CS 577 – Spring 2015
Programming Project 3
The Small World Problem

Background.

The attached article describes a famous experiment by the social psychologist Stanley Milgram. His main conclusion is that the distance between two people tends to be fairly small. Here, we consider two people to be linked if they know each other personally, and the distance is just the minimum number of such links needed for one to reach the other.

Why did Milgram become interested in this? One reason is that social networks form a kind of invisible "glue" behind more formal economic and political entities. For example, Milgram cites work explaining how the breakdown of intercity communication led to the Dark Ages in Europe. Moving forward to our own time, it is natural to speculate about how well a large computer network such as the Internet, or its informally maintained population of web pages, is connected. Information of this type could be useful to web page designers, librarians, or advertisers.

In this project, you will work with some graphs that might display some of the features observed by Milgram and his successors. These features include global connectivity and the frequency of short paths between nodes.

Techniques and Skills Needed.

You should be familiar with algorithms for searching undireccted graphs, and for finding shortest paths. These are described in Chapters 3 and 4 of our text. You should be able to choose between graph representations, depending on the likely edge density of your graph. Finally, you should know how to implement the algorithms you need in the representation(s) you choose to use. You may use any language or system that you find convenient.

Generating Input Data.

For your experiments, you will need to generate graphs exhibiting different properties. We will first describe three different models commonly used for this purpose, and then ask you to make one of your own.

1. Independent edges.

Let n denote the number of vertices. For each of the $\binom{n}{2}$ possible edges, flip a coin that has

$$\Pr[\text{ head }] = p, \qquad \Pr[\text{ tail }] = 1 - p.$$

These coin flips are independent. For each coin that came up heads, add the edge. This model has one parameter p, the probability of "edgeness." Note that $0 \leq p \leq 1$. The idea of this model is to generate graphs without imposing any additional structure.

2. Local preference.

This model reflects the idea that you are more likely to know the people who live near you than the people who live far away. Start by picking n points uniformly distributed over the rectangle $[0,1]^2$. You can do this by just choosing x and y coordinates uniformly between 0 and 1. Now, let $L = \sqrt{2}$, the maximum possible distance between pairs. If p, q are points that are distance d apart, the probability of a link between p and q is

$$\beta e^{-d/(L\alpha)}.$$

Here, α and β are parameters of the model, satisfying $\alpha > 0$ and $0 \leq \beta \leq 1$. Note that β controls the overall number of edges, and α controls the mix of long and short edges.

3. Preferential attachment.

In this model, nodes are added one at a time. Start with $n_0/2$ pairs of edges. Now, for each of the nodes $n_0 + 1, \ldots, n$, sample (with replacement) m nodes from the existing network, using the law

$$\Pr[\text{ choose node } i\ ] = \frac{d_i}{\sum_j d_j}.$$

Here, $d1, d2, \ldots$ are the degrees of the already existing nodes, and the sum is over previously created nodes. Create a new node and add an edge linking it to each node appearing in the sample. Parameters of this model are the initial population n_0 and the maximum number m of new edges per stage. The idea behind this model is to simulate the so-called "Matthew effect," whereby new resources are more likely to be alloted to those already having them.

4. Other models.

Please add to the three models above by studying a different scheme of your own design. Start by describing a real-world situation you find interesting, in which questions of "connectedness" naturally appear. Then, decide how to generate graphs in a way that models this situation. Finally, specify your model precisely enough that a reader could do experiments with it.

Studies to Do.

A. Edge density.

When using different models, we would like to guarantee that the graphs are roughly similar. One way to do this is to control the edge density

$$\delta = \frac{\#\text{ of edges observed}}{\#\text{ of edges possible}}$$

Note that the denominator is always $n(n-1)/2$. For any graph, $0 \leq \delta \leq 1$.

Intuitively, we would expect that graphs exhibiting larger values of δ are better connected. For this reason, when comparing models, you should test graphs with similar

values of δ against each other. To do this effectively you will need to know how the edge density changes with variations in the parameters of a model. This can be studied by simulation. For each of the models above, try to answer the following type of question. For a given number n of nodes, and desired edge density δ, how should I set parameters in each of my models?

B. Connectivity.

Milgram speculated about whether the social connection graph of a large country is likely to be fully connected. This means that any two citizens could be linked by a path, albeit a fairly long one. Choose an algorithm that you can use to test this hypothesis on a particular graph. For each of your models, how dense does the graph have to be before full connectivity appears?

