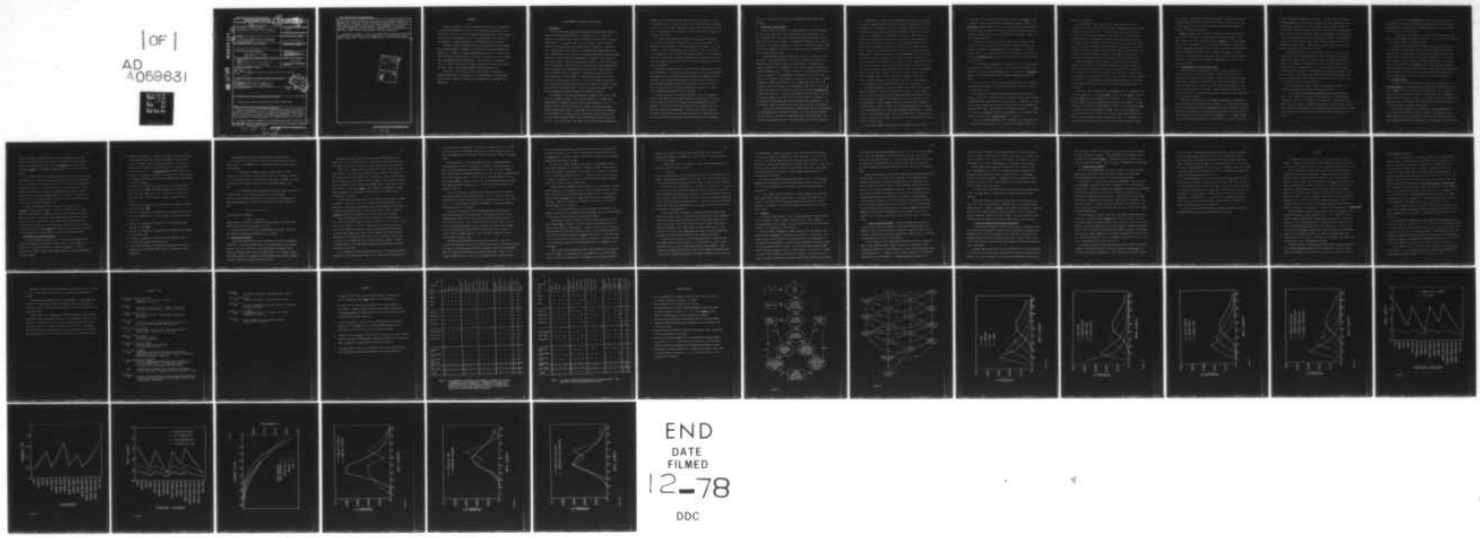
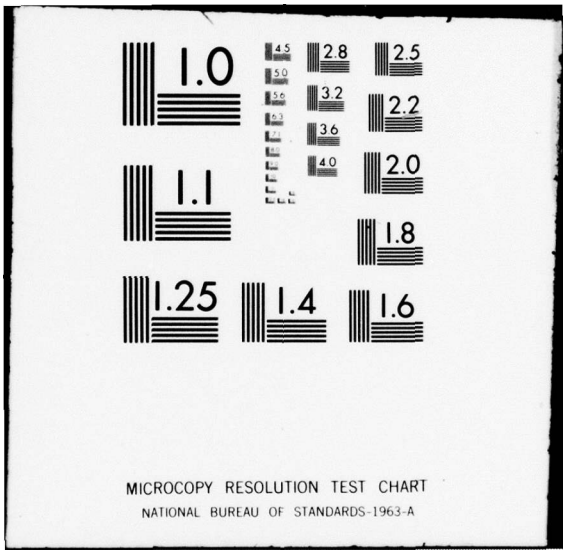
AD-A059 631 WEST VIRGINIA UNIV MORGANTOWN F/G 12/2
A PSEUDOMODEL OF THE SMALL WORLD PROBLEM.(U)
AUG 78 P D KILLWORTH, H R BERNARD N00014-75-C-0441
UNCLASSIFIED KB-117-78 NL



END
DATE FILMED
12-78
DDC


1.0
1.1
1.25
4.5
5.0
5.6
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
1.4
1.6
MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

SECURITY CLASSIFICATION OF THIS PAGE (When Date Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: KB-117-78	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): A Pseudomodel of the Small World Problem	5. TYPE OF REPORT & PERIOD COVERED: interim
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Peter D. Killworth (University of Cambridge); H. Russell Bernard (West Virginia University)	8. CONTRACT OR GRANT NUMBER(s): N000014-75-C-0441-P00001
9. PERFORMING ORGANIZATION NAME AND ADDRESS: ONR, Code 452, Arlington, VA 22217	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: West Virginia University, Morgantown, WV 26506	12. REPORT DATE: August, 1978
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Interim repts	15. SECURITY CLASS. (of this report): unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

46 p

Approved for public release, distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

N000014-75-C-0441

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

small world, networks, communication networks, Markov model

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A model is presented of the decision-making process used by intermediaries in small world experiments in the U.S. This involves allocating each of the population of the U.S. to one of 16 categories; the membership of each population is a function of the target in the small world experiment. This is shown to be equivalent, for modeling purposes, to a Markov process with 16 states.

The Markov transition probabilities are derived partly from reverse small → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Date Entered)

world data and partly by guesswork, but using as few disposable parameters as possible (3). Statistics of chain lengths from various types of starter (e.g. those far from the target, those in the target's occupation, etc.) are derived and compared favorably with observations. The possibility of incompleted chains is included by allowing a constant probability of loss at every step in the chain. Again, there is good agreement with most observations.

A discussion is given as to how such a model might be validated by suitable observations; in particular, a set of experiments is described which should produce a great deal of additional information about the small world experiment.

ABSTRACT

A model is presented of the decision-making process used by intermediaries in small world experiments in the U.S. This involves allocating each of the population of the U.S. to one of 16 categories; the membership of each population is a function of the target in the small world experiment. This is shown to be equivalent, for modeling purposes, to a Markov process with 16 states.

The Markov transition probabilities are derived partly from reverse small world data and partly by guesswork, but using as few disposable parameters as possible (3). Statistics of chain lengths from various types of starter (e.g. those far from the target, those in the target's occupation, etc.) are derived, and compared favorably with observations. The possibility of incompleted chains is included by allowing a constant probability of loss at every step in the chain. Again, there is good agreement with most observations.

A discussion is given as to how such a model might be validated by suitable observations; in particular, a set of experiments is described which should produce a great deal of additional information about the small world experiment.

A PSEUDOMODEL OF THE SMALL WORLD PROBLEM

I. Introduction

The Small World (SW) technique was invented by Stanley Milgram (1967) as a means of gathering data about social relations in the U.S. Briefly, a Starter (S) is asked to send a folder to a Target (T) if and only if S "knows T personally." If S does not know T, then S is asked to send the folder to someone he or she knows who "has a better chance" of knowing T. Milgram and his associates used the SW technique to study mean chain lengths between randomly chosen persons in the U.S. as well as between persons who were deliberately chosen as Ss and Ts because they were members of different races, classes, or occupational groups. Others have used the SW technique to study organizational structure (Lundberg, 1975); and several theoretical papers have appeared which treat the mathematical problems associated with the basic SW goal, namely to find out whether everyone is connected to everyone else in the world by a path through a set of links, where the links are other people, who are also connected to everyone else... and so on. The first such paper was written 20 years ago by Pool and Kochen, and has enjoyed an "invisible college" following ever since (Pool and Kochen, 1978). Indeed, it was Pool and Kochen's paper which led Milgram to develop his famous data gathering technique.

The SW method has provided some important information about social structure. For example, as hoped, it has produced statistics on the mean chain length between individuals and across socioeconomic boundaries in the U.S. An extensive review of the SW literature and findings is presented elsewhere (Bernard and Killworth, 1978). In this paper we try to extend the usefulness of the SW method by a) drawing on what is known about social structure; and b) constructing a plausible model of how a folder travels between links in a SW chain. We will postulate a plausible list of criteria which SW experiment

participants use in selecting their choice of person to whom they will send the proverbial folder. This list is based on our findings from an experiment we conducted on the "reverse small-world problem" (RSW, Killworth and Bernard, 1978). In other words, we will present a simple algorithm for moving along the chain, from S to T.

The form of this algorithm is equivalent to placing all the people in the U.S. into a collection of categories, for a given SW experiment. The membership of each category is a function purely of step target. Under this assumption, a second SW experiment, with a different T would produce identical categories, but with different memberships. Because of this assumption, the movement of the folder in any SW experiment can be modeled, on the macro scale, as a Markov process, with "states" replacing "categories" as the unit under investigation.

The transition matrix for such a process must be obtained both from data and from guesswork. The data from the RSW experiment provide some information; the arguments of Pool and Kochen (1978) provide more; but guesswork remains necessary. Now, the Markovian assumption is only a model, and models with many adjustable parameters can trivially be tuned to fit the real world. Hence, we have used as few disposable parameters as possible, in order to provide something approaching a test of the model.

Although our model fits the extant date well, a real test remains to be done. For example, we do not know whether the categories of choice which we have created are the "real" categories used by participants in a SW experiment. We will show how the data necessary for a stringent test of the model may be acquired experimentally. We hope that this will extend the usefulness of the SW method by allowing predictions to be made about other observable phenomena from the data it produces. We hope, in other words, to

be able to use data about SW chains to say something concrete about social structure.

II. Operational Decision Making

In this section, we consider how a person anywhere in a SW chain chooses the next link in that chain. Some information is available already. The work of Milgram and his associates (Travers and Milgram, 1969; Korte and Milgram, 1970) has shown that there is a tendency for choosing the next link on the basis of the target's occupation -- at least for stockbroker targets.[1] They also found a tendency for each link to be nearer (at least in geographic distance) than all previous ones.

The RSW experiment added more information. Except for targets which are perceived to be in some way "near" the starter, location and occupation are confirmed as the overriding reasons for choosing links. Furthermore, the higher the status of T's occupation, the more likely occupation is used as a reason for making a choice. In fact, a linear fit can be made (accounting for 37% of the variance) to the probability of choosing on the basis of either location or occupation. This depends on the size of T's town (small or large), distance of T's town from the starter (in a neighboring state or not), and occupation level of the target on the Duncan-Reiss (1961) scale. Finally, the RSW data show that on many occasions Ss made choices who were associated with a particular occupation or location, even though the choices neither practiced T's precise occupation, nor lived in T's precise location. For example, our respondents often explained that they were choosing their pharmacist as a link to a medical technician, because they did not know any nurses or physicians.[2]

In order to use the RSW information, we shall assume that all intermediaries in a SW experiment can be treated as Starters, i.e., they are not concerned with the history of the chain in making their choice of the next link. This includes the particular case of intermediaries who are in the same occupation

as T. Our assumption is that these intermediaries are no more or less likely to make an occupation-oriented choice than any other intermediaries. In RSW we found that the probability (over 58 Ss) of making occupation choices for stockbroker Ts was identical to Travers and Milgram's findings for SW chains to an actual stockbroker. In other words, there is no evidence (yet), on whether an intermediary's occupation has any bearing on whether his choice will be made on the basis of occupation. Of course, intuitively, one feels that lawyers, for example, would be very likely to choose another lawyer to get to a lawyer T, but that this would not be the case for, say, carpenters-to-carpenters. However, this remains to be demonstrated; hence our assumption that chain history is irrelevant. This allows the RSW data to be applied to each intermediary link as well as to the starter. We also assume that Ss make their choice of the next link in a SW experiment purely on the basis of location and/or occupation. Two important cases are excluded from this assumption. The first case is when S knows T. The second is when S believes that the choice he makes knows T. We term this choice the "penultimate link." Of course, S may be mistaken: the "penultimate link" may not actually know T. These assumptions will be discussed critically in the conclusions.

