely to have many acquaintances near her home, however not all of her neighbors are necessarily her acquaintances as in the model of Watts and Strogatz. This person is also likely to have some acquaintances who geometrically live far away. However she is likely to have fewer acquaintances who live hundreds or thousands of miles away then those who live close by. This geometric model extends past social networks, however. When a person hooks a computer up to the Internet, it is desirable to keep the physical cable connecting that computer to the Internet as short as possible for expense reasons. Say one was to add a server to the Internet, and for robustness would like to connect to four other servers. It makes sense to run cables to the three physically closest servers and then one additional cable to a major, centrally located server. Our model works to encapsulate this behavior.
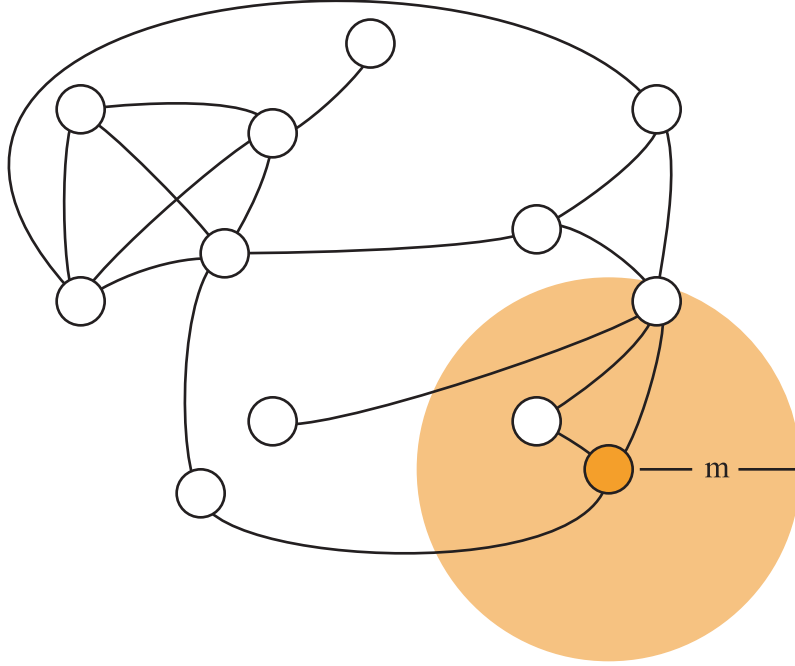
Figure 4: Geometric network. The highlighted area around the orange node represents the zone of close nodes.

One interesting aspect of our geometric model is that it can be modified to create power-law networks. Rather than placing the nodes randomly on a plane, arrange n nodes on a lattice. We then modify the probability of forming an edge between two nodes. Instead of taking the Euclidean distance between each nodes, we use the distance between a node and the center of the lattice. Then the probability P an edge will be created between nodes u and v is

$$P(u, v) = \frac{1}{(d_{center}(v))^r}$$

where r is a constant that can be modified to alter γ, the "slope" of degree distributions. For instance, when $\gamma = 2$, the degrees distribution has a much steeper falloff then when $\gamma = 1/2$. Thus the geometric method of constructing power-law networks is quite flexible.

3 Search Applications

In recent years, there has been considerably effort to develop local or decentralized search strategies for networks. This method of searching has very little information at each step, knowing almost nothing of the global topology of the network. We first cover the decentralized search model of Kleinberg. Next we discuss search in power-law networks. And finally, we present our own decentralized search model.

3.1 Decentralized Search

Recently the special properties of small-world networks have been exploited to find efficient search strategies. The algorithmic aspect of small-world networks was examined by Jon Kleinberg when he

showed an important theoretical result about decentralized search in small-world networks [15, 16]. Kleinberg's basic results show how to build the long-range connections in a lattice network to minimize delivery time of a message. He showed that a network constructed by placing nodes on an $n \times n$ grid, then connecting each node to its four neighbors can have a logarithmic decentralized search algorithm if a single long-range connection is added in a special manner to each node. Kleinberg's algorithm for sending a message across the network is very simple. A message starts at node s, then moves to a node connected to s such that the distance to the target t is minimized. Without long range contacts, it is easy to see that the expected delivery time of a message in an $n \times n$ lattice network is $\Theta(n)$. Kleinberg showed how to add a single long-range contact to each node to decrease this bound to a logarithmic one.