C. Short Paths.

Try to replicate Milgram's experiment in a random graph setting. This means you should pick two nodes and then try to find a path between them. For each model (and appropriate choices of graph size and parameters), you should investigate the chance of a successful connection, and the distribution of path lengths found. Do your conclusions support Milgram's observation that the number of links between two people tends to be small?

Turning it In.

This project can be used in lieu of Homework Assignment 5.

Please submit your code and data in the Learn@UW drop box. Also, for each experiment you did, write two paragraphs indicating your methods (what algorithms, etc.) and your main conclusions. These paragraphs should be supported by graphs or tables, made from your data. This part can be submitted in paper form; use the plastic box outside 4382 CS.

Submit your work by 4 PM Thursday, April 16. (One submission per group, please.)

References.

[1] Stanley Milgram, The Small World Problem, Psychology Today, v. 1, no. 1, May 1967, pp. 61-67. [attached]

[2] Wikipedia page on Small-World Experiment.

The Small-World Problem

By Stanley Milgram

Fred Jones of Peoria, sitting in a sidewalk cafe in Tunis, and needing a light for his cigarette, asks the man at the next table for a match. They fall into conversation; the stranger is an Englishman who, it turns out, spent several months in Detroit studying the operation of an interchangeable-bottlecap-factory. "I know it's a foolish question," says Jones, "but did you ever by any chance run into a fellow named Ben Arkadian? He's an old friend of mine, manages a chain of supermarkets in Detroit . . ."

"Arkadian, Arkadian," the Englishman mutters. "Why, upon my soul, I believe I do! Small chap, very energetic, raised merry hell with the factory over a shipment of defective bottlecaps."

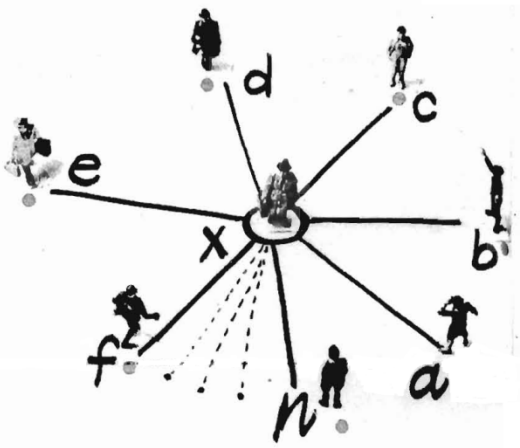
"No kidding!" Jones exclaims in amazement.

"Good lord, it's a small world, isn't it?"

Almost all of us have had the experience of encountering someone far from home, who, to our surprise, turns out to share a mutual acquaintance with us. This kind of experience occurs with sufficient frequency so that our language even provides a cliché to be uttered at the appropriate moment of recognizing mutual acquaintances. We say, "My it's a small world."

Psychology Today, vol. 1, no. 1, May 1967, pp61-67

Random dispersement of people in the small world.



Each person's first-hand acquaintances are shown, A through N.

The simplest way of formulating the small-world problem is: Starting with any two people in the world, what is the probability that they will know each other? A somewhat more sophisticated formulation, however, takes account of the fact that while persons X and Z may not know each other directly, they may share a mutual acquaintance—that is, a person who knows both of them. One can then think of an acquaintance chain with X knowing Y and Y knowing Z. Moreover, one can imagine circumstances in which X is linked to Z not by a single link, but by a series of links, X-*a*-*b*-*c*-*d* . . . *y*-Z. That is to say, person X knows person *a* who in turn knows person *b*, who knows *c* . . . who knows *y*, who knows Z.

Therefore, another question one may ask is: Given any two people in the world, person X and person Z, how many intermediate acquaintance links are needed before X and Z are connected?

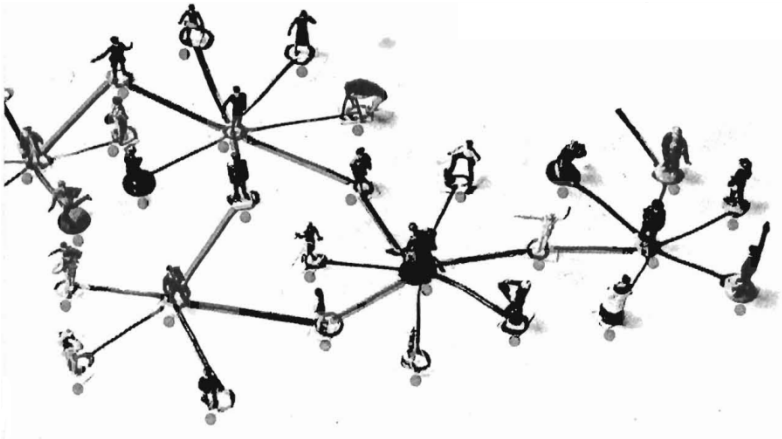
Concern with the small-world problem is not new, nor is it limited to social psychologists like myself. Historians, political scientists, and communication specialists share an interest in the problem. Jane Jacobs, who is concerned with city planning, describes an acquaintance chain in terms of a children's game:

