Third Year Essay Project
SOCIAL NETWORK ANALYSIS: THE SMALL-WORLD PHENOMENON

Si Wu
00829371
18/01/2016

Abstract: Since the 1960s, psychologists, physicists, mathematicians and computer scientists have developed an interest in "the small-world phenomenon". What "the small-world phenomenon" essentially means is, anyone in the world is six or fewer steps away from you. This leads to the concept of "six degrees of separation". This essay discusses three landmark papers in the history by Stanley Milgram (1967), Watts-Stogatz (1998) and Jon Kleinberg (2001) and the conclusions drawn from these models.

Introduction: Networks can be used to represent brain neurons, electric power grids, collaboration graph of film actors, etc. The small-world phenomenon is a common phenomenon arising in a range of networks in society, nature and technology. It is also crucial in the evolution of the world wide web, including popular websites such as Facebook, LinkedIn and Twitter etc. The understanding of networks also helps us to study phenomena such as violence, social epidemics and biological diseases.

In this essay, we start from Milgram's experiment, set up a mathematical model with local and long-range links based on his experiment (Watts-Stogatz model), predict that the clustering exponent controls the "randomness" of the long-range links based on the model (Jon Kleinberg), which can be proven with real data from LiveJournal and Facebook. This is how we will expect the procedures of experiments, theories and measurements to display. [6]

Milgram's small-world experiment: Stanley Milgram (1933-1984), a psychologist at Harvard University, conducted his famous small-world experiment. He was one of the first people to come up with the concept of "six degrees of separation" (although he did not use this phrase himself). He started with one question: what is the probability of any two people in the world knowing each other? He gathered a group of about 300 volunteers around Nebraska and Kansas and asked them to send a letter to a person in Boston. 64 of the letters reached the target. [1] The initial success rate was low, but after some modifications the average number of intermediates in a successful chain was found to be 5 over many trials. That is to say, there is at most six degree of separation between any two people in the world. This was stated in the famous article "The Small World Problem" by Milgram published on Psychology Today in 1967.

His experiment demonstrated two interesting facts about social networks. First, it stated that short paths do exist in abundance. [2] Path is defined as a sequence of nodes with each consecutive pair in the sequence being connected by an edge. [3] The nodes here, in social network science, represent people and the edges represent the interactions between people. So the "short paths" here just means short chains of friends or people. This property is what we mean by saying a network is a small world. [3] Second, the experiment demonstrated that these short paths can be found without any kind of global "map" of the network, i.e., using only local information of the network. [2]

Watts and Stogatz's model: Small-world networks are highly clustered, and have short path lengths. [4] This means that the nodes are connected such that any node can be reached from any other in an average of only a few steps. [4] So can we come up with a model with both of the two features? In 1998, Duncan Watts and Steve Strogatz proposed with such a model in their joint paper in *Nature*. Duncan Watts is a Principal Research Scientist at Yahoo! Research.

He also teaches sociology at Columbia University in New York. [5] Steve Strogatz is a professor of Applied Mathematics in the Department of Theoretical and Applied Mathematics at Cornell University. Their research was first inspired by the synchronization of cricket chirps. [5] They thought that such a model has two basic social-network properties: homophily and weak ties. Homophily is the tendency of individuals to connect with people who are like ourselves. Weak ties connect us to acquaintances that are in the parts of the network that are far away from us, i.e. when people have more distant and less strongly connected acquaintances. Homophily is the reason that many clusters of people are formed in the network, while the weak ties create the widely branching structure that reach many nodes with only a few intermediates. [6] In the Watts-Strogatz model, they first assumed that everyone lives on a two-dimensional grid. The grid can be a model of geographic or social proximity, or any other similar concept that can lead to the formation of links (figure 1-a). If two nodes are right next to each other we say that they are one grip step apart. They then gave the network two types of links: one due to homophily and the other due to weak ties. As shown in figure 1-b, the regular structure represents homophily and the random links represent the weak ties. Watts and Strogatz argued that, if we start at a single node v, with each node having k random links, it is very unlikely for us to see the same node twice in the first few steps. That is because the weak ties out from the nodes are chosen uniformly at random, which results in us reaching a large number of nodes in a small number of steps. [6] This is how the hybrid networks can lead to short paths. Surprisingly, we only need an even smaller amount of randomness to have the same effect. As seen in figure 1-c, instead of having each person having k random friends that are far away, we have now one person with one random friend out of k people. Even with this type of network, we can find shortest paths between all pairs of nodes. Here, then, is the key conclusion of the Watts-Strogatz model: only a small amount of randomness is needed to make the world "small", that is to say, there is short path between every pair of nodes.