With these assumptions, how does S choose his next link? The mechanism which we postulate is shown in flow-chart form in Fig. 1. We must stress that this stems specifically from one single--and plausible--assumption, that Ss do the best job they can of choosing their next link. (However, even if this mechanism is totally incorrect, this will not invalidate the model's results, as will be seen). In Fig. 1, S first decides whether he or she knows T or a penultimate link; if S does, then he chooses appropriately. If S knows neither T nor a penultimate link, then the next best choice is someone who is "near T" and who is in the same occupation as T. (An operational definition of "near T" is given below).

If S can not combine both occupation and location in selection of an

intermediary, then a decision is made whether to choose from either of these two domains. We do not know how such a decision is reached. However, the probability of making either decision is fairly well known from the RSW data, so for this model we assume that a decision is made with probability P_ℓ of choosing on location, and $P_o = (1-P_\ell)$ of choosing on occupation. (P_ℓ is obtained from the RSW data).

If S decides to choose on the basis of location, he attempts to "get closer" to T. It is not obvious that "closer" is necessarily related to geographic distance, but we assume that is is for simplicity. If S knows someone close to T, he chooses the closest such individual. If not, he looks for someone associated with T's location, and chooses the most likely person as the next link.

As in the pharmacist example above, informants often told us they didn't "know anyone in California" but their uncle or friend "used to work in San Francisco." The uncle or friend might now live in Iowa, but the association with a T in California seemed to be the reason for making the particular choice.

If all else fails, S chooses a "more likely" person who is at least not further away from T than S is; typically this would be someone who "knows more people than S does."

If S chooses on the basis of occupation, a similar set of decisions follows. Thus, if T is a dentist, and if S chooses a dentist, then S chooses the geographically closest dentist to T that he knows. S may well, for reasons only he knows, choose a dentist further from T than S is; but presumably not too much further away. (For clarity, this is omitted from Fig. 1). As above, is S chooses on occupation, but does not know anyone in T's occupation, then someone is selected who is associated with T's occupation.

Finally, the same "desperation move" is available to S when choosing on occupation as when choosing on location: S may choose someone nearby who

"knows a lot of people."

In setting up this model of the decision making process, the role of occupation as a reason for choice has been made reasonably operational, whereas distance and "near T" have not. In RSW the simple concept of near vs. far was used in order to make geographical distance discrete. "Near" meant "in bordering states" and "far" meant "everywhere else." Since SW chains will, in general, reach to the center of large cities as well as to very small towns, such a crude measure as near vs. far is probably not very useful. Two people living a mile apart in a small town are far more likely to know each other than two people living a mile apart in Chicago. This suggests that population density is a more relevant measure for distance, and will be adopted here.

Clearly, population density (however it is defined) is a continuous variable. For esthetic, sociological and mathematical reasons, however, we have divided the population of the U.S. into discrete sizes.[3] We argue that most of the U.S. is "far" (on any measure) from a specific T. However, many people do live in the same state or city or town as a given T, and are therefore "closer" to T. In fact, Travers and Milgram (1969) found that Ss in the same city (Boston) as T were one link closer to T than Ss who were "far away" (in Nebraska).

We choose to split distance into four categories based on population size, as follows: far from T; in the same region as T; in the same area as T; and in the same town as T, in increasing order of proximity to T. The relevant population sizes of these four categories must (eventually) be guessed. Let far be taken to include most of the population of the U.S. (2×10^8), and let town be defined here to be a population of 10,000. The U.S. Census Bureau uses 2,500 as the cutoff for an "urban" area. Most observers would probably find a town of, say, 3,000 population to be anything but urban. The average size of a census tract in large cities in the U.S. (over 50,000 population) is about

4,000 persons. The model we present below assumes that cities are collections of neighborhoods (here called "towns") each of which is taken to be of order 10,000 population. As it turns out, our model is relatively insensitive to changes in this lower bound value.

Region and area sizes are derived by assuming a constant multiplicative factor between the population of each category and the next higher. This factor turns out to be 27, giving the size of a region as 7.4×10^6 and area as 2.7×10^5. Conveniently, 7.4×10^6 is about the population of New York City, or Washington and Oregon, or North and South Carolina, or Massachusettes, Maine, New Hampshire and Vermont, etc. Similarly, the area size is about the size of a large suburb or a small city. Again, these population sizes are insensitive to the precise size assumed for town population, because of the cube root involved.

III. A Representation of the Decision Process

We can now transform the flow-chart of Fig. 1 into a picture of the likely movements of a folder in a SW chain. The population of the U.S. is divided into 17 categories, on the basis of location (far, region, area, and town); plus a collection of penultimate links. The likely movements of the folder are shown in Fig. 2, which reflect the likely choice patterns in Fig. 1. Other choices are possible: for example, someone far may know someone in the same area as T; this is unlikely, but possible -- after all, this is what makes the SW problem of interest.

These other possibilities are too complex to represent on a diagram, but they are easily seen in matrix form, as in Table 1. All likely moves are coded as L. The unlikely moves have been subdivided into three categories. The first consists of those which are so unlikely that they almost never occur, such as a far S knowing a link in the same town and occupation as T. Although this can occur, we set such probabilities to zero arbitrarily, in order to reduce the

number of disposable parameters in the model. Of course, people from Boston and Omaha do meet in a cafe in Paris, and do find out that they have a common acquaintance in San Francisco. This is what led Pool and Kochen (1978) to their classic reasoning about the SW problem. However, the overwhelming number of times that such discoveries do not take place must be the more relevant fact for understanding social structure, and must therefore be represented in a model of social structure. Hence, for parsimony, extremely unlikely probabilities are set to zero. Setting them to a small finite value like .001 makes no noticeable difference to the results.

The second and third of the unlikely types of moves in a SW chain occur sufficiently often to warrant attention. Some of these moves are "helpful" because they actually advance the folder closer to T; these are coded as H in Table 1. For example, it is possible that a far S may know someone in the same area (2.7×10^5 population) as T. If this is the case, S will almost certainly choose that person. The remainder of the moves are unhelpful, coded as U in Table 1. Such moves include what we have called "desperation moves," as well as moves which send the folder farther from T than S in order for S to use occupation as the criterion of choice.

Most of the pattern in Table 1 should now be apparent; however, some further observations on our choice of entries are in order.

1.) We have arranged the matrix as symmetrically as intuition would allow. Note, though, the omission of the ULH pattern for Ss associated with location and choices in T's occupation. We feel that an S associated with a region is just as likely, if choosing on occupation, to make a choice in either far-occupation or region-occupation. Lacking information about such "associates," there seems little to be done except to assign the same probability to both choices.

2.) An S in a particular geographic location (e.g. region, area, etc.) is assumed not to choose a person "associated with" that location, even as a desperation move. A person living in North Carolina will not choose someone who "used to work in North Carolina."

Similarly, S would generally not choose a person "associated with" his own location and the the target's occupation. If S is "far" from T, however, we allow such a choice as a desperation move.

3.) There are many probabilities connected with penultimate links and Ts. Obviously, most are zero. Once an S is in the same town (10,000 population) as T, it is very likely (see below) that he knows at least a penultimate link. A person in the same town and occupation as T is also quite likely to know T. Note, finally, that because the penultimate link is only defined to be someone who is perceived as such by the previous S, several moves are actually possible from such people: 1) penultimate links may genuinely know the target; 2) they may think they know other penultimate links; or 3) they may simply choose others in the same town, or town and occupation.

IV. The Markov Process

It should be clear by this stage that Table 1 can be thought of as a transition matrix P in a Markov process, in which the 17 categories of Ss are the 17 states in which the process may be at any stage. Such a concept is not new in the SW literature. Indeed, it forms the basis of a fundamental paper by Hunter and Shotland (1974). In that paper, they divided Michigan State University into 11 categories, defined by the ordinary university structure from administration to freshmen undergraduates.

The transition probabilities were obtained from SW experiment data collected by Shotland (1976). Using an extension of the Kemeny and Snell (1960) analysis, Hunter and Shotland were able to predict means and s.d. of path lengths between any pair of categories in the university. The means were,

in fact, predicted very well; but the s.d. were far higher than observed. Hunter and Shotland explain the discrepancy by the fact that individual networks are not necessarily within or between categories, but rather between or amongst friends. This problem presumably becomes more acute as the folder nears T.

The difference between the Michigan State study and ours is: in the former, the membership of the categories was defined a priori, whereas the membership of our 17 categories is purely a function of T. In other words, changing the occupation or location of T directly affects the choices made by all Ss, and therefore the membership of, say, the region-occupation category. At the local level, the change in category membership is greater for a change in T, because of such local linkage factors as ethnicity, hobbies, religion, political activity, schooling, children, etc.

Although the membership of each category is a function of T, the existence of each category is not. Of course, the probability of a folder going from one category to another is a function of T (although, we believe, not a strong one). As a result, the mean path lengths to enormous numbers of Ts will turn out to be identical, purely because the transition matrices P are identical (or nearly so). Thus many people in the U.S. are, by our model, identical insofar as SW path lengths are concerned, even though the motivations and reasoning behind the choices made by each S are violently different.

V. The Entries in the Transition Matrix

It is now necessary to translate the U, L and H entries in Table 1 into specific probabilities of transition from one state to another (not the probabilities of an S knowing someone in a specific category). In order to avoid having to guess too many of the probabilities, nine explicit assumptions are made, namely:

1) A likely (L) probability of transition from any state to any location or location-associated state has the value α(α is constant within rows but not columns and will be determined below). We assume, in other words, that there are equal probabilities of choosing either a link closer to T, or a link associated with somewhere closer to T.

2) A likely (L) probability of transition from any state to any occupation or occupation-associated state has the value β (again constant within rows but not columns).

3) An unhelpful (U) probability of transition to a location-oriented state has the value $\gamma\alpha$, where γ is an empirical constant whose value is normally taken to be 0.05. In other words, on about 1 in 40 occasions when a location choice is made, a desperation choice occurs.

4) An unhelpful (U) probability of transition to an occupation-oriented state has the value $\gamma\beta$.

5) A helpful (H) probability of transition to a location-oriented state has the value αp, where p is another empirical constant, also taken to be 0.05.

6) A helpful (H) probability of transition to an occupation-oriented state has the value βp.

7) A helpful (H) probability of transition to penultimate link or target has a probability γ.

8) A likely (L) probability of transition to the target is e, where e is an empirical constant guessed to be 0.6.

9) A likely (L) probability of transition to penultimate link is e if transition to T has probability γ, and e/2 if transition to T has probability e.

These assumptions produce the transition matrix shown in Table 2.

Note that the α and β entries in each row of the matrix may be found by using the fit to the probability of choosing on the basis of location, given in RSW, namely

$$P = 0.056 \text{ size} - 0.0031 \text{ occupation} - 0.092 \text{ distance} + 0.818$$

where size = 1 if T lives in a well known big city, and 2 otherwise: distance = 1 if T is local and 2 if T is far (local is here taken to include town, area, and region); and occupation is defined on the Duncan-Reiss (1961) scale. Then

Σ (all probabilities of transitions from a given state to a location-oriented state) = P_ℓ [1 - prob(penultimate link) - prob(target)]

where the latter probabilities are those of transition to penultimate link and target respectively. Thus, for the first row, (i.e. transitions from far)

$$\gamma\alpha + \alpha + \alpha\rho + \alpha + \alpha\rho = P_\ell$$

and similarly for occupation

$$\beta + \beta\rho + \beta + \beta\rho = P_o$$

define α and β for that row. Similar calculations apply to the other rows. Typical values of α and β are about 0.2 to 0.3.

Numerical justification of the probabilities P_{ij} in Table 2 is given in the Appendix, although much guesswork is still required.

VI. Results From the Model

Results from any Markov model are best presented in terms of observable quantities. For the SW problem, these are the distributions of path lengths from Ss to Ts, since these have already been measured by Milgram (1967), Travers and Milgram (1969), Korte and Milgram (1970), and Lin et al. (1978). The relevant formulae for computing the model predictions are given by Kemeny and Snell (1960) and by Hunter and Shotland (1974).