> When my sister and I first came to New York from a small city, we used to amuse ourselves with a game we called Messages. I suppose we were trying, in a dim way, to get a grip on the great, bewildering world into which we had come from our cocoon. The idea was to pick two wildly dissimilar individuals—say a head hunter in the Solomon Islands and a cobbler in Rock Island, Illinois—and assume that one had to get a message to the other by word of mouth; then we would each silently figure out a plausible, or at least possible, chain of persons through which the message could go. The one who could make the shortest plausible chain of messengers won. The head hunter would speak to the head man of his village, who would speak to the trader who came to buy copra, who would speak to the Australian patrol officer when he came through, who would tell the man who was next slated to go to Melbourne on leave, etc. Down at the other end, the cobbler would hear from his priest, who got it from the mayor, who got it from a state senator, who got it from the governor, etc. We soon had these close-to-home messengers down to a routine for almost everybody we could conjure up . . .

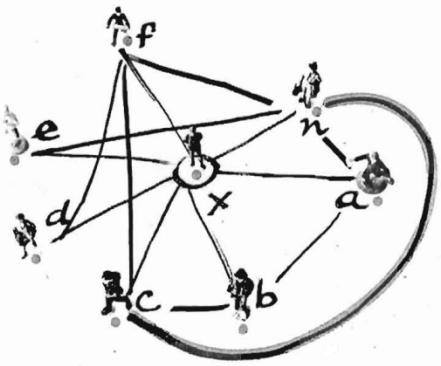
The importance of the problem does not lie in these entertaining aspects, but in the fact that it brings under discussion a certain mathematical structure in society, a structure that often plays a part, whether recognized or not, in many discussions of history, sociology, and other disciplines. For example, Henri Pirenne and George Duby, important historians, make the point that in the Dark Ages communication broke down between cities of western Europe. They became isolated and simply did not have contact with each other. The network of acquaintances of individuals became constricted. The disintegration of society was expressed in the growing isolation of communities, and the infrequent contact with those living outside a person's immediate place of residence.

There are two general philosophical views of the small-world problem. One

The network spreads, with complicated inter-connections.



With group inbreeding, X's acquaintances feed back into his own circle, normally eliminating new contacts.



view holds that any two people in the world, no matter how remote from each other, can be linked in terms of intermediate acquaintances, and that the number of such intermediate links is relatively small. This view sees acquaintances in terms of an infinitely intersecting arrangement that permits movement from any social group to another through a series of connecting links.

The second view holds that there are unbridgeable gaps between various groups and that therefore, given any two people in the world, they will never link up because people have circles of acquaintances which do not necessarily intersect. A message will circulate in a particular group of acquaintances, but may never be able to make the jump to another circle. This view sees the world in terms of concentric circles of acquaintances, each within its own orbit.

The Underlying Structure

Sometimes it is useful to visualize the abstract properties of a scientific problem before studying it in detail; that is, we construct a model of the main features of the phenomenon as we understand them. Let us represent all the people in the United States by a number of blue points. Each point represents a person, while lines connecting two points show that the two persons are acquainted. [*See top illustration, opposite page.*] Each person has a certain number of first-hand acquaintances, which we shall represent by the letters *a, b, c, . . . n.* Each acquaintance in turn has his own acquaintances, connected to still other points. The exact number of lines radiating from any point depends on the size of a person's circle of acquaintances. The entire structure takes on the form of a complex network of 200 million points, with complicated connections between them [*see bottom left illustration, opposite page*]. One way of restating the small-world problem in these terms is this: Given any two of these points chosen at random from this universe of 200 million points, through how many intermediate points would we pass before the chosen points could be connected by the shortest possible path?

Research at M.I.T.

There are many ways to go about the study of the small-world problem, and I shall soon present my own approach to it. But first, let us consider the important contributions of a group of workers at The Massachusetts Institute of Technology, under the leadership of Ithiel de Sola Pool. Working closely with Manfred Kochen of IBM, Pool decided to build a theoretical model of the small-world, a model which closely parallels the idea of points and lines shown. However, unlike my own model, which is purely pictorial, Pool and Kochen translate their thinking into strict mathematical terms.