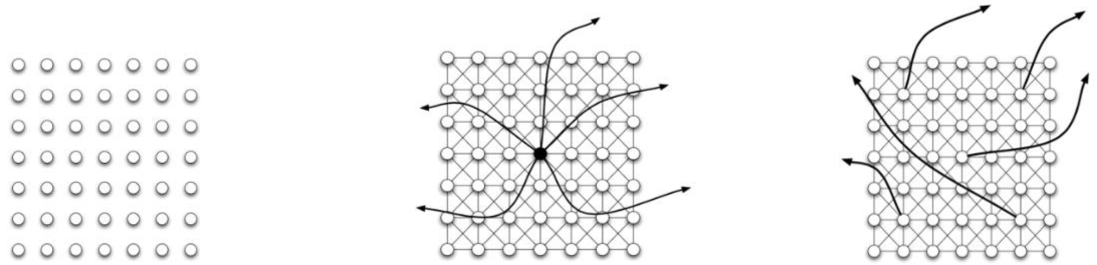
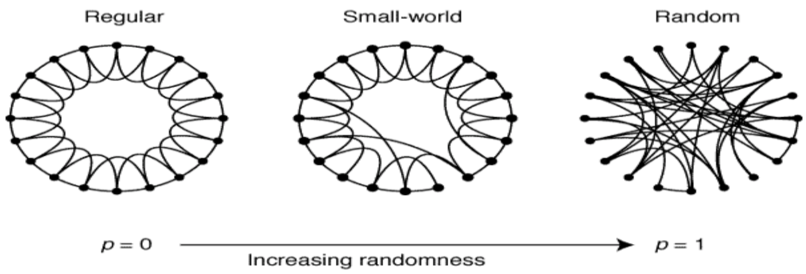


Figure 1-a　　　　Figure 1-b　　　　Figure 1-c

Figure 1: Watts and Strogatz's grid-based model [6]

Their model is well known as the beta model after being published in Watts' book *Six Degrees: The Science of a Connected Age.* This was also explained in their famous joint paper in *Nature*. They argued that there are two extreme cases of networks: completely regular or completely random. Most biological, technological and social networks are somewhere in between the two extreme cases. Small-world networks, which lie in between the two extremes, are highly clustered like the regular lattices (or say have a high clustering coefficient) and also have small characteristic path lengths like the random graphs, so that any node can be reached from any other node in a few steps. [4] This is illustrated in figure 2. Some examples of small-world networks include neural networks, electric power grids and the collaboration graph of film actors. [7]

Figure 2 shows the process of rewiring edges at random, with an increasing amount of disorder going from the regular lattice (left) to the random networks (right). They assumed that the ring lattice in figure 2 has n vertices and k edges per vertex. And they assigned each random edge a probability p. That is the reason why the regular lattice has p = 0 and the random lattice has p = 1. They also defined two quantities: characteristic path length (L) and clustering coefficient (C). Characteristic path length is defined as the number of edges of the shortest path between two nodes, averaged over all pairs of nodes of the whole network. For the clustering coefficient, suppose that a vertex v has k_v nearest neighbours. Then the highest number of edges that can exist between all those neighbours of v is $\frac{k_v\,(k_v-1)}{2}$. This happens when every neighbour of v is connected to every other neighbour of v. If C_v is the fraction of all these edges out of all edges, then the clustering coefficient C is an



C	high	high	low
L	long	short	short

Figure 2: Watts and Strogatz's beta model [4]

average of all C_v. Characteristic path length measures the separation between two nodes (a global property), and clustering coefficient measures the cliquishness of a neighbourhood (a local property). For example, in a friendship network, the meanings of these terms can be listed as follows: L reflects the average number of friendships in the shortest chain between two people; C_v reflects how many friends of v are also friends with each other; therefore C measures the cliquishness of a friendship cycle. We are interested in networks that have vertices with sparse connections, but not too sparse for the graph to be disconnected. [7] We need $n \gg k \gg \ln(n) \gg 1$. $k \gg \ln(n)$ guarantees that a random graph will be connected.[10] We find that as $p \to 0$, $L \sim n/2k \gg 1$, $C \sim 3/4$. As $p \to 1$, $L \approx L_{random} \sim \ln(n)/\ln(k)$ and $C \approx C_{random} \sim k/n \ll 1$. So at p = 0, the regular lattice is highly clustered, L grows linearly with n where it represents a large world. At p = 1, the random network is badly clustered, L only grows logarithmically with n where it represents a small world. One of the main results for this model is that the small-world network is highly clustered like a regular lattice and has a small characteristic path length like a random network. [7]

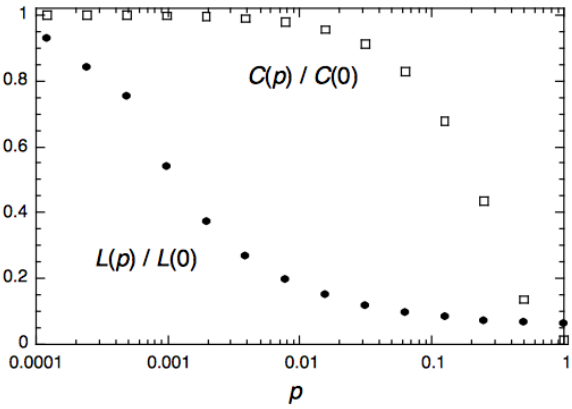


Figure 3: The behaviour of L(p) and C(p) [7]