To examine Kleinberg's results further, we need to formalize the notion of lattice distance. We begin with a set of lattice points on an $n \times n$ square, $V = \{(i,j) : i \in \{1, 2, ..., n\}, j \in \{1, 2, ..., n\}\}$. Then define the *lattice distance* between nodes (i,j) and (l,k) as $d((i,j),(l,k)) = |k-i| + |l-j|$. This is simply the number of steps on the lattice separating the two nodes and can be thought of as "taxicab geometry." Then to construct the lattice network, each node will have an edge to each other node distance one away. That is, if $d((i,j),(l,k)) = 1$, then (i,j) and (l,k) are connected. Now, a single, directed long-range contact is created from u to v with probability proportional to $d(u,v)^{-r}$. This must be normalized by dividing by the sum of the distances from u to each other node in the graph. So, we add an edge from u to v with probability

$$Pr(u,v) = \frac{d(u,v)^{-r}}{\sum_{w \in V} d(u,w)^{-r}}$$

So, if $r = 0$, we have a uniform random distribution of long-range contacts across the grid. However, Kleinberg showed that only one value of r builds the long-range contacts in an optimal manner for sending information through the network.
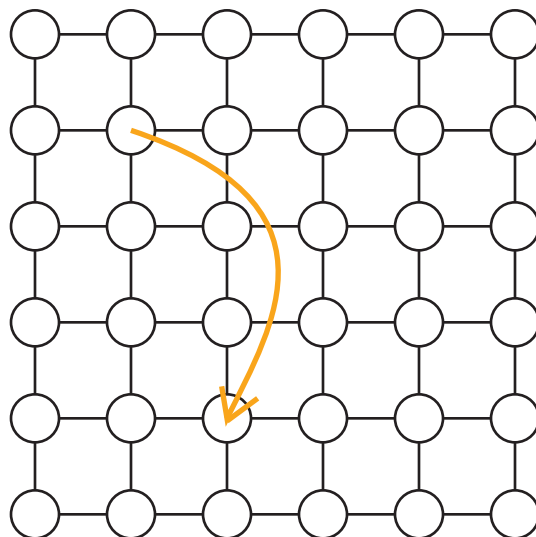
Figure 5: Kleinberg's small-world model. Each node is connected to its neighbors distance one away; the highlighted connection is a long-range contact.

Before stating Kleinberg's main results, we need to discuss how a message will move across the network. We wish to send a message in a decentralized manner, with the current message holder

knowing as little global knowledge as possible. So we will allow the node u currently holding the message to have knowledge of

1. the set of local contacts among all nodes; and

2. the location of the target node t.

With this minimal knowledge, the message will move from node to node, always minimizing the distance to the target t (eg. the message will not move *away* from the target, only towards it).

We are now ready to present Kleinberg's main result.

Theorem 1 (Kleinberg 2000) *Using the decentralized algorithm described above, there is a constant α, independent of n, so that when $r = 2$, the expected delivery time of a message is at most $\alpha(\log n)^2$.*

The proof of Theorem 1 is too complicated to present here. We will comment, however, that Kleinberg's result is especially interesting, since $r = 2$ is the *only* value for which the decentralized algorithm will deliver a message in logarithmic time. This shows that long-range contacts need to be arranged in a very specific manner. When $r < 2$, the long-range contacts of a node u are too far away to be useful, and when $r > 2$ the long-range contacts of u are too close. This result can be extended to arbitrary dimensions, so if we have a k-dimensional lattice of nodes, the optimal delivery time will be reached when $r = k$.

3.2 Power-Law Search

While Kleinberg's results are for a very idealized network model, they can be applied to increase the search speed in real world networks. Authors have applied Kleinberg's search algorithms to peer-to-peer (P2P) networks such as Gnutella [2, 9] and Freenet [8, 22]. In addition to showing how the small-world model can speed up search in the Gnutella P2P network, Adamic et al. give local search algorithms to find a message in sublinear time [1, 2]. The search strategy Adamic et al. give will visit an entire power-law network with n nodes in approximately $\ln^2(n)$ steps. This result is quite promising, since many other networks (eg. the WWW) that one would like to search also have a power-law degree distribution.

3.3 Searching with a Modified Degree

The search strategies in [2] give sublinear search methods for power-law networks, however the current message holder u must have knowledge of its neighbors as well as its neighbors' neighbors. We have investigated a search strategy that doesn't require as much information at each step, but is still useful for non-ideal networks.

To do this, we propose a new idea for measuring the degree of a vertex u. The standard definition for the degree of a node is the number of edges that include u. In [2], the search strategy sent a message to nodes with continuously higher degrees, until the message had reached the node of maximum degree. Our searching strategy modifies the notion of degree, so that the degree of a vertex is dependent on the position of the message in the network.

Our notion of degree incorporates Euclidean geometry, so we begin with n nodes on a plane. When node u has the message, only nodes adjacent to u will be assigned a degree. And then, only those that are in the direction of the target node t will be assigned a degree. More precisely, define $d(w, t)$ to be the Euclidean distance between nodes w and t. Then if $d(u, t) < d(w, t)$, w will not be given a degree. This means that we only consider nodes closer to the target than the current

message holder. Then the degree of w adjacent to u with $d(w,t) < d(u,t)$ is the number of nodes adjacent to w *closer to t than w*. Stated simply, the degree of w is the number of nodes connected to w that could move the message closer to the target t. This notion of degree is shown in Figure 6. If no nodes adjacent to u are closer to t than u, the message will be sent to the node of maximum degree connected to u using the traditional notion of degree.
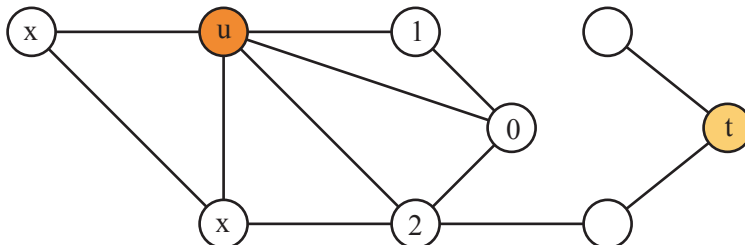

Figure 6: Degrees of nodes connected to u for target t. In this case, the message would be sent to the node with degree two, since that is maximal for nodes adjacent to u.