To build such a model they needed certain information. First, they had to know how many acquaintances the average man has. Surprisingly, though this is a very basic question, no reliable answers could be found in the social science literature. So the information had to be obtained, a task which

The beginning of a typical chain (#111) in the Nebraska Study.

STARTING PERSON
Widowed clerk in Omaha, Nebraska

Michael Gurevitch, then a graduate student at M.I.T., undertook. Gurevitch asked a variety of men and women to keep a record of all the persons they came in contact with in the course of 100 days. It turned out that on the average, these people recorded names of roughly 500 persons, so that this figure could be used as the basis of the theoretical model. Now, if every person knows 500 other people, what are the chances that any two people will know each other? Making a set of rather simple assumptions, it turns out that there is only about one chance in 200,000 that any two Americans chosen at random will know each other. However, when you ask the chances of their having a mutual acquaintance, the odds drop sharply. And quite amazingly, there is better than a 50-50 chance that any two people can be linked up with two intermediate acquaintances. Or at least, that is what the Pool-Kochen theory indicates.

Of course, the investigators were aware that even if a man has 500 acquaintances, there may be a lot of inbreeding. That is, many of the 500 friends of my friend may be actually among the people I know anyway, so that they do not really contribute to a widening net of acquaintances; the acquaintances of X simply feed back into his own circle and fail to bring any new contacts into it [*see bottom right illustration, opposite page*]. It is a fairly straightforward job to check up on the amount of inbreeding if one uses only one or two circles of acquaintances, but it becomes almost impossible when the acquaintance chain stretches far and wide. So many people are involved that a count just isn't practical.

So the big obstacle one runs up against is the problem of social structure. Though poor people always have acquaintances, it would probably turn out that they tend to be among other poor people, and that the rich speak mostly to the rich. It is exceedingly difficult to assess the impact of social structure on a model of this sort. If you could think of the American population as simply 200 million points, each with 500 random connections, the model would work. But the contours of social structure make this a perilous assumption, for society is not built on random connections among persons but tends to be fragmented into social classes and cliques.

A Harvard Approach

The Pool and Kochen mathematical model was interesting from a theoretical standpoint, but I wondered whether the problem might not be solved by a more direct experimental approach. The Laboratory of Social Relations at Harvard gave me $680 to prove that it could. I set out to find an experimental method whereby it would be possible to trace a line of acquaintances linking any two persons chosen at random.

Let us assume for the moment that the actual process of establishing the linkages between two persons runs only one way: from person A to person Z. Let us call person A the *starting* person, since he will initiate the process, and person Z the *target* person, since he is the person to be reached. All that would be necessary, therefore, would be to choose a starting person at random from the 200 million people who live in the United States, and then randomly choose a target person

This is how the study was carried out. The general idea was to obtain a sample of men and women from all walks of life. Each of these persons would be given

the name and address of the same target person, a person chosen at random, who lives somewhere in the United States. Each of the participants would be asked to move a message toward the target person, using only a chain of friends and acquaintances. Each person would be asked to transmit the message to the friend or acquaintance who he thought would be most likely to know the target person. Messages could move only to persons who knew each other on a first-name basis.

As a crude beginning, we thought it best to draw our starting persons from a distant city, so we chose Wichita, Kansas for our first study and Omaha, Nebraska for our second. (From Cambridge, these cities seem vaguely 'out there,' on the Great Plains or somewhere.) To obtain our sample, letters of solicitation were sent to residents in

1st REMOVE
Self-employed friend in Council Bluffs, Iowa

these cities asking them to participate in a study of social contact in American society. The target person in our first study lived in Cambridge and was the wife of a divinity school student. In the second study, carried out in collaboration with Jeffrey Travers, the target person was a stockbroker who worked in Boston and lived in Sharon, Massachusetts. To keep matters straight, I will refer to the first study as the Kansas Study, and the second as the Nebraska Study. These terms indicate merely where the starting persons were drawn from.