In figure 3, L(p) is the characteristic path length, C(p) is the clustering coefficient. In this graph, the data are averaged over 20 random rewiring processes in figure 2 in this essay. They were normalized by L(0), C(0) for a regular lattice. There are n = 1000 vertices and an average of k = 10 edges per vertex for all the graphs. There is a logarithmic scale on the horizontal axis in order to resolve the rapid drop in L(p). This drop is due to the small-world phenomenon. These small-world networks result from the immediate drop in L(p) are caused by a few long-range edges. These long-range edges (or "short cuts") are the same as the "weak ties" shown in figure 1-b or 1-c. They connect vertices that would otherwise be much further apart than L_{random} (characteristic path length for a random network). At small p, each of these long-range edges has a highly nonlinear effect on L. It contracts the distance between the two vertices it connects, and their nearest neighbours, neighbours of neighbours and so on. While for C(p), a long-range edge from a clustered neighbourhood only has at most a linear effect on C(p). So as we can see from figure 3, for small p, even if L(p) drops rapidly, C(p) almost remains unchanged. Then we can conclude that at the local level (reflected by C(p) because it is the local property), we cannot detect the transition to a small world. But we must remember that the rewired edges must connect vertices that would otherwise be much further apart than L_{random}. [7]

The above analysis shows the important effect of the long-range edges. It tells us with only a small amount of short cuts needed, the small-world phenomenon appears commonly in sparse networks with many vertices. In order to test that idea, Watts and Strogatz calculated L and C based on three real networks: the collaboration graph of film actors, the neural network of the nematode worm C.elegans, and the electric power grid of the western United States. [11] The graph of film actors is similar to a social network. C.elegans is an example of a neural network and the graph of a power grid shows the efficiency and robustness of power networks. [7]

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Table 1: L and C values for three different networks

As shown in table 1, L $\gtrsim$ L_{random} but C $\gg$ C_{random} for all three networks, illustrating the small-world phenomenon.

They then investigated the effect of small-world phenomenon for the spreading of an infectious disease. This is an example of a dynamical system. The structure of population is modelled by the graphs in figure 2. When t = 0, an infected person is introduced into a group of healthy people. After a period of time, the infective individuals are removed permanently, either by immunity or death. We give each of these infective individuals a probability of infecting its healthy neighbours r. The disease either ends up infecting the whole population along the edges of the graph, or dies out, only infecting some of the population.

We have two results. First, as you can see from figure 4-a, the critical infectiousness r_{half}, at which the disease infects half of the population, decreases significantly at small p (note the logarithmic scale on the x-axis). This shows that the disease spreads much more easily in a small world because the lower the critical infectiousness is, the easier the disease will spread. The second result was, for a disease that can easily infect the entire population no matter what the structure is like, the time T(p) required for global infection matches the L(p) curve. This is shown in the right graph in figure 4. This tells us that the disease spreads much more quickly in a small world. The important and less obvious point is how few short cuts are needed to make the world small. Even if only a few percent of edges in the original lattice are randomly rewired, the time to global infection is nearly as short as a random graph. [12]

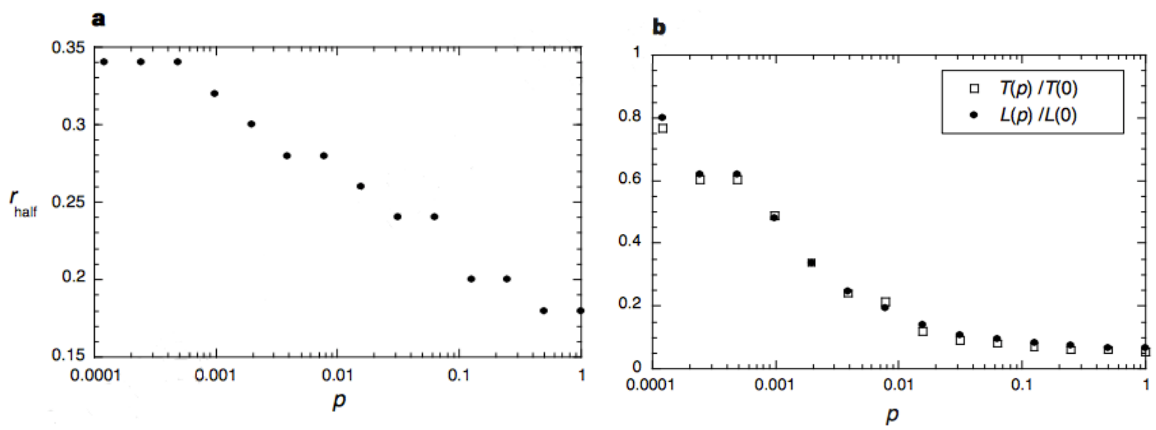


Figure 4: The results for dynamical systems [7]

Unlike all the other models of disease spreading, which all show that the network structure affects the speed and extent of disease transmission, Watts-Strogatz model indicates that the dynamics is an explicit function of structure (figure 4). [13] Kretschmar and Morris also showed that the increase in the number of sexual partners can rapidly increase the propagation of a sexually-transmitted disease that spreads along the edges of a graph. [8] Duncan and Steve also examined the functional significance of small-world connectivity on three more dynamical systems: cellular automata, the iterated, multi-player 'Prisoner's dilemma' and the coupled phase oscillators. [14] [15] [16]