The probability that a path from a given starter category will be of length 1, 2, ..., 10 is shown in Fig. 3. The results in our model depend only very weakly on the size of T's town or on T's occupation level on the Duncan (1961) scale. Making a very large change to T (from a small town, occupation level 10, to a large town, occupation level 90) causes a decrease in path length of order 0.05 intermediaries. Hence Fig. 3, and all other results, refer uniformly to a T in a large town with occupation level 90--the archetypal "Boston stockbroker" used by Travers and Milgram (1969). (Note that in Fig. 3, and elsewhere, we count path length, and not number of intermediaries, which seems customary in the experimental literature; hence to compute number of intermediaries, subtract 1).

The rapid decrease in path length as the starter category nears the target is apparent. These results can not be compared directly with Travers and Milgram; this is because the probabilities of actual completion are much lower in Travers and Milgram, due to the high attrition rate. However, judged as relative probabilities of completion, a limited comparison is possible.

Travers and Milgram had three starter categories: 1) Nebraska random (i.e. "far" in our terminology); 2) Nebraska stockholders (i.e. far-occupation); and 3) Boston random (i.e. region). Comparison of our theoretical results in these three categories with their actual data shows agreement, although our calculated path lengths tend to be too long. Comparison with Korte and Milgram (1970) is more difficult; while all of their Ss are "far" from T, there is no way to tell whether any are in our far-occupation category. However, the general shape of the chain length distribution again agrees well with the predictions of our model.

Fig. 4 summarizes the path length statistics from our 16 different starter categories. The predicted path lengths for far, region, and far-occupation Ss

are 7.28, 5.62, 6.57 respectively. These may be compared with 6.72, 5.40, and 6.42 found by Travers and Milgram (1969); 6.61 for the far category found by Korte and Milgram (1970); and 5.04 for a locale of 7.5×10^5 found by Lin et al, (1978).

With the exception of the far category, which is, as previously noted, overestimated by the model, the agreement is excellent. The spread about the mean is also well-predicted, with modeled s.d. of 1.98, 1.78, 1.93, compared with observed values from Travers and Milgram of 1.73, 2.32, and 1.79 respectively, and 2.07 by Lin et al. (1978). Only the s.d. of the Boston random Ss is seriously underestimated. The rest of Fig. 4 is a set of testable predictions about path lengths from other categories of Ss. How such tests may be made is considered in section 7.

It is, of course, quite straightforward to ensure a good fit to data by adjusting the entries in the transition matrix. It was in order to avoid this trivializing of the model that only four free parameters were used in P_{ij}. However, it is still necessary to estimate how dependent are the model results on the specific entries in the matrix.

To do this, a collection of statistical manipulations were performed, each testing sensitivity in different ways. The crudest such test involved changing each non-zero entry in P_{ij} by $\pm$ 20% in a random fashion (followed by a division of P_{ij} by P_{ij} to restore row sums to unity). This led to an increase in mean path lengths of only 0.1 (further iterations of this randomising having successively weaker effects; the mean path from far Ss was, of course, the most strongly affected). Hence the model is <u>not</u> very sensitive to gross random changes in P_{ij}.

Extremely small probabilities had been omitted a priori. They can be included by allocating a random value between 0.0 and 0.05 to all entries initially less than 0.01; most mean path lengths are increased by about one intermediary, except for paths from far Ss, which actually become shorter. There are two (linked) reasons for this. First, this procedure produces a general evening

out of the transition probabilities towards the system in which transition to any state is equally likely. Second, it becomes more likely that folders may move away from T. Therefore, accurate modeling of the SW means that very small probabilities must remain very small.

The next tests involve systematic modification of the parameters rather than random "adjustment" of the entries. Reducing p and γ to zero merely increased path lengths, and decreased s.d.s, by about 0.1. Similarly, increasing p to 0.1 reduced path lengths by about 0.1. Hence, the model is not sensitive to changes in p and γ ; in other words, most of the time, folders go through likely paths, and not through unlikely ones.

The most important parameter in the system is e (i.e. the chance of certain states knowing T). Increasing e from 0.6 to 0.66 (the maximum value it could take in this model) only decreased mean path lengths by 0.2. As long as penultimate links are likely to know T, how likely they are is not particularly important. Reducing e, however, to 0.2 makes it very difficult to reach T, and path lengths soar to about 12.

Thus, in the entire formulation, only two quantities are essential: the number of transition states between far and penultimate link (or town-occupation), and the chance of the penultimate link knowing the target.

Everything we have said so far involves the "life-or-death" SW problem, where folders must be passed on. The real world of SW experiments is not so pure, however, because folders get lost. A model of social structure based on the SW concept can, and should, ignore attrition. This is because the SW technique itself is trying to gather data about the real world. On the other hand, it is a trivial matter to include attrition in the model, if one wants to model the data which are obtained from SW experiments (but not, presumably, social structure).

White (1970), Hunter and Shotland (1974), and Feinberg and Lee (1975) all make the simplest null hypothesis, that the probability of the folder being lost at any step is a constant value, independent of both category and path

history. White (1970) finds that a loss rate of 25% fits the Travers and Milgram data reasonably, although there is some evidence of variation of loss rate with path length.

If an 18th state ("lost") is permitted, with a constant probability K of transition from any state i, so that

$$P_{i,18} = K$$

and the other probabilities P_{ij} are scaled down by (1 - K), an equivalent 18 x 18 matrix is produced which includes the loss state. (This is the precise inverse of the procedure used by Hunter and Shotland, 1974, to remove the loss category).

This adds one more free parameter to the system. In order to fix its value, we use the result from Travers and Milgram that only 18% of chains originating from the Nebraska random group were completed. (As noted by Lundberg, 1974, the correct figure is 18%, and not 24% as stated by Travers and Milgram who neglected the 79 folders which never got beyond the Ss). A value of K of 0.22 (c.f. White's estimate of 0.25) gives a completion rate, for far Ss, of the 18% observed; for our purposes, this determines K.

A test of the model is therefore to compare loss rates from other populations, as shown in Fig. 5. Chains with far-occupation Ss (such as Nebraska stockholders) have a predicted completion rate of 22%, compared with the observed value of 24%. Chains with regional Ss (such as Boston random) have a predicted completion rate of 27%, compared with the observed value of 22%. The latter fit is poor, as are many predictions about the Boston random group. In section 7 we discuss what we believe to be causing this problem.

The effect of attrition is to reduce the mean length of completed paths (the longer the path, the more likely it is to terminate in a loss). Fig. 6 shows mean and s.d. of path lengths (both complete and incomplete). Mean complete path lengths from far, far-occupation, and region Ss now become 6.48, 5.82, and 4.99, compared with Travers and Milgram's findings of 6.72, 6.42, and

5.40 respectively. Predicted s.ds are a little lower than observed, by about 0.1, except for Boston (i.e. regional) Ss, where the model's s.d. is seriously in error. (Of the three means and s.ds in the extant data, the Boston s.d. is the only one which differs significantly (.05 level) from the model's predictions).[4] Of course, since complete chains are a statistically rare event, a vast number of SW starters would be necessary to provide enough complete chains for a stringent test of any model.

The pattern of predicted incomplete path lengths (Fig. 7) is in excellent agreement with observations; note particularly that, provided Ss are suitably "far" from T, the probability of termination is not particularly dependent on category of S.

A more rigorous test of the model is a direct comparison of predictions about complete path lengths (when attrition is included) with extant data, as in Fig. 8. Not that far and far-occupation path distributions are in good agreement with the data. As usual, the comparison with regional (i.e. Boston) data is poor.

VII. Critique

1) Why use a Markov model? There is a great temptation to describe many social processes in terms of a Markov transition model. Markov theory is well-understood and easy to apply. But this does not mean that such models are relevant for a description of the SW process.

One of the basic assumptions of Markov theory is that the transition probabilities are independent of the history of the process. But the only circumstance in which the full history of the SW process is of any importance is in the conduct of SW experiments! All other network processes, we believe, involve at best the previous link in a chain. It is true that rumors sometimes begin with "I heard from so-and-so that ...." And one may, in fact, choose to squelch the rumor if the credibility of "so-and-so" is suspect. But where does this stop? We choose to assume that, in general, only a very limited (if any) history

is important for understanding communication flow. This is, of course, testable with a set of SW experiments, one which provides Ss with the chain history (as usual), and one which does not. It is entirely possible that attrition may be reduced by not providing the chain history. Long chains (if known) may frustrate intermediaries and make them more inclined to drop out of the experiment.

Another serious objection to the Markov hypothesis was raised by Hunter and Shotland (1974), in a critique of their own model. They argued that their categories and associated transition probabilities did not represent how Ss actually made their choices. They felt that the Markov model, while parsimonious, did not reflect psychological realities. This is probably true, but if one wants to describe the behavior in a social process (rather than the presumed psychological forces which drive the process) then psychological realities may be irrelevant. On the other hand we believe that the 16 categories of our model <u>are</u> relevant to how an S makes a choice in a SW process: both because RSW provides some evidence of at least the <u>types</u> of category we suggest, and also because the membership of each category is not a static quantity, but a function of the target. How this can be tested is discussed below.

2) <u>Why not use a simple model</u>? It might be argued that the degree of complexity of our model is too great; a much simpler Markovian model would generate statistics that, with suitable parameter tuning, would still fit the data well. Specifically, why should location and occupation, both of target and of intermediaries be retained in the theory?

One can construct a very simple model in which the folder is presumed to move in four preordained steps (region - area - town - penultimate link), followed by transition to T with probability e, and to another penultimate link with probability 1 - e. Taking e to be 0.4 gives a mean and s.d. of the path length of 6.5 and 1.9 respectively. Although these are very good fits to

observed results, the shape of the path distribution is radically different from the observed. Furthermore, if a constant loss rate of 0.2 is added, then e must be reduced to 0.25 in order to fit the data. This seems to us to be an unacceptably low value for the probability of a penultimate link knowing T.

Another simple model assumes that the same five steps (region - area - town - penultimate link - T) must be traversed, but with the probability of moving one link down the chain being e at every step (i.e. a random walk which may not proceed backwards). Choosing e to fit the observed mean path length of 6.5 gives e = 0.77. The resulting s.d. of the path length distribution is, again, seriously in error.

It is clear that the SW process must be more complex than the above simple models. But <u>how</u> complex? Is the concept of occupation categories really necessary?

Although RSW showed that target occupation has a strong influence on S's next choice, Section 6 showed that it produced little, if any, effect on path length statistics. However, the <u>types</u> of path (location-oriented vs. occupation-oriented) were a strong function of target occupation. (For low target occupations, hardly any occupation choices are made). We believe, therefore, that occupation must be retained in any adequate description of the SW process.

3) <u>Are the 16 categories in the model appropriate</u>?

Another basic assumption of Markov theory is that its component states are discrete and well-defined. Frankly, we do not know whether categories are independent of S, or even if categories exist. Obviously, we know even less regarding whether four discrete population sizes (far - region - area - town) is the "right" number to use, or even whether modeling "distance" by population size is legitimate.

We suspect that North and South Carolina combined are not really equivalent as a social unit to New York City, despite the fact that their populations are similar. The lack of success of our predictions concerning the Boston starters

may only reflect a lack of fine tuning; or it may reflect a genuine misrepresentation of Boston in the model. Sufficiently varied SW experiments should be capable of distinguishing types of subpopulations, and the quantities on which such definitions depend. Can the concept of "neighborhood" be quantified?

4) Is any of this testable?

One test is straightforward: many numerical predictions have been made regarding SW chains from Ss other than those reported in the literature. Obvious experiments can be concocted to verify these predictions.

The second test is more difficult, but more fundamental. Can one derive a set of meaningful categories from data? The answer is "no," given currently available data. We believe, however, that experiments can and should be performed to acquire data which will either yield (or disprove the existence of) a collection of categories for the SW process. Along the way, of course, much other useful information will be produced. Specifically, a collection of Ss, who may all reside in the same locale, are presented a short (order 100) list of Ts. Some Ts very locally (in the same "town" as the S), some "further away," and some "very far away."