While this method of searching gives more information to the current message holder than Kleinberg's model, the assumption that a message holder would know *something* of its friends' friends is reasonable. Consider the original experiments of Milgram. He asked the message holder at any step to send the message to someone they thought would know the target person. Even knowledge of where the target person lives helps the message holder decide who to send the letter to next. Milgram's original target lived in Massachusetts, so it's reasonable that any person holding the message would send it to someone they knew had connections in Massachusetts. The person holding the message may not know her friends' friends, but it's likely she knows something about her friends that would help her decide where to send the letter next. This knowledge could be geographic knowledge, or even knowledge of the target's career. If the target is a stock broker, it makes sense for the current message holder to send the letter to a friend who is also a stock broker. Our search model works to incorporate this limited knowledge. Our model sends the message to a node with maximum knowledge of nodes related to the target, without actual knowledge of nodes more than one step away.

We conjecture that an algorithm based off our search methods would deliver a message from a source s to a target t in sublinear time in simple power-law network model. To deliver a message quickly with our search method, we need to construct a power-law network with a geometrically even spacing of nodes with many connections. From our empirical results, we theorize that our search method can deliver a message in logarithmic time for a network with few connections per node on average, however we have not been able to prove this yet.

4 Summary

In this paper, we have shown some recent developments concerning the small-world phenomenon, presented new small-world models, and gave ideas to improve search in idealized networks. Our geometric model improves upon existing models, while retaining a flexible average node degree and an adaptable clustering coefficient. We believe our search ideas help to incorporate many of the structural cues present in a social network.

While much of the work done with small-world networks has been motivated by empirical studies,

much work remains to be done to explain these empirical results. The concept of the "small world" originated in the social sciences, but has grown to an active area for mathematical research. Much work is yet to be done on improved network models. Even more, decentralized search is an area ripe for scientific investigation. For more on small-world networks, we recommend [4, 18, 20].

5   Acknowledgments

References

[1] L. A. Adamic, R. M. Lukose, and B. A. Huberman, "Local search in unstructured networks," in S. Bornholdt and H. G. Schuster (eds.), *Handbook of Graphs and Networks*, (Wiley-VCH, Berlin, 2003).

[2] L. A. Adamic, R. M. Lukose, A. R. Puniyanai, and B. A. Huberman, "Search in power-law networks," *Physical Review E*, 64 (2001).

[3] R. Albert, H. Jeong, and A. L. Barabási, "Diameter of the world-wide web," *Nature* 401 (1999).

[4] A. L. Barabási, *Linked: The New Science of Networks*, (Perseus, Cambridge, MA, 2002).

[5] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science* 286 (1999).

[6] A. L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: The topology of the World Wide Web", *Physica A* 281 (2000).

[7] B. Bollobás, *Random Graphs* (Academic Press, London, 1985).

[8] I. Clarke, O. Sandberg, B. Wiley, and T. W. Hong, "Freenet: A distributed anonymous information storage and retrieval system in designing privacy enhancing technologies," *International Workshop on Design Issues in Anonymity and Unobservability*, LNCS 2009 (2001).

[9] Clip2 Company, Gnutella. http://www.clip2.com/gnutella.html

[10] C. Korte and S. Milgram, "Acquaintance networks between racial groups: Application of the small world method," *Journal of Personality and Social Psychology*, 15 (1978).

[11] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the Internet topology," *ACM SIGCOMM* (1999).

[12] J. Guare, *Six Degrees of Separation: A Play* (Vintage Books, New York, 1990).

[13] B. A. Huberman, *The Laws of the Web*, (MIT Press, Cambridge, MA, 2001).

[14] J. Kaiser, Ed., "It's a small Web after all," *Science* 285 (1999).

[15] J. Kleinberg, "The small-world phenomenon: An algorithmic perspective," *Proc. 32nd ACM Symposium on Theory of Computing*, (2000).

[16] J. Kleinberg, "Navigation in a Small World," *Nature* 406 (2000).

[17] S. Milgram, "The small world problem," *Psychology Today* 1 (1967).

[18] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, 45 (2003).

[19] J. Travers and S. Milgram, "An experimental study of the small world problem," *Sociometry* 32 (1969).

[20] D. J. Watts, *Small Worlds*, (Princeton University Press, Princeton, 1999).

[21] D. Watts, S. Strogatz, "Collective dynamics of small-world networks," *Nature* 393 (1998).

[22] H. Zhang, A. Goel, and R. Govindan, "Using the small-world model to improve freenet performance," *IEEE Infocom* (2002).