Each person who volunteered to serve as a starting person was sent a folder containing a document, which served as the main tool of the investigation. Briefly, the document contains:

1. The name of the target person as well as certain information about him. This orients the participants toward a specific individual.
2. A set of rules for reaching the target person. Perhaps the most important rule is: *"If you do not know the target person on a personal basis, do not try to contact him directly. Instead, mail this folder . . . to a personal acquaintance who is more likely than you to know the target person . . . it must be someone you know on a first-name basis."* This rule sets the document into motion, moving it from one participant to the next, until it is sent to someone who knows the target person.
3. A roster on which each person in the chain writes his name. This tells the person who receives the folder exactly who sent it to him. The roster also has another practical effect; it prevents endless looping of the folder through participants who have already served as links in the chain, because each participant can see exactly what sequence of persons has led up to his own participation.

In addition to the document, the folder contains a stack of 15 business reply, or "tracer" cards. Each person receiving the folder takes out a card, fills it in, returns it to us, and sends the remaining cards along with the document to the next link.

Several other features of the procedure need to be emphasized. First, each

2nd REMOVE
Publisher in Belmont, Mass

participant is supposed to send the folder on to one other person only. Thus the efficiency with which the chain is completed depends in part on the wisdom of his choice in this matter. Second, by means of the tracer card, we have continuous feedback on the progress of each chain. The cards are coded so we know which chain it comes from and which link in the chain has been completed. The card also provides us with relevant sociological characteristics of the senders of the cards. Thus, we know the characteristics of completed, as well as incompleted, chains. Third, the procedure permits experimental variation at many points.

In short, the device possesses some of the features of a chain letter, though it does not pyramid in any way; moreover it is oriented toward a specific target, zeroes in on the target through the cooperation of a sequence of participants, and contains a tracer that allows us to keep track of its progress at all times.

Would It Work?

The question that plagued us most in undertaking this study was simply: Would the procedure work? Would any of the chains started in Kansas actually reach our target person in Massachusetts? Part of the excitement of experimental social psychology is that it is all so new we often have no way of knowing whether our techniques will work or simply turn out to be wispy pipe dreams.

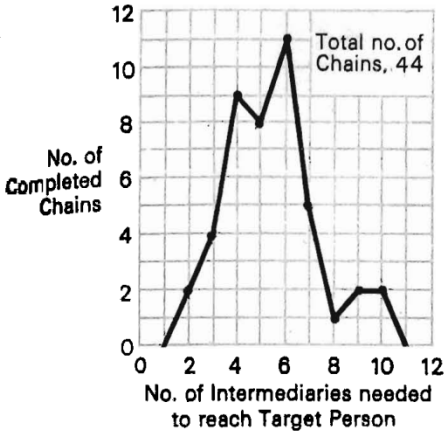
The answer came fairly quickly. It will be recalled that our first target person

3rd REMOVE
Tanner in Sharon, Mass.

was the wife of a student living in Cambridge. Four days after the folders were sent to a group of starting persons in Kansas, an instructor at the Episcopal Theological Seminary approached our target person on the street. "Alice," he said, thrusting a brown folder toward her, "this is for you." At first she thought he was simply returning a folder

4th REMOVE
Sheet metal worker in Sharon, Mass.

that had gone astray and had never gotten out of Cambridge, but when we looked at the roster, we found to our pleased surprise that the document had started with a wheat farmer in Kansas. He had passed it on to an Episcopalian minister in his home town, who sent it



In the Nebraska Study the chains varied from two to 10 intermediate acquaintances with the median at five.

in the minister who taught in Cambridge, who gave it to the target person. Altogether the number of intermediate links between starting person and target person amounted to *two!*

How Many Intermediaries?

As it turned out, this was one of the shortest chains we were ever to receive, for as more tracers and folders came in, we learned that chains varied from two to 10 intermediate acquaintances, with the median at five [*see illustration above*]. A median of five intermediate persons is, in certain ways, impressive, considering the distances traversed. Recently, when I asked an intelligent friend of mine how many steps he thought it would take, he estimated that it would require 100 intermediate persons or more to move from Nebraska to Sharon. Many people make somewhat similar estimates, and are surprised to learn that only five intermediaries will—on the average—suffice. Somehow it does not accord with intuition. Later, I shall try to explain the basis of the discrepancy between intuition and fact.

5th REMOVE
Dentist in Sharon, Mass.

On a purely theoretical basis, it is reasonable to assume that even fewer links are essential to complete the chains. First, since our participants can send the folder to only one of their 500 possible contacts, it is unlikely that even through careful selections, they will necessarily and at all times, select the contact best able to advance the chain to the target. On the whole they probably make pretty good guesses but surely, from time to time, they overlook some possibilities for short cuts. Thus, the chains obtained in our empirical study are less efficient than those generated theoretically

Second, by working on a highly rational basis, each intermediary moves the folder toward the target person. That is, a certain amount of information about the target person—his place of employment, place of residence, schooling, and so forth—is given to the starting subject, and it is on the basis of this information alone that he selects the next recipient of the folder. Yet, in real life, we sometimes know a person because we chance to meet him on an ocean liner, or we spend a summer in camp together as teenagers, yet these haphazard bases of acquaintanceship cannot be fully exploited by the participants.