For each T each S is asked to make his or her choice of the next link in A SW chain. Initially, S knows nothing about T (even T's name). S may ask an unlimited number of questions about T until S can make a choice of an intermediary. If the investigator does not know the answer to a particular question about T, then T is called on the phone and the answer is elicited. Preliminary data show that a rather amazing list of queries is generated for local T's (i.e. T's living in the same town as Ss). For example, Ss asked about T's hair color, use of contraception, age of oldest child, etc.

At the end of the first data collection, each question asked is allocated a number. The characteristics of Ss and their choices, corresponding to as many of these questions as possible, is then acquired by further interviewing

of the Ss, as are their reasons for making each choice. Coding the absence or presence of each possible question for each S and T combination as zero or one, respectively, yields many data arrays of the form: Starter information; Target information; questions asked by S; information about choice; and reasons for choice. These arrays may then be factored, for example, to yield such things as number and type of "choice categories," a list of independent questions, the important target characteristics, and so on.

A second experiment could then determine whether knowing all the information solicited in the first experiment about any T would help a second group of Ss make "better" choices of intermediaries. To test this, two SW experiments are conducted. One group of Ss (and all their intermediaries, all the way down the chain to T) is given all the information elicited about T in the first experiment. A control group of Ss is given only the information which they request. Will the chains from the full-information group be significantly shorter than those from the control group?[5]

Clearly, the SW and RSW techniques are capable of yielding a great deal more data about social structure than they have in the past.

APPENDIX

1) Consider first the P_{ij} which do not involve penultimate links or targets. A typical row is the first, namely transitions from the "far" state. Now, there are 27 regions in the U.S., as we have defined them ($2 \times 10^8/7.4 \times 10^6 = 27$). Also, RSW showed that Ss have about 200 "useful" people who are available as choices in a SW experiment. Assume that, say, 100 of these are location choices (the actual figure in RSW was 95) and 100 are occupation choices (c.f. 99 in RSW). Then the probability that all 100 location choices have no connection with a specific region is $(1 - 1/27)^{100}$, which is vanishingly small. (This assumes a random distribution of choices over regions). Similarly, if there are, say, 50 essentially different occupations, the probability that none of the 100 occupation choices have no connection with T's occupation is $(1 - 1/50)^{100} = 0.13$ which is again very small. Thus it is extremely likely that a choice can be made which moves the folder towards T (either by location or by occupation). In Table 2 this probability is $1 - \gamma\alpha \simeq 0.99$.

Movement from far to area is more unlikely. There are 730 areas, and the equivalent calculation yields a probability of $1 - (1 - 1/730)^{100} = 0.13$ of knowing someone connected with T's region (again, this assumes a random distribution). This seems, intuitively, to be much too high; indeed, Pool and Kochen's (1978) arguments about social strata (which can, of course, be reinterpreted as geographical strata) would reduce this probability drastically. Hence we choose, perhaps arbitrarily, to leave this probability as αp (typically about 0.01 to 0.02). The effects of increasing p (and with it such probabilities) are investigated in the text.

Movement from far to town (there are 19,742 towns) is highly unlikely on any structural assumptions; hence the zero probability allocated. Similar plausibility arguments can be given for the other P_{ij} in Table 2, involving

varying degrees of guesswork.

2) It is also necessary to justify the probabilities of reaching penultimate links or T; clearly the model results will depend strongly on how easy or difficult it is to reach T. Consider first the probabilities of reaching T from the "town" state in the matrix. Gurevich's (1961) study showed that an individual "knows", on average, about 500 people locally. In a town of population 10,000 the odds that one of these is T is 0.05 (i.e. γ). However, not all the 500 would be of use as potential penultimate links since many of these are casual acquaintances about whose networks S knows nothing. Thus the number of potentially useful first links reduces to the 100 assumed above. Then, using Pool and Kochen's (1978, p. 29 and p. 33) arguments, with n = 100, N = 10^4, the probability P_1 that two people have at least one common acquaintance, (i.e. that S can choose a correct penultimate link) is 0.63 or 0.59, depending on the argument used. Hence, our value of 0.6 for e.

It was our intuition that, at the local level, knowing a penultimate link would be very likely -- and knowing T would be possible, but not likely. The above arguments led to our selection of 10,000 as the smallest unit of population.

The probability of transition from area to penultimate link again derives from Pool and Kochen. With n = 100, N = 2.7 x 10^5, p_1 = 0.04 (taken again as γ).

When occupation is also involved in defining S, it is even less clear what values of n or N to use for estimation purposes. Using estimates from the U.S. Statistical Abstracts for 1977, there are, on average, about 200,000 people in any given occupation. Assuming an even distribution across the U.S., 7,400 of these are in each of our regions; 270 in each area; and 10 in each town.

Assume that an S in T's occupation knows 20 people in that occupation who are closer to T. Then transition from region-occupation to penultimate link again yields a probability γ (n = 20, N = 7,400, p_1 = 0.05).

Similarly, transition from area-occupation to penultimate link and T yields ($n = 20$, $N = 270$) probabilities 0.73 and 0.07 respectively, taken here as e and γ.

Transitions from penultimate link or town-occupation to penultimate link or target are, from the above figures, extremely likely. We give them the high values in Table 2 without any justification. Obviously, they remain to be tested empirically.

Finally, there is no information yet available regarding transition from "associate" states to penultimate link or T. We use the same device as in the main body of the matrix, which is to allocate the same probabilities as for a non-associate who is one distance unit further removed from T. For example, transition from someone associated with T's town to penultimate link is given the same probability as transition from an S in T's area to penultimate link.

REFERENCES CITED

Bernard, H.R. and P.D. Killworth
1978 A review of the small world literature
CONNECTIONS, Vol. 2, no. 1.

Duncan, O.D.
1961 Socioeconomic Index Scores. In Albert J. Reiss, Jr.,
Occupations and Social Status. New York: Free Press.

Feinberg, S.E. and S.K. Lee
1975 Small world statistics. Psychometrika and supplements.
40:219-228.

Gurevich, N.
1961 The social structure of acquaintanceship networks. Ph.D.
dissertation: MIT, Cambridge, Massachusettes.

Hunter, E. and R.L. Shotland
1974 Treating data collected by the "small world" method as a
Markov process. Social forces. 52:321-332.

Kemeny, J.G. and J.L. Snell
1960 Finite Markov chains,
Princeton: Van Nostrand.

Killworth, P.D. and H.R. Bernard
1978 The reverse small-world experiment.
Social Networks (in press).

Korte, C. and S. Milgram
1970 Acquaintance links between white and negro populations:
application of the small world method. Journal of personality
and social psychology. 15:101-118.

Lin, N., P. Dayton and P. Greenwald
1977 The urban communication network and social stratification: a
"small world" experiment. In: Communication Yearbook I,
R.D. Ruben (ed.), New Brunswick: Transaction Books.

1978 Analyzing the instrumental use of relations in the context
of social structure. Sociological Methods and Research, in press.

Lundberg, C.C.
1975 Patterns of acquaintanceship in society and complex organization:
a comparative study of the small world problem. Pacific socio-
logical review. 18"206-222.

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

Pool, I. and M. Kochen
1978 Contacts and influence. Social Networks. 1:1-48.

Shotland, R.L.
1976 University communication networks: the small world method. New York: John Wiley.

Travers, J. and S. Milgram
1969 An experimental study of the small world problem. Sociometry 32:425-43.

White, C.
1970 Search parameters for the small world problem. Social forces. 49:259-64.

FOOTNOTES

1. Lin et al. (1978), however, show a slight tendency to choose the next link of occupational status higher than T, for middle-class Ts.

2. Of course, sex of T and sex of intermediary are strongly related to sex of choice. However, in what follows we will neglect this because, for modeling purposes, it merely doubles the size of the matrices without contributing any further information.

3. It would be possible to set up a model with population as a continuous variable, but the mathematics of a continuous Markov process would be both harder and less clear as to their sociological meaning.

4. Comparison with Lin et al.'s (1978) data for a locale of 7.5×10^5 -- neither region nor area in our terminology -- shows good agreement for the mean but, again, a significant underestimate by our model for the s.d.

5. Lin et al. (1977) examine a very restricted form of this question (using race and occupation only) but unfortunately do not give any results.

To Category / From	Far	Region	Area	Town	Far-Occn	Region-Occn	Area-Occn	Town-Occn	Assoc-Region	Assoc-Area	Assoc-Town	Assoc-Far Occn	Assoc-Region Occn	Assoc-Area Occn	Assoc-Town Occn	Pen. Link	Target
Far	U	L	H	O	L	H	O	O	L	H	O	L	H	O	O	O	O
Region	O	U	L	H	U	L	H	O	O	L	H	O	O	H	O	O	O
Area	O	O	U	L	O	U	L	H	O	O	L	O	O	O	H	H	O
Town	O	O	O	U	O	O	U	L	O	O	O	O	O	O	O	L	H
Far-Occn	U	L	H	O	U	L	H	O	L	H	O	O	L	H	O	O	O
Region-Occn	O	U	L	H	O	U	L	H	O	L	H	O	O	L	H	H	O
Area-Occn	O	O	U	L	O	O	U	L	O	O	L	O	O	O	L	L	H
Town-Occn	O	O	O	U	O	O	O	U	O	O	O	O	O	O	O	L	L
Assoc-Region	U	L	H	O	L	L	H	O	O	L	H	O	O	H	O	O	O
Assoc-Area	O	U	L	H	O	L	L	H	O	O	L	O	O	O	H	O	O
Assoc-Town	O	O	U	L	O	O	L	L	O	O	O	O	O	O	O	H	O
Assoc-Far Occn	U	L	H	O	L	H	O	O	L	H	O	O	L	H	O	O	O
Assoc-Region Occn	O	U	L	H	U	L	H	O	O	L	H	O	O	L	H	O	O
Assoc-Area Occn	O	O	U	L	O	U	L	H	O	O	L	O	O	O	L	H	O
Assoc-Town Occn	O	O	O	U	O	O	U	L	O	O	O	O	O	O	O	L	H
Pen. Link	O	O	O	U	O	O	O	U	O	O	O	O	O	O	O	L	L
Target	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	1

Table 1. **The likelihood of transition from category to category in the model. L represents a likely probability; H represents an unlikely, but helpful probability; U represents an unlikely and unhelpful probability; and O represents a probability that is so unlikely that it can be set to zero without loss of generality.**

To Category From	Far	Region	Area	Town	Far-Occn	Region-Occn	Area-Occn	Town-Occn	Assoc-Regic	Assoc-Area	Assoc-Town	Assoc-Far Occn	Assoc-Regic Occn	Assoc-Area Occn	Assoc-Town Occn	Pen. Link	Target
Far	$\gamma\alpha$	α	αp	0	β	βp	0	0	α	αp	0	β	βp	0	0	0	0
Region	0	$\gamma\alpha$	α	αp	$\gamma\beta$	β	βp	0	0	α	αp	0	0	βp	0	0	0
Area	0	0	$\gamma\alpha$	α	0	$\gamma\beta$	β	βp	0	0	α	0	0	0	βp	γ	0
Town	0	0	0	$\gamma\alpha$	0	0	$\gamma\beta$	β	0	0	0	0	0	0	0	e	γ
Far-Occn	$\gamma\alpha$	α	αp	0	$\gamma\beta$	β	βp	0	α	αp	0	0	β	βp	0	0	0
Region-Occn	0	$\gamma\alpha$	α	γp	0	$\gamma\beta$	β	βp	0	α	αp	0	0	β	βp	γ	0
Area-Occn	0	0	$\gamma\alpha$	α	0	0	$\gamma\beta$	β	0	0	α	0	0	0	β	e	γ
Town-Occn	0	0	0	$\gamma\alpha$	0	0	0	$\gamma\beta$	0	0	0	0	0	0	0	e	e
Assoc-Region	$\gamma\alpha$	α	αp	0	β	β	βp	0	0	α	αp	0	0	βp	0	0	0
Assoc-Area	0	$\gamma\alpha$	α	αp	0	β	β	βp	0	0	α	0	0	0	βp	0	0
Assoc-Town	0	0	$\gamma\alpha$	α	0	0	β	β	0	0	0	0	0	0	0	γ	0
Assoc-Far Occn	$\gamma\alpha$	α	αp	0	β	βp	0	0	α	αp	0	0	β	βp	0	0	0
Assoc-Region Occn	0	$\gamma\alpha$	α	αp	$\gamma\beta$	β	βp	0	0	α	αp	0	0	3	βp	0	0
Assoc-Area Occn	0	0	$\gamma\alpha$	α	0	$\gamma\beta$	β	βp	0	0	α	0	0	0	β	γ	0
Assoc-Town Occn	0	0	0	$\gamma\alpha$	0	0	$\gamma\beta$	β	0	0	0	0	0	0	0	e	γ
Pen. Link	0	0	0	$\gamma\alpha$	0	0	0	$\gamma\beta$	0	0	0	0	0	0	0	e	e
Target	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Table 2. **The model transition probabilities for the Markov process. The quantities α and β are determined row by row.**