Watts-Strogatz model inspired people to know more about small-world networks. Although people may not have paid much attention to small-world architecture at the time, Duncan and Steve suggested that it would turn out to be common in biological, social and technological networks, as well as some dynamical systems. [7]

Kleinberg's model: Jon Kleinberg is a professor of Computer Science at Cornell University. He was one of the first people to realize Stanley Milgram's small-world experiment implied not only that short paths exist between people, but also people seem to be good at finding these paths (the second basic aspect of the small-world experiment). He asked a question: Watts and Strogatz's research only showed that short paths exist between an arbitrary pair of vertices, but can we guarantee that one can find such a path for communication? [6] Or, put another way, Watts-Strogatz model is a small world, does it also allow efficient decentralized search? "Decentralized" here means operating with local information only, or knowing only the locations of their direct acquaintances. [9]

To understand what "decentralized" means in more details, we can have a look at a model for decentralized search. It is not difficult to model the type of decentralized search from Milgram's small-world experiment. Starting with Watts and Strogatz's grid-based model, we suppose that a starting node s needs to pass a message to a target node t through all the edges along the network. So what "decentralized" actually means is, s only knows the location of t on the grid, but, crucially, it does not know the random edges out of any node other than itself. Each intermediate along the path only has its partial information, and it must choose which of its neighbours to send the message to next. Just like the volunteers in the Milgram experiment collectively constructed paths to the target in Boston, these choices form a collective procedure for constructing a path from s to t. We will analyse all search procedures based on their delivery time. The delivery time here means the expected number of steps required to reach the target, over a set of randomly selected starting and target nodes, and long-range contacts. [6]

Unfortunately, with this set-up, it turns out that the decentralized search in the Watts-Strogatz model happens to require a large number of steps to reach the target. It is in fact much larger than the actual length of the shortest path. Hence, as a mathematical model, the Watts-Strogatz model is good at finding the density of clusters of people and the existence of short paths, but not the ability of people, working together in the network to actually find those paths. [6] The fundamental difficulty is that you are trying solve a global problem – finding a short path – using only local information about the network. [4]

We now know that decentralized search cannot be performed efficiently with the structure of the Watts-Strogatz model, Jon Kleinberg therefore came up with a model that exhibits both properties required: the networks have short paths, and these short paths can be found with decentralized search. [6]

In the Kleinberg model, he layered a set of "long-range" links on top of the lattice network. He assumed that it is a k-dimensional lattice. In this lattice, each vertex is connected to all of its nearest neighbours. So each vertex has 2k local connections, and they are bidirectional: if u connects to v, then v also connects to u. Each vertex u has a single long-range link (u,x) other than this, where x is chosen uniformly among the vertices some distance d away (see figure

5 below). This model is very similar to the Watts-Strogatz model if we choose d uniformly at random (rich in local connections, with a few long-range connections). [2]

Kleinberg showed the probability of adding a random link follows a power law with scaling exponent r. When r equals the dimensionality of the lattice k, i.e. r = k, then according to a greedy routing algorithm, the packets will be delivered in O($log^k\ n$) steps. This greedy routing algorithm is the same as the procedure in Milgram's experiment. The volunteers examine all of their neighbours (the 2k local neighbours and the 1 long-range neighbour) and deliver the packet to the neighbour whose remaining distance to the target is the smallest. The process is repeated at the new location until the destination is reached. [2]

He began with a two-dimensional grid (i.e. r = 2) and allowed edges to be directed. In this grid, the lattice points in this n × n square (as shown in figure 5) represent individuals in the social network. [9]

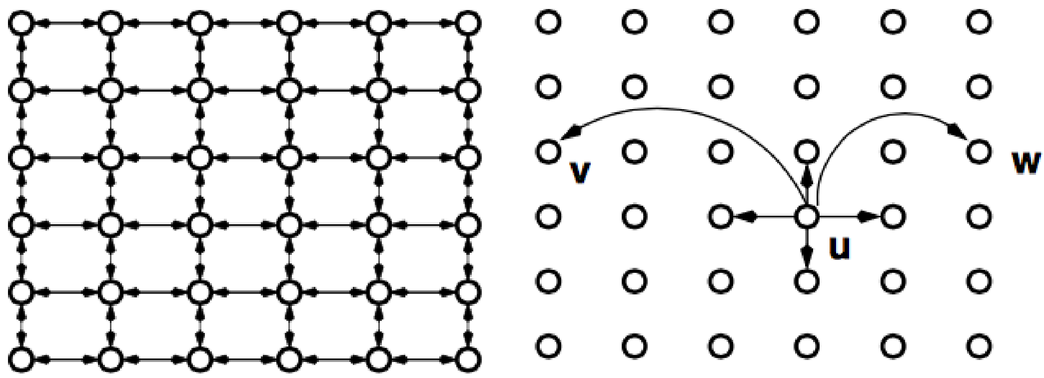


Figure 5: The Kleinberg model [2]