There is one factor, however, that could conceivably have worked in the opposite direction in our experiments, giving us the illusion that the chains are shorter than they really are. There is a certain decay in the number of active chains over each remove, even when they do not drop out because they reach the target person. Of 160 chains that started in Nebraska, 44 were completed and 126 dropped out. These chains die before completion because on each remove a certain proportion of participants simply do not cooperate and fail to send on the folder. Thus, the results we obtained on the distribution of chain lengths occurred within the general drift of a decay curve. It is possible that some of the incompleted chains would have been longer than those that were completed. To account for this possibility, Harrison White of Harvard has constructed a mathematical model to show what the distribution of chain lengths would look like if all chains went through to completion. In terms of this model, there is a transformation of the data, yielding slightly longer chains.

6th REMOVE
Printer in Sharon, Mass.

Examining the Chains

Several features of the chains are worth examining, for they tell us something about the pattern of contact in American society. Consider, for example, the very pronounced tendency in our Kansas Study for females to send the folder on to females, and males to send it on to males. Of the 145 participants involved in the study, we find:

Female	⇒	Female	56
Male	⇒	Male	58
Female	⇒	Male	18
Male	⇒	Female	13

Thus participants were three times as likely to send the folder on to someone of the same sex as to someone of the opposite sex. Exactly why this is so is not easy to determine, but it suggests that certain kinds of communication are strongly conditioned by sex roles.

7th REMOVE
Clothing merchant in Sharon, Mass.

Participants indicated on the reply cards whether they were sending the folder on to a friend, a relative, or an acquaintance. In the Kansas Study, 123 sent the folder to friends and acquaintances, while only 22 sent it to relatives. Cross-cultural comparison would seem useful here. It is quite likely that in societies which possess extended kinship systems, relatives will be more heavily represented in the communication network than is true in the United States. In American society, where extended kinship links are not maintained, acquaintance and friendship links provide the preponderant basis for reaching the target person. I would guess, further, that within certain ethnic groups in the United States, a higher proportion of

familial lines would be found in the data. Probably, for example, if the study were limited to persons of Italian extraction, one would get a higher proportion of relatives in the chain. This illustrates, I hope, how the small world technique may usefully illuminate varied aspects of social structure.

Spaced throughout the preceding text is a series of illustrations showing the kinds of people found in a typical chain (number 111) from the Nebraska Study.

Common Pathways

Each of us is embedded in a small-world structure. It is not true, however, that each of our acquaintances constitutes an equally important basis of contact with the larger social world. It is obvious that some of our acquaintances are more important than others in establishing contacts with broader social realms; some friends are relatively isolated, while others possess a wide circle of acquaintances, and contact with them brings us into a far-ranging network of additional persons.

Referring to our Nebraska Study, let us consider in detail the pattern of convergence crystallizing around the target person—the stockbroker living in Sharon, Massachusetts, and working in Boston [*see top illustration opposite page*]. A total of 64 chains reached him. (44 chains originated in Nebraska and 20 chains, from an auxiliary study, originated in the Boston area). Twenty-four of the chains reached him at his place of residence in the small town outside of Boston. Within Sharon, 16 were given to him by Mr. Jacobs, a clothing merchant in town. Thus, the clothing merchant served as the principal point of mediation between the broker and a larger world, a fact which came as a considerable surprise, and even something of a shock for the broker. At his place of work, in a Boston brokerage house, 10 of the chains passed through Mr. Jones, and five through Mr. Brown. Indeed, 48 percent of the chains to reach the broker were moved on to him by three persons: Jacobs, Jones, and Brown. Between Jacobs and Jones there is an interesting division of labor. Jacobs mediates the chains advancing to the broker by virtue of his residence. Jones performs a similar function in the occupational domain, and moves 10 chains enmeshed in the investment-brokerage network to the target person.

More detail thus fills in the picture of the small world. First, we learn that the target person is not surrounded by acquaintance points, each of which is equally likely to feed into an outside contact; rather, there appear to be highly popular channels for the transmission of the chain. Second, there is differentiation among these commonly used channels, so that certain of them provide the chief points of transmission in regard to residential contact, while others have specialized contact possibilities in the occupational domain. For each possible realm of activity in which the target person is involved, there is likely to emerge a sociometric star with specialized contact possibilities.