FIGURE CAPTIONS

1. How an intermediary is assumed to choose the next link in a SW chain. (One stage is omitted, for clarity. See text).
2. Likely movements between categories in the model.
3. The model probabilities of path lengths originating from each category of starter, taking the values 1, 2, ... 10. (Path lengths, and not number of intermediaries, are used consistently throughout).
4. Mean and standard deviation of path lengths from each starter category.
5. Completion probabilities for paths originating from each category, with a loss rate of 22%.
6. Mean and standard deviation of complete and incomplete paths, originating from the 16 categories of the model.
7. Predicted and observed probabilities of incomplete path lengths taking the values 1, 2, ... 10. For clarity, only values for the three starter categories which allow comparison with observed path lengths are shown.
8. Predicted and observed probabilities of complete path lengths (with a 22% loss rate) taking the values 1, 2, ... 10. a) "far" starters; b) "far-occupation;" c) "region."

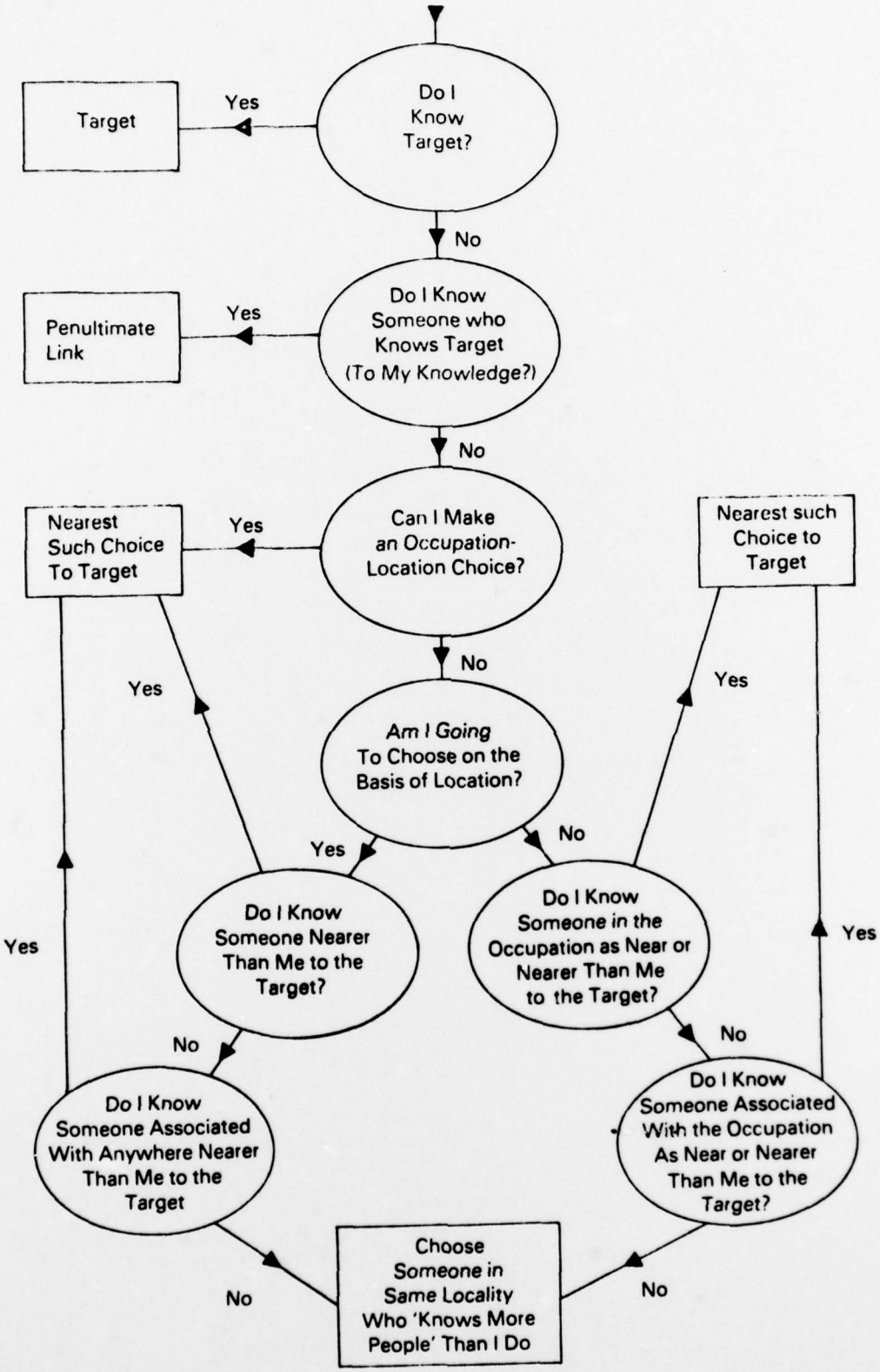

Do I Know Target?
Yes
Target
No
Do I Know Someone who Knows Target (To My Knowledge?)
Yes
Penultimate Link
No
Can I Make an Occupation-Location Choice?
Yes
Nearest Such Choice To Target
No
Am I Going To Choose on the Basis of Location?
Yes
No
Yes
Do I Know Someone Nearer Than Me to the Target?
Do I Know Someone in the Occupation as Near or Nearer Than Me to the Target?
Yes
Nearest such Choice to Target
Yes
No
No
Do I Know Someone Associated With Anywhere Nearer Than Me to the Target
Do I Know Someone Associated With the Occupation As Near or Nearer Than Me to the Target?
No
No
Choose Someone in Same Locality Who 'Knows More People' Than I Do

Figure 1.

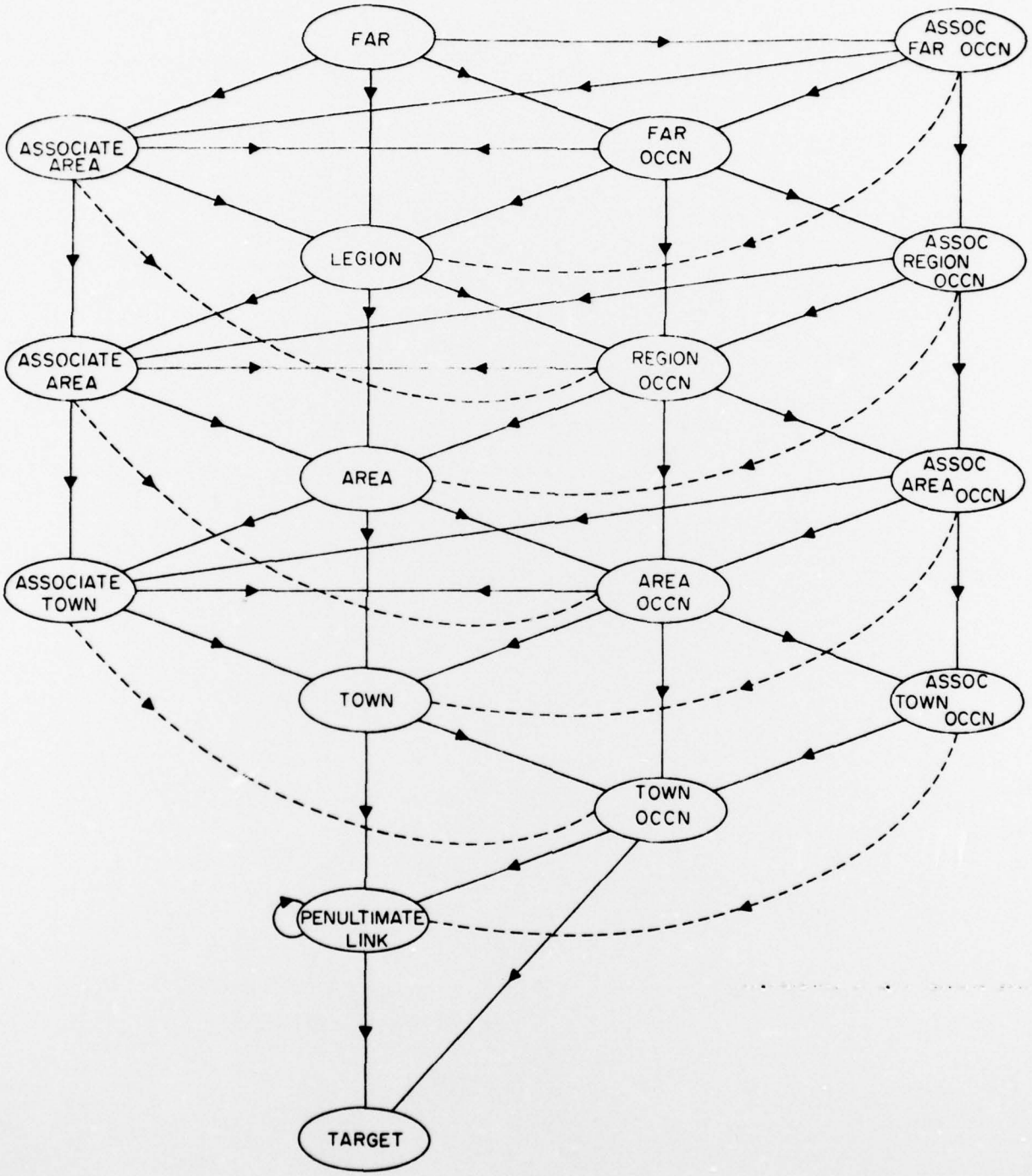

FAR
ASSOC FAR OCCN
ASSOCIATE AREA
FAR OCCN
LEGION
ASSOC REGION OCCN
ASSOCIATE AREA
REGION OCCN
AREA
ASSOC AREA OCCN
ASSOCIATE TOWN
AREA OCCN
TOWN
ASSOC TOWN OCCN
TOWN OCCN
PENULTIMATE LINK
TARGET

Figure 2.