1. The node u has a directed edge to every other node within lattice distance p (p ≥ 1, and p is a constant). These are its local constants.
2. For universal constants q ≥ 0 and r ≥ 0, we also give u a number of q other directed edges (the long-range contacts) with independent random trials. The i^{th} directed edge from u has endpoint v with probability proportional to $[d(u,v)]^{-r}$. We divide it by $\Sigma_v[d(u,v)]^{-r}$ (normalising constant) to obtain a probability distribution. Here we define the distance d(u, v) as Manhattan distance (the number of grid steps between them) on the lattice between u and v, i.e., d(u, v) = $\sum_{i=1}^{k} |u_i - v_i|$. Remember r is the exponent (or clustering exponent). In this analysis we are looking at a two dimensional lattice, therefore k = 2. thus optimal routing occurs when r = k = 2. [2] [9]

When r = 0, we have the uniform distribution over long-range contacts, the distribution used in the Watts-Strogatz model - long-range contacts are chosen independently of their positions on the lattice. As r increases, the long-range contacts of a node become more and more clustered around the node. Thus, r measures how widely "networked" the underlying society of nodes is. [9] Or, put another way, the original grid-based model (figure 1-a) corresponds to r = 0. This is when the links are chosen uniformly at random.

Varying r is like turning a knob that controls how uniform the random links are. When r is small, the long-range links are “too random to be found”, and cannot be used for decentralized search; when q is large, the long-range links are “not random enough”, since there are not enough long-distance jumps that are required to make the world small. [6]

We choose two arbitrary nodes in the network u and v. We make u the starting node, and v the target. We want to pass a message from u to v in as few steps as possible - based on Milgram’s experiment. [9] Particularly, the message holder u in a given step has knowledge of i) the set of local contacts among all nodes (i.e. the underlying grid structure); ii) where the target t is on the lattice; and iii) the locations and long-range contacts of all nodes that have come in contact with the message. [9] So the question becomes: is there an optimal operating point for the network, where the distribution of long-range links is balanced between the two extremes (“too random to be found” and “not random enough”) to allow efficient decentralized search? [6]
In fact, there is. The result turns out to be, in a two-dimensional space, and in the limit of large network size, decentralized search is most efficient when r = 2 (so that the random links follow an inverse-square distribution). [6]

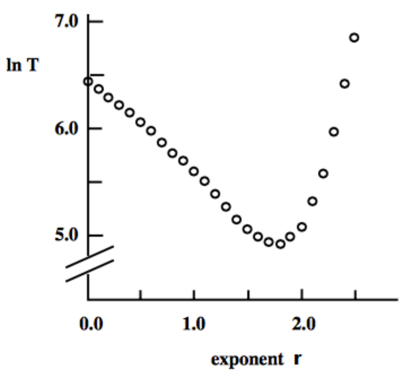


Figure 6: The behavior of lnT for a network of several hundred million nodes. [6]

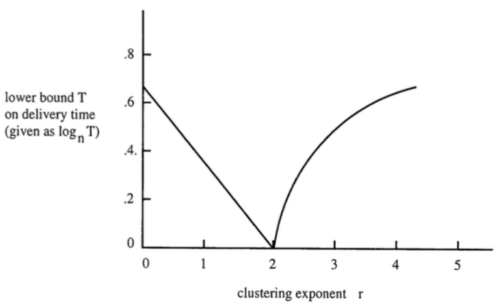


Figure 7: The behavior of lnT when the network size goes to infinity. [2]

Figure 6 shows the behaviour of decentralized search in the grid-based model for different values of r, for a network of several hundred million nodes. With this size of the network, i.e. several hundred million nodes, the exponent r has roughly the same efficiency from 1.5 to 2.0. And the best efficiency turns out to be slightly below 2.0. So we can conclude that the overall trend is, as the network size gets bigger, the optimal efficiency occurs as r approaches 2.0. Figure 7 shows the performance of this decentralized search method when the network size goes to infinity (best efficiency occurs at exactly 2!). [6]

Now you might wonder: what’s so special about the exponent r = 2 that makes the decentralized search most efficient? Here, we describe a short calculation that suggests why 2 is important. [6]

In the real world, where Milgram’s small-world experiment was conducted, we arrange distances into different “scales of resolution”: something can be around the world, across the country, across the state, across town, or down the block. For example, if we look at a node v, a good way to understand these "scales of resolution" is to consider the groups of all nodes at increasingly large distances from v: nodes at distance 2-4, 4-8, 8-16 and so on.[6]