Geographic and Social Movement

The geographic movement of the folder from Nebraska to Massachusetts is striking. There is a progressive closing in on the target area as each new person is added to the chain. [*See bottom illustration, opposite page.*] In some cases, however, a chain moves all the way from Nebraska to the very neighborhood in which the target person resides, but then goes round and round, never quite making the necessary contact to complete the chain. Some chains died only a few hundred feet from the target person's house, after a successful journey of 1000 miles. Thus we see that social communication is sometimes restricted less by physical distance than by social distance.

TARGET PERSON
Stock broker living in Sharon, Mass.

The next step is to see what happens when we change the relationship between the starting person and the target person. That is, if the two are drawn from different class backgrounds, does this then decrease the probability of completing the chain? Does it increase the number of links?

In collaboration with Charles Korte, I am now applying the small-world method to the study of communications between subgroups in American society—Negro and white. We will have both Negro and white starting persons, but only Negro target persons, and try to trace the lines of communication between them. First, we want to ask: In what degree are the racial lines surmounted? Can any sizeable fraction of the communications get through the racial barrier? If the answer is yes, we then want to identify the typical locus of transmission. Does it occur at the neighborhood level, or at the place of work? We are particularly interested in the persons who serve as links between Negro and white groups. In what way do they differ from others in the chain? Do they tend to occupy particular professional categories, such as minister, teacher, and so forth? Is the communication flow between Negroes and whites easier in Northern or in Southern locales? Perhaps some new light can be cast on the structural relationships between Negro and white communities by probing with the small-world method.

Intuition and Fact

As we saw above, many people were surprised to learn that only five intermediaries will, on the average, suffice to link any two randomly chosen individuals, no matter where they happen to live in the United States. We ought to try to explain the discrepancy between intuition and fact.

The first point to remember is that although we deal directly with only five intermediaries, behind each of them stands a much larger group of from 500 to 2500 persons. That is, each participant has an acquaintance pool of 500 to 2500 persons from which he selects the person who, he thinks, is best able to advance the chain. Thus we are dealing only with the end product of a radical screening procedure.

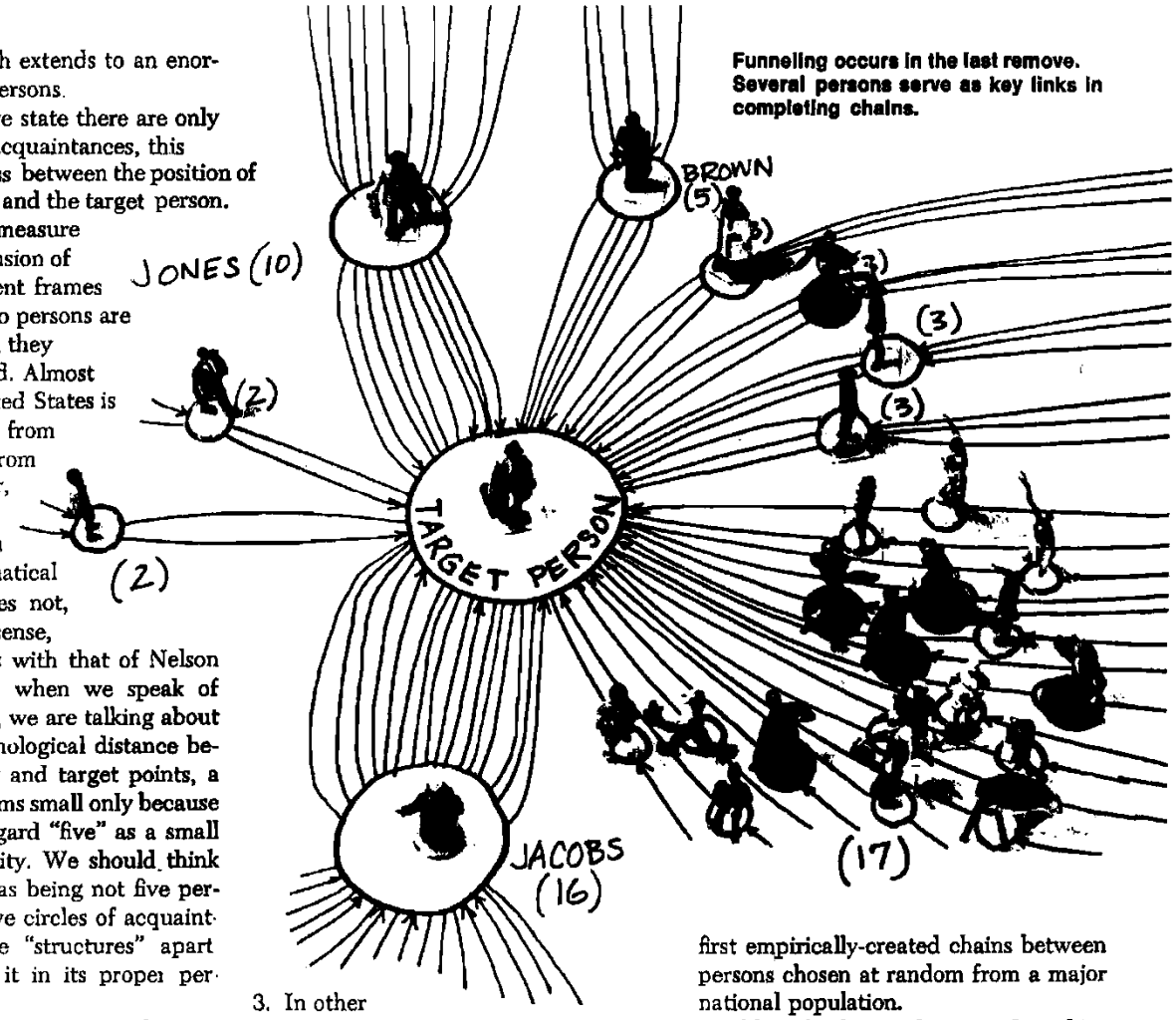
The second thing to remember is that geometric progression is implicit in the search procedure, but nothing is more alien to mathematically untutored intuition than this form of thinking. As youngsters, many of us were asked the question: If you earned a penny a day and the sum were doubled each day, how much would you have earned by the end of a 30-day working period? Most frequently people give answers on the order of $1.87 or $6.45, when in fact the sum is more than $10 million for one 30-day working period, the last day alone yielding $5,368,709.12. Elements of geometric progression with an increase rate far more powerful than mere doubling underlie the small-world search procedure, and thus, with only a few