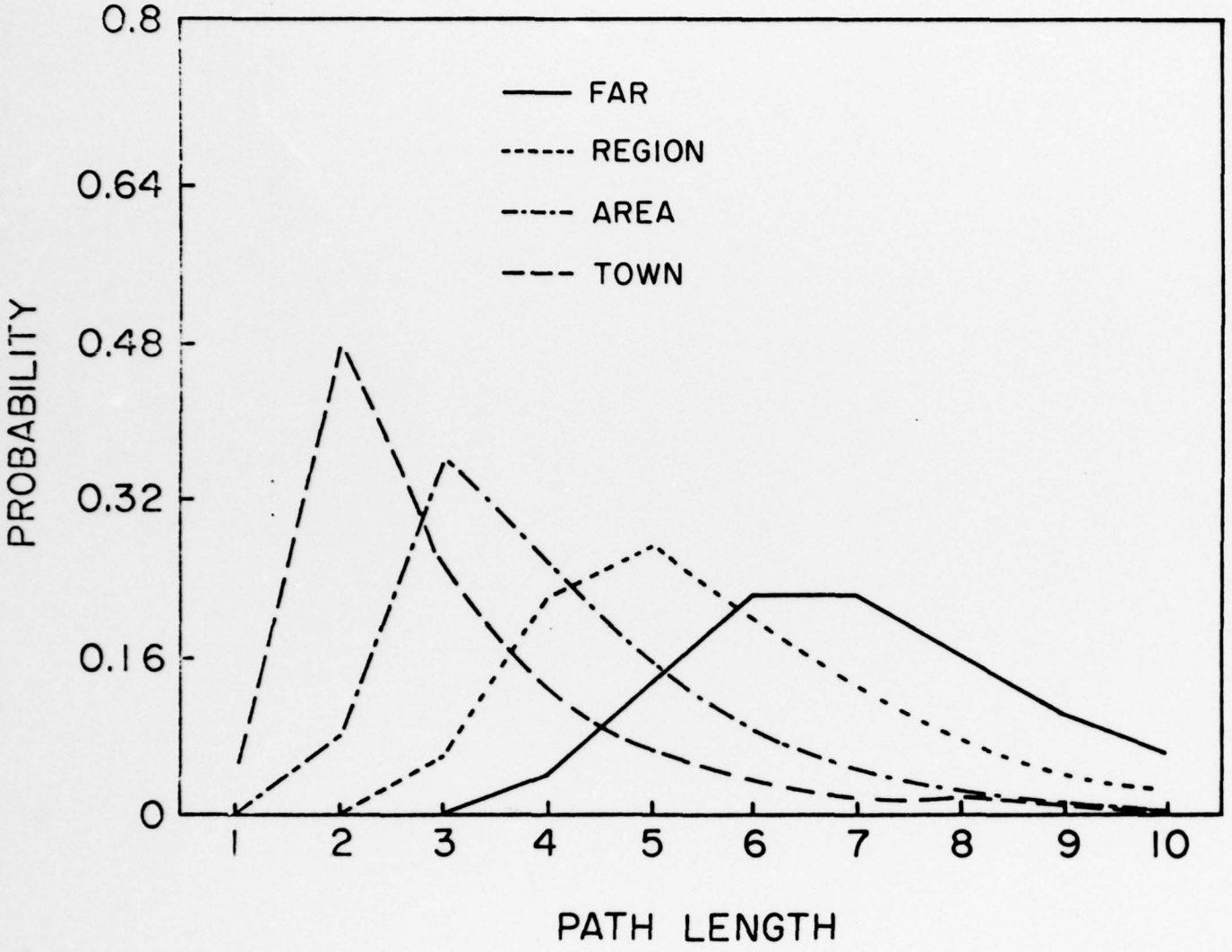

0.8
0.64
0.48
0.32
0.16
0
PROBABILITY
FAR
REGION
AREA
TOWN
1
2
3
4
5
6
7
8
9
10
PATH LENGTH

Figure 3a.

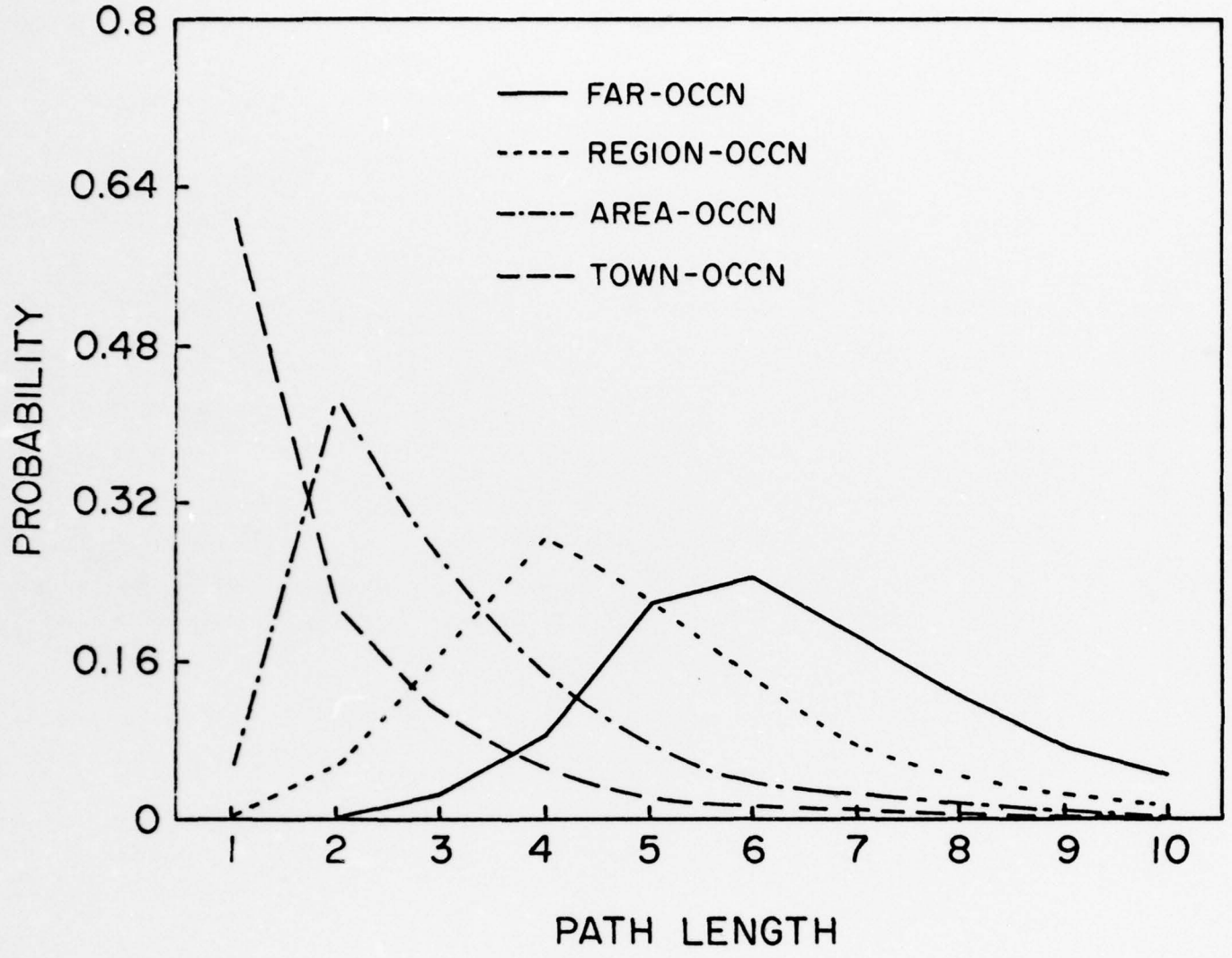

0.8
0.64
0.48
0.32
0.16
0
PROBABILITY
FAR-OCCN
REGION-OCCN
AREA-OCCN
TOWN-OCCN
1
2
3
4
5
6
7
8
9
10
PATH LENGTH

Figure 3b.

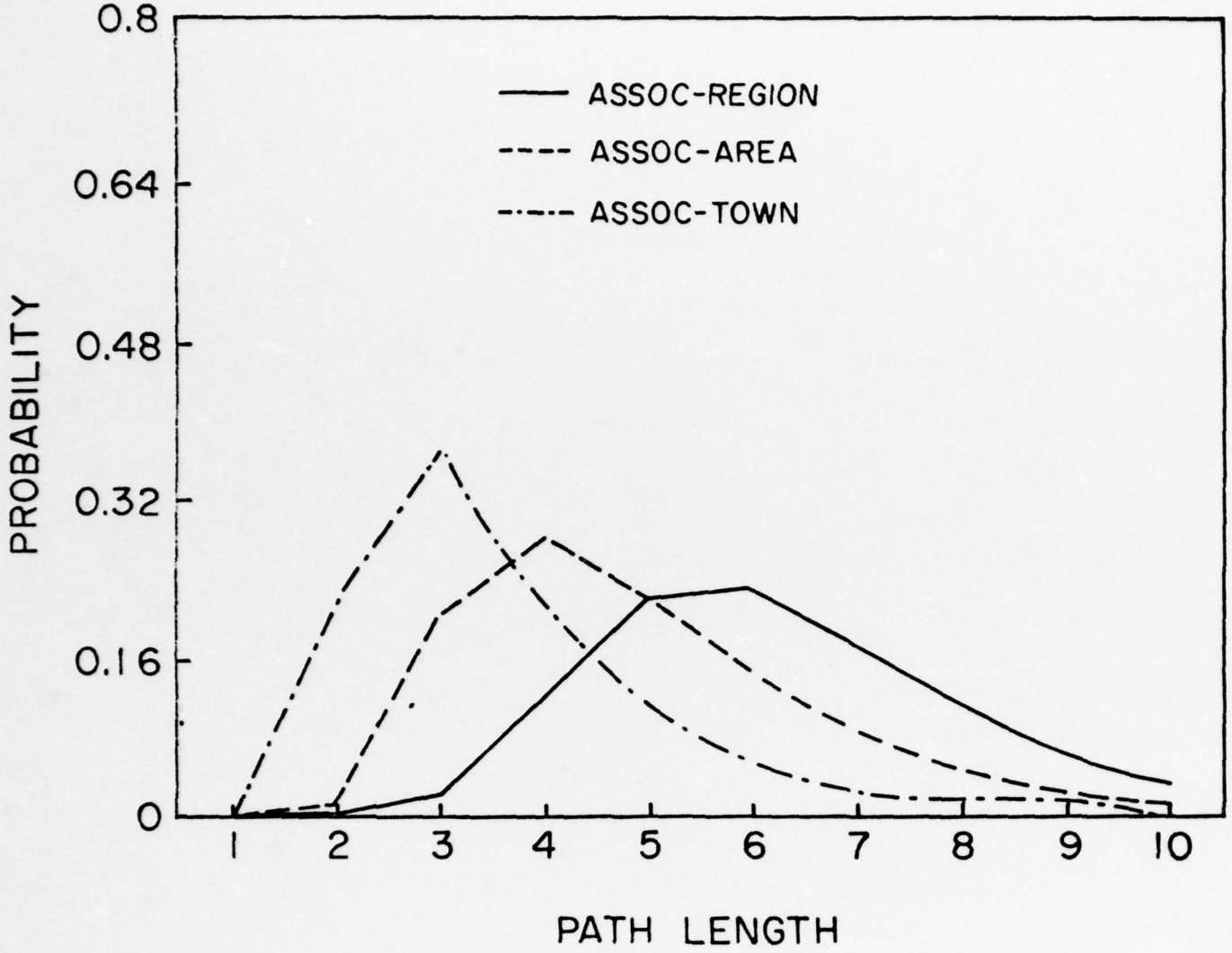

0.8
0.64
0.48
0.32
0.16
0
PROBABILITY
ASSOC-REGION
ASSOC-AREA
ASSOC-TOWN
1
2
3
4
5
6
7
8
9
10
PATH LENGTH

Figure 3c.