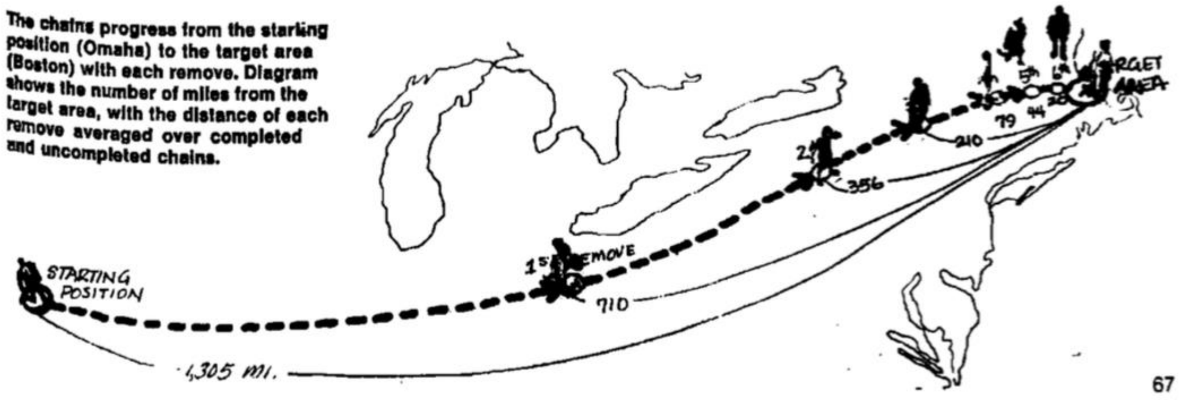


Figure 8: Milgram's small-world experiment [1]

The connection between this organizational scheme and the decentralized search can be shown by figure 8 above: effective decentralized search "funnels inward" through these different scales of resolution. The distance from the package to the target is reduced by approximately a factor two with each step. [6]

So, how does the inverse-square exponent r = 2 interact with these "scales of resolution"? Let's start with a single scale by taking a node v in the network, and a fixed distance d from the node v. Now we consider the group of nodes lying between d and 2d from the node v, as shown in figure 9 below. [6]

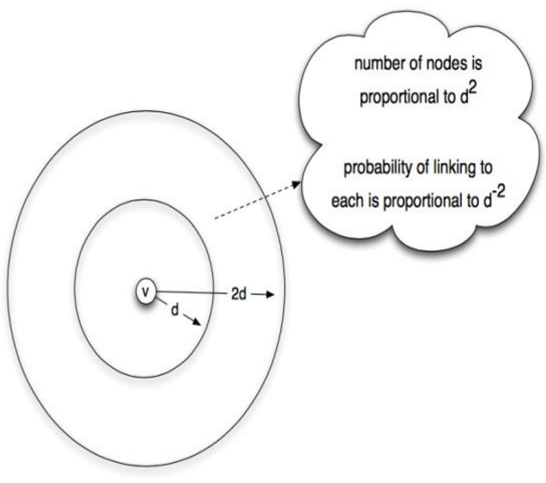


Figure 9: The concentric scales of resolution around a node [6]

What is the probability that v forms a link to some node inside this group? We know that area in the plane increases with the square of the radius, so the total number of nodes in this group is proportional to d^2. Also, the probability that v forms a link to any node in the group depends on how far it is from v, and each individual probability is proportional to d^{-2}. These two terms – the total number of nodes in the group and the probability of v forming a link to any one of them - approximately cancel out. So we can conclude: the probability that a random edge links to some node in this ring is approximately independent of the value of d. [6]

So, looking at this qualitatively, when the exponent r = 2: long-range weak ties are spread roughly uniformly over all different scales of resolution. [6] That means, in terms of social network, an individual is likely to have the same number of friends in each region, or scale. In other words, you should expect yourself to have the same number of friends in your neighbourhood as in the rest of the city, the same number of friends in the rest of the state, the same number again in the remainder of the country, and so on, right up to the scale of the whole world. You are roughly as likely to know someone who lives on another continent as someone who lives down the street. [4] In this case, when people try to deliver a package, they can consistently find ways to reduce their remaining distances to the target person, no matter how near or far they are. This is unlike how the postal service uses the address on an envelope to deliver a message - the address specifies the country, county, town, street, and finally the house number. But we need to remember that there is a large amount of money spent at the postal system to do precisely this job; the corresponding system of the inverse-square network only arise from a completely random patterns of links. [6]

The essence of Kleinberg's result is that, when this condition of equal connectivity at all length scales is satisfied, we know that not only do the short paths exist between all pairs of nodes, but also message senders can find the paths if each of them simply forwards the message to the person they know who seems closest to the target. All a sender has to worry about is passing the message onto the next phase of the search. You only need to get it to the right part of the world, and then let someone else worry about it. In so doing, you are assuming that the next person that you are passing the message to, being closer to the target, has more precise information than you do, and so is better able to advance the search to its next phase. This is essentially what the exponent r = 2 guarantees. [4]

As mentioned before, when exponent r = the dimensionality of the lattice k, then a greedy routing algorithm will deliver packets in O($log^k\ n$) steps. We prove this now.
First we set the probability distribution that u chooses v as its long-range contact to follow a power law form (the terms and the equation were defined in paragraph 2 on page 8):

$$\Pr(u \rightarrow v) = \frac{d(u,v)^{-r}}{\sum_{u \neq v} d(u,v)^{-r}} \qquad \text{eq(1)}$$