removes, the search extends to an enormous number of persons.

Finally, when we state there are only five intermediate acquaintances, this connotes a closeness between the position of the starting person and the target person. But this is in large measure misleading, a confusion of two entirely different frames of reference. If two persons are five removes apart, they are far apart indeed. Almost anyone in the United States is but a few removes from the President, or from Nelson Rockefeller, but this is true only in terms of a particular mathematical viewpoint and does not, in any practical sense, integrate our lives with that of Nelson Rockefeller. Thus, when we speak of five intermediaries, we are talking about an enormous psychological distance between the starting and target points, a distance which seems small only because we customarily regard "five" as a small manageable quantity. We should think of the two points as being not five persons apart, but "five circles of acquaintances" apart – five "structures" apart This helps to set it in its proper perspective.

There is a very interesting theorem based on the model of the small world It states that if two persons from two different populations cannot make contact, then no one within the entire population in which each is embedded can make contact with any person in the other population. In other words, if a particular person, *a*, embedded in population A (which consists of his circle of acquaintances), cannot make contact with a particular person, *b*, embedded in population B, then:

1. No other person in A can make contact with *b*.
2. No other person in A can make contact with any other person in B.
3. In other words, the two sub-populations are compeltely isolated from each other Conceivably, this could happen if one of the populations were on an island never visited by the outside world In principle, any person in the United States can be contacted by any other in relatively few steps, unless one of them is a complete and total hermit, and then he could not be contacted at all.



Funneling occurs in the last remove. Several persons serve as key links in completing chains.

In sum, perhaps the most important accomplishment of the research described here is this: Although people have talked about the small world problem, and have even theorized about it, this study achieved, as far as I know, the first empirically-created chains between persons chosen at random from a major national population.

Although the study started with a specific set of questions arising from the small-world problem, the procedure illuminates a far wider set of topics. It reveals a potential communication structure whose sociological characteristics have yet to be exposed. When we understand the structure of this potential communication net, we shall understand a good deal more about the integration of society in general. While many studies in social science show how the individual is alienated and cut off from the rest of society, this study demonstrates that, in some sense, we are all bound together in a tightly knit social fabric. ☐

The chains progress from the starting position (Omaha) to the target area (Boston) with each remove. Diagram shows the number of miles from the target area, with the distance of each remove averaged over completed and uncompleted chains.