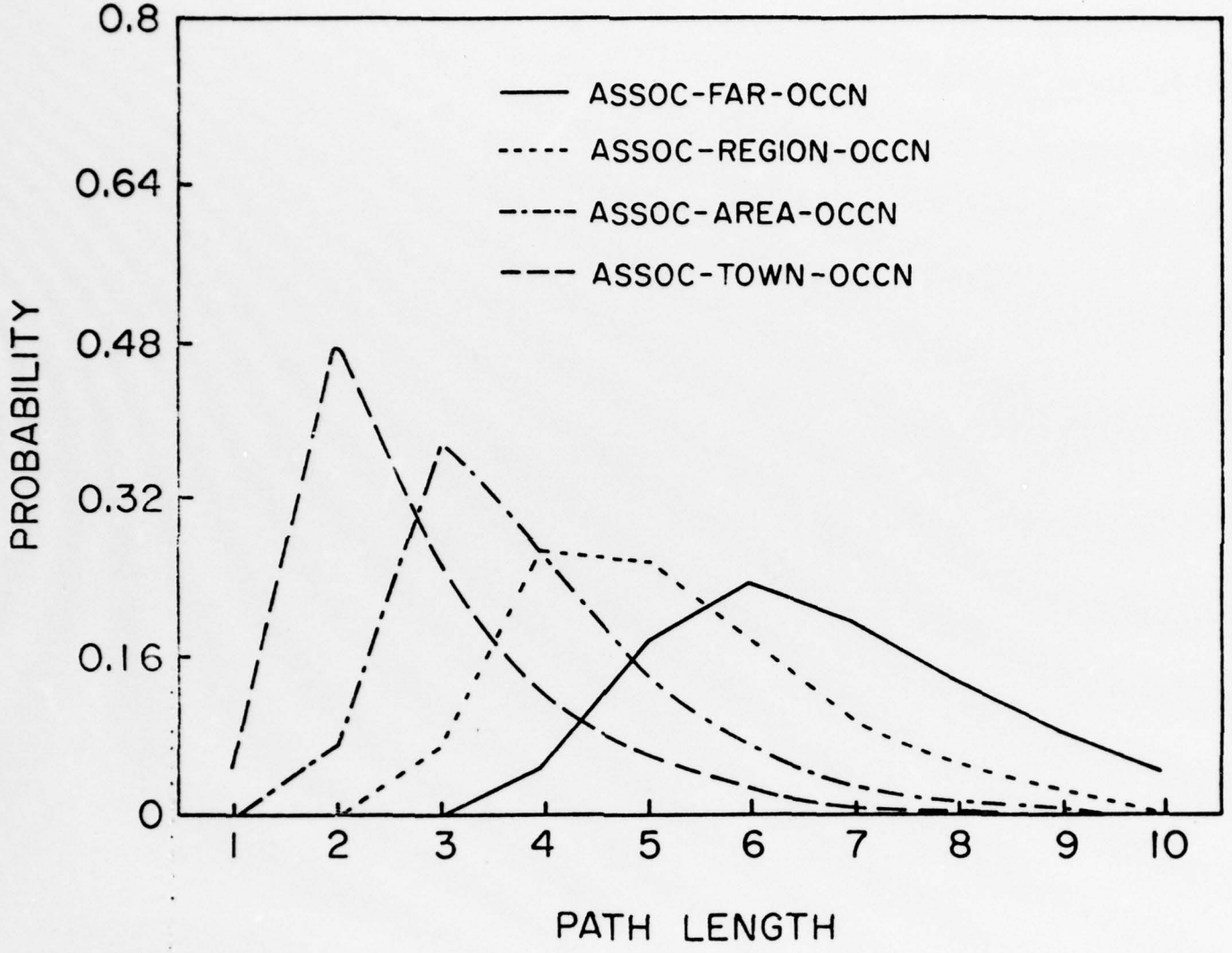

0.8
0.64
0.48
0.32
0.16
0
PROBABILITY
ASSOC-FAR-OCCN
ASSOC-REGION-OCCN
ASSOC-AREA-OCCN
ASSOC-TOWN-OCCN
1
2
3
4
5
6
7
8
9
10
PATH LENGTH

Figure 3d.

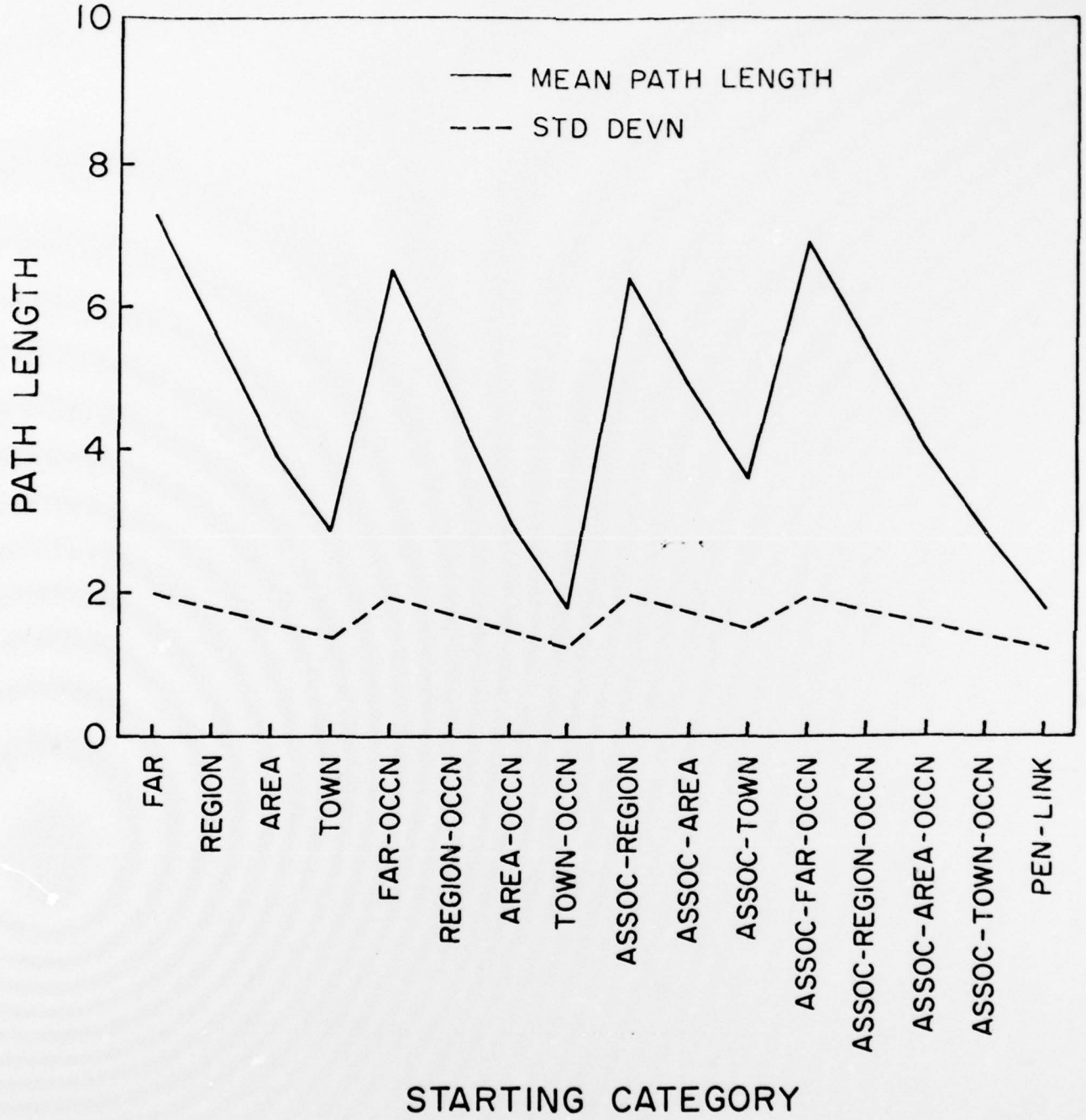

MEAN PATH LENGTH
STD DEVN
PATH LENGTH
0
2
4
6
8
10
FAR
REGION
AREA
TOWN
FAR-OCCN
REGION-OCCN
AREA-OCCN
TOWN-OCCN
ASSOC-REGION
ASSOC-AREA
ASSOC-TOWN
ASSOC-FAR-OCCN
ASSOC-REGION-OCCN
ASSOC-AREA-OCCN
ASSOC-TOWN-OCCN
PEN-LINK
STARTING CATEGORY

Figure 4.

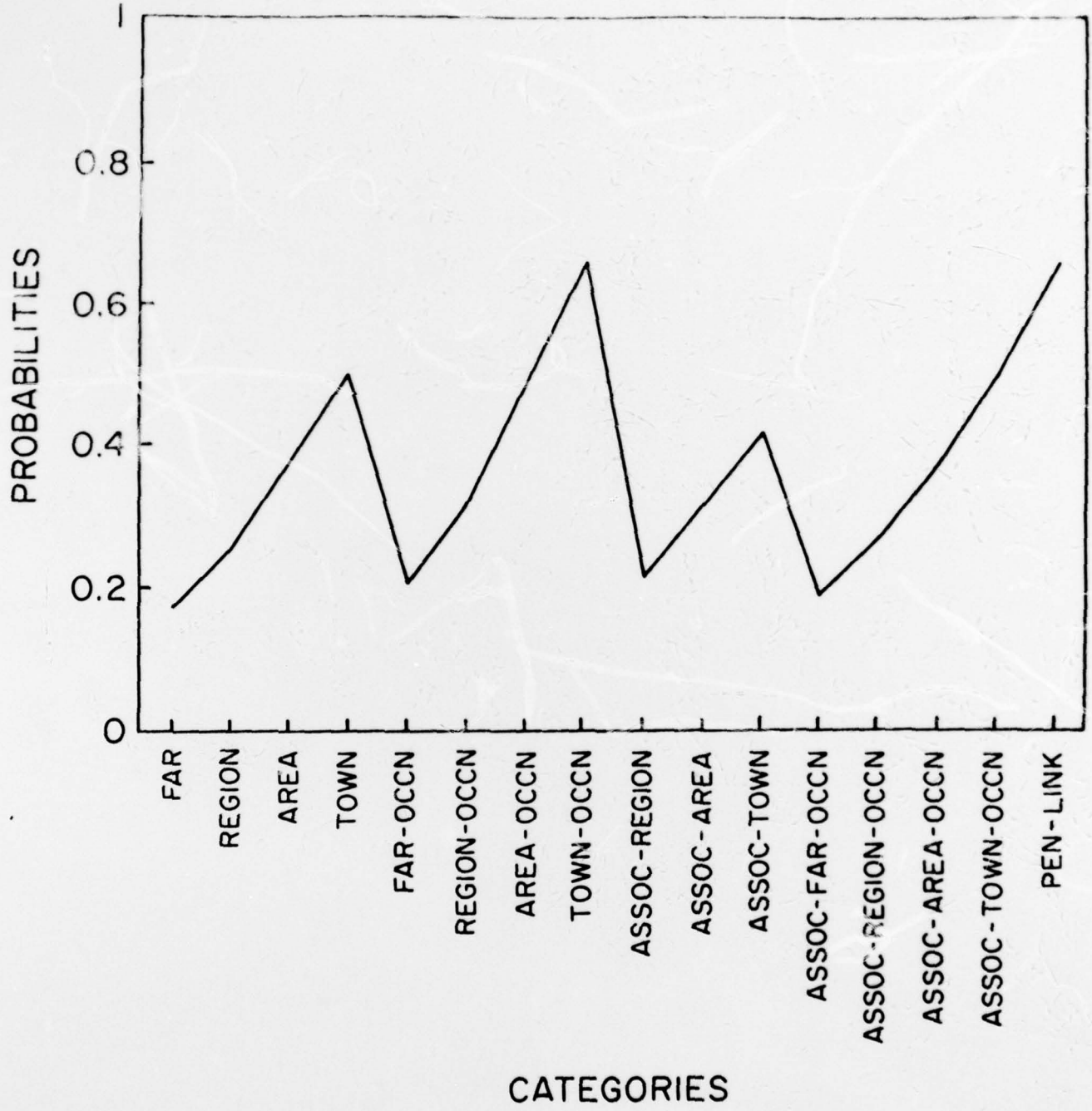

PROBABILITIES
1
0.8
0.6
0.4
0.2
0
FAR
REGION
AREA
TOWN
FAR-OCCN
REGION-OCCN
AREA-OCCN
TOWN-OCCN
ASSOC-REGION
ASSOC-AREA
ASSOC-TOWN
ASSOC-FAR-OCCN
ASSOC-REGION-OCCN
ASSOC-AREA-OCCN
ASSOC-TOWN-OCCN
PEN-LINK
CATEGORIES

Figure 5.