Because it's a two dimensional lattice (k = 2), we can therefore simplify its denominator:

$$\sum_{u \neq v} d(u,v)^{-2} \leq \sum_{j=1}^{2n-2} (4j)(j^{-2})$$

where the first term is the number of vertices at a distance j in a two-dimensional lattice, and the second term is the probability that u forms a link to a vertex at a distance j. Simplifying it further, we have

$$\begin{aligned}\sum_{u \neq v} d(u,v)^{-2} &= 4 \sum_{j=1}^{2n-2} j^{-1} \\ &\leq 4\ (1 + \ln(2n-2)) \\ &\leq 4\ (\ln 3 + \ln(2n)) \\ &\leq 4\ \ln 6n\end{aligned}$$

Then we can rewrite eq(1)

$$\Pr(u \to v) \geq \frac{d(u,v)^{-r}}{4ln6n}$$

which provides a normalized distribution. [2]

Now consider a packet being delivered from a node u to a node v. We divide the procedure into a set of "phases". We define phase j as the packet being at some vertex x such that $2^j <$ d(x,v) ≤ 2^{j+1}. So the 0th phase begins when d(x,v) ≤ 2, and the packet is at most two steps away from the target. A phase ends when the distance between the packet and the destination has been halved. This can be understood as a type of binary search (as shown in figure 10 below), and j ≤ log n. [2]

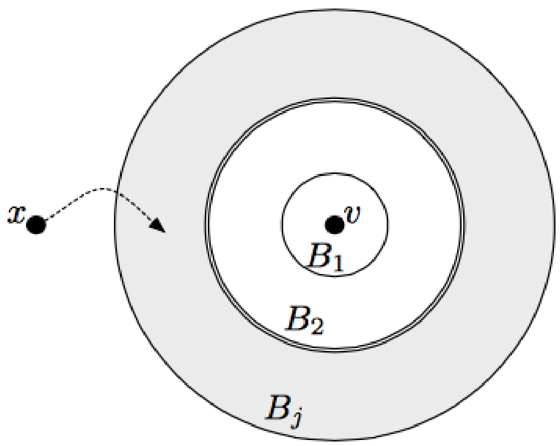


Figure 10: a type of binary search, where the procedure is divided into a set of "phases" [2]

When does the j^{th} phase end? From the definitions above, we can easily show that the j^{th} phase ends when d(x,v) < 2^j. Every time the packet is passed along a person in the chain, it gets a new chance to find a long-range link that will end the phase. This happens when the long-range link connects to a vertex which is within distance 2^j of the target v. These long-range links are independent of each other. So we say that the probability of such an event happening is the probability that x connects to some $\omega \in B_j$, where B_j is the set of vertices that are within distance 2^j of the target v. Vertex x can be connected to any of the vertices in ω. The number of vertices in ω is

$$1 + \sum_{i=1}^{2^j} i = 2^{2j-1} + 2^j + 1$$
$$= \frac{1}{2} 2^{2j} + \frac{1}{2} 2^j + 1$$
$$> 2^{2j-1}$$

We also need to node that these vertices are all within distance $2^{j+1} + 2^j < 2^{j+2}$ of x. We therefore can write Pr(u→v) as

$$\Pr(u \to v) = \frac{d(u,v)^{-2}}{\sum_{u \neq v} d(u,v)^{-2}}$$
$$\geq [(2^{2j+4})(4 \ln 6n)]^{-1}$$

We can therefore deduce that the probability the phase ends at vertex x is

$$\Pr(j^{th} \text{ phase ends at x}) \geq (2^{2j-1})\ [(2^{2j+4})\ (4 \ln 6n)]^{-1}$$
$$= \frac{1}{128\ ln6n}$$

If we make X_j the number of steps in the j^{th} phase, the expected value of X_j is

$$E[X_j] = \sum_{i=1}^{\infty} \Pr\ (X_j \geq i)$$
$$\leq \sum_{i=1}^{\infty} (1 - \frac{1}{128 \ln 6n})^{-1}$$

$= 128 \ln 6n$

And now we can see that Pr(jth phase ends at x) = 1/E[X_j].

The total number of steps to the packet is the sum of the number of steps of all the phases. Remember that because each phase halves the remaining distance to the target, there can be a maximum of $log_2 n$ phases. The expected number of steps required for delivery is therefore

$$E[X] = E[\sum_{j=0}^{log_2 n} X_j]$$
$$= \sum_{j=0}^{log_2 n} E[X_j]$$
$$\leq (1+log_2 n)\ (128 \ln 6n)$$
$$\leq \alpha_2 log^2 n = O(log^2 n).\ [2]$$

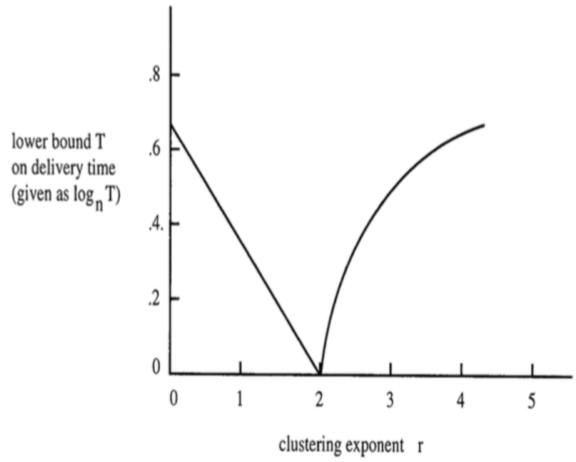


Figure 11: The delivery time reaches its best value when r=2

When r = 0, that means all nodes in the lattice are equally likely to be random contacts. In this case, plenty of short paths exist, but they are "too random" to be found. When r is large, the probability of a random shortcut decreases so rapidly with distance that only the nodes nearby (on the lattice) can be connected. In this case long-range contacts are impossible, there are no short paths to be found. What Kleinberg showed was, in a two-dimensional lattice, when r equals two, the network reaches the best balance between the navigational convenience of the lattice and the distance-cutting power of long-range edges. That is to say, when r equals exactly two, the network possesses short paths that individuals can actually find. [4]

Conclusion: In this essay, we first discussed the Milgram's small-world experiment and its outcome, which leads to the Watts and Strogatz's model. Their model can be expressed in two forms: grid-based model and beta model. In their paper, they also discussed the effect of the small-world phenomenon in a dynamical system (disease spreading). Jon Kleinberg then showed the relationship between the clustering exponent and the "randomness" of the long-range links of the network, and decentralized search is most efficient when the random links follow an inverse-square distribution (in a two-dimensional case). He also calculated that for exponent r = 2 in a two-dimensional network, the packets will be delivered in O($log^k\ n$) steps. The analysis of the small-world phenomenon can go beyond the content of this essay by validating Jon Kleinberg's theory on real data from Facebook and LiveJournal, after generalizing the model to use rank-based friendship.

People have been interested in the small-world phenomenon since 1960s, but now we can describe it in a much better way. The results we get from all the experiments and papers also help us to understand some social, natural and technological networks, as well as the development of some popular websites such as LinkedIn, Facebook and Twitter. It clearly plays an important role in modern culture.

At the end of the essay, I want to thank all the scientists that contributed and all the effort they put in. Without Jon, we would never have been able to know how to think about the search problem – we would not have known which door to walk through. And without Watts and Strogatz's models and explanations on the small-world phenomenon, Jon would never have thought about the problem at the beginning. Without Milgram, we would not have known what it is that we were trying to work out. When we look back, everything seems obvious, but the truth is that the small-world problem could only have been solved through the contributions from many different thinkers, coming from all different angles and with an amazing diversity of skills, ideas and perspectives. [4]

References:

[1] Milgram, S. The small world problem. *Psychology Today,* 2, 60-67 (1967)
[2] Lecture 10, Aaron Clauset, *Network Analysis and Modeling lecture notes* (2013).
[3] Chapter 2, David Easley and Jon Kleinberg, *Networks, Crowds and Markets: Reasoning about a Highly Connected World* (2010), Cambridge University Press, ISBN: 978-0521195331.
[4] Duncan Watts *Six Degrees: The Science of a Connected Age* (2004) Vintage Press, ISBN: 978-0099444961.
[5] December 2008 – Author Commentaries: Duncan Watts & Steve Strogatz, by ScienceWatch.com.
[6] Chapter 20, David Easley and Jon Kleinberg, *Networks, Crowds and Markets: Reasoning about a Highly Connected World* (2010), Cambridge University Press, ISBN: 978-0521195331.
[7] Duncan J. Wawtts and Steven H. Strogatz, Collective dynamics of 'small-world' networks. *Nature,* Vol.393, 440-442, 4th June 1998.
[8] Kretschmar, M. & Morris, M. Measures of concurrency in networks and the spread of infectious disease. *Math. Biosci*. **133**, 165-195 (1996).
[9] J.Kleinberg, The Small-World Phenomenon, an Algorithmic Perspective. *Proc. 32nd ACM Symposium on Theory of Computing* (2000).
[10] Bollabás, B. *Random Graphs* (Academic, London, 1985).
[11] Achacoso, T.B. & Yamamoto, W.S. *AY's Neuroanatomy of C.elegans for Computation* (CRC Press, Boca Raton, FL, 1992)
[12] Longini, I.M. Jr A mathematical model for predicting the geographic spread of new infectious agents. *Math. Biosci*. **90,** 341-366 (1988).
[13] Hess, G. Disease in metapopulation models: implications for conservation. *Ecology* **77,** 1617-1632 (1996).
[14] Das, R., Mitchell, M. & Crutchfield, J. P. *in Parallel Problem Solving from Nature* (eds Davido, Y., Schwefel, H.-P. & Männer, R.) 344-353 (Lecture Notes in Computer Science 866, Springer, Berlin, 1994).
[15] Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359,** 826-829 (1992).
[16] Kuramoto, Y. *Chemical Oscillations, Waves, and Turbulence* (Springer, Berlin, 1984).