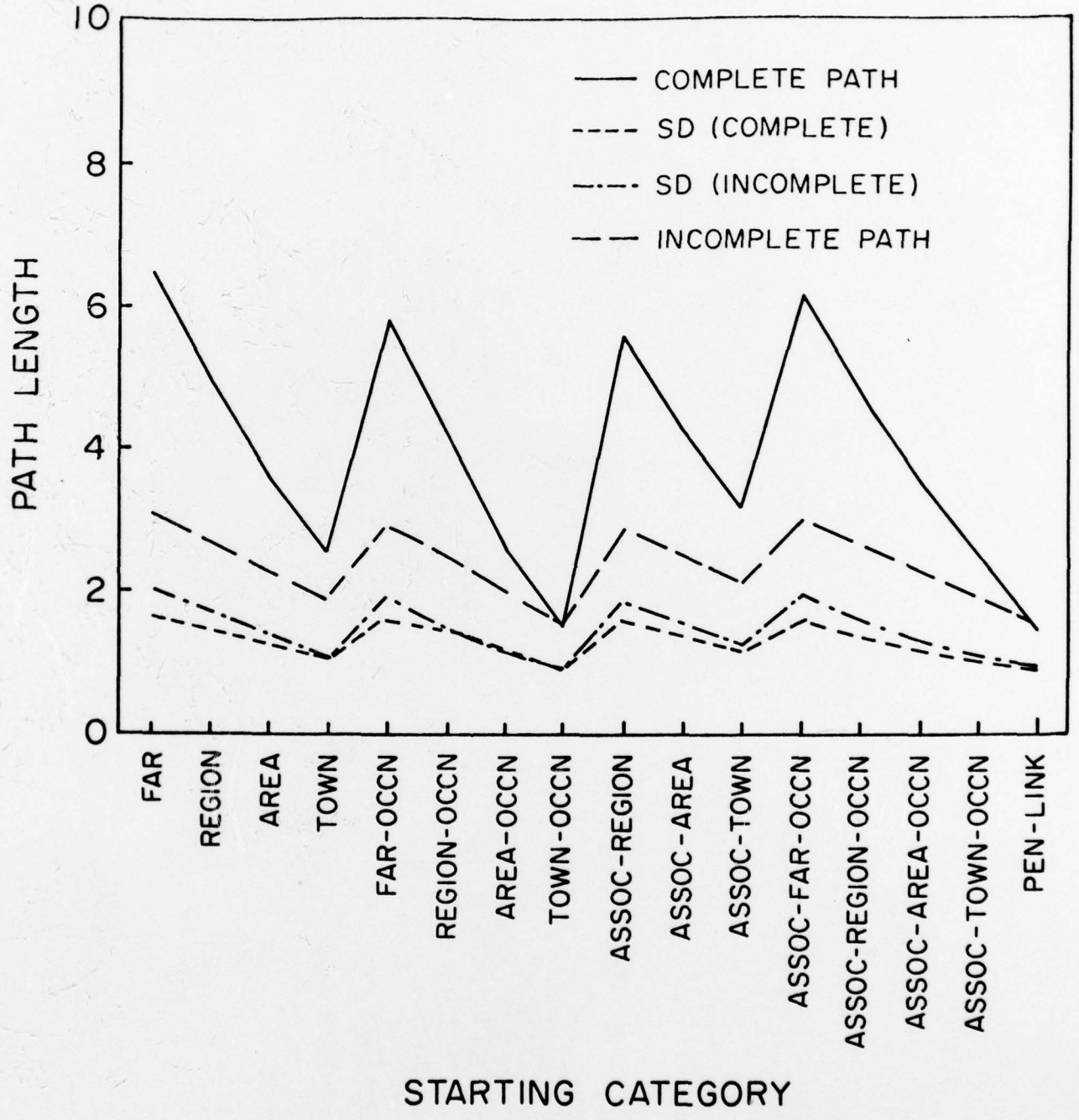

10
8
6
4
2
0
PATH LENGTH
COMPLETE PATH
SD (COMPLETE)
SD (INCOMPLETE)
INCOMPLETE PATH
FAR
REGION
AREA
TOWN
FAR-OCCN
REGION-OCCN
AREA-OCCN
TOWN-OCCN
ASSOC-REGION
ASSOC-AREA
ASSOC-TOWN
ASSOC-FAR-OCCN
ASSOC-REGION-OCCN
ASSOC-AREA-OCCN
ASSOC-TOWN-OCCN
PEN-LINK
STARTING CATEGORY

Figure 6.

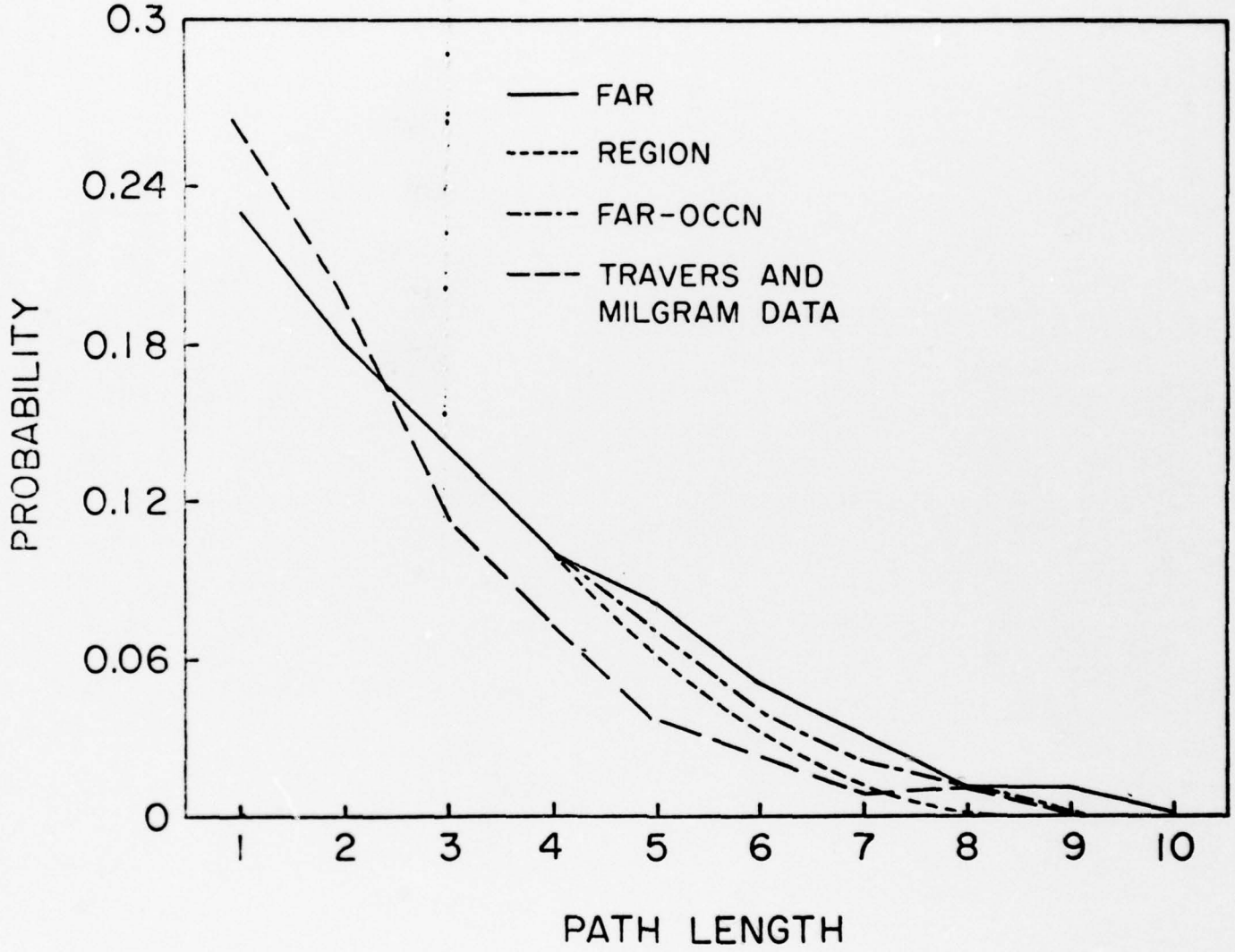

0.3
0.24
0.18
0.12
0.06
0
PROBABILITY
FAR
REGION
FAR-OCCN
TRAVERS AND
MILGRAM DATA
1
2
3
4
5
6
7
8
9
10
PATH LENGTH

Figure 7.

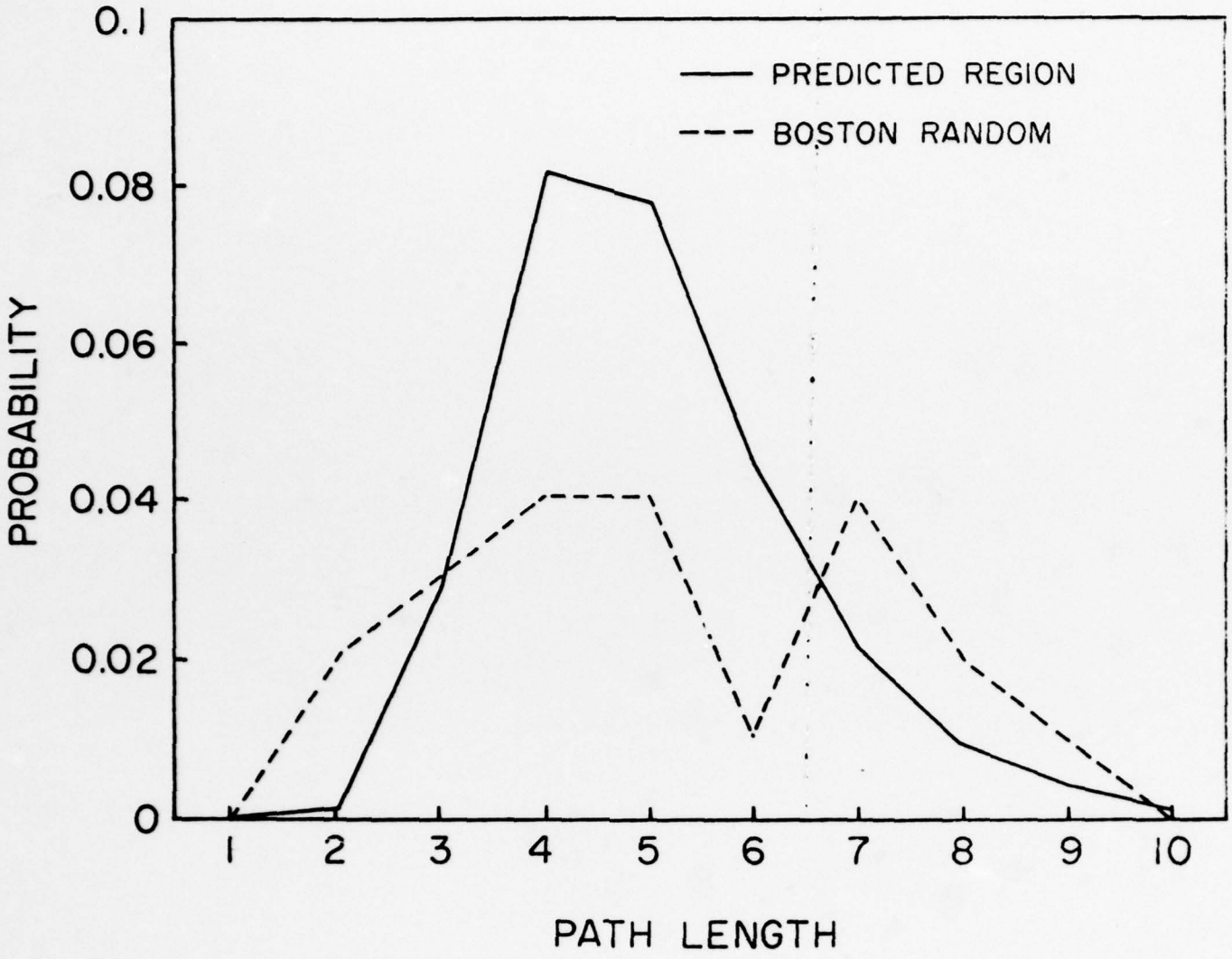

0.1
0.08
0.06
0.04
0.02
0
PROBABILITY
PREDICTED REGION
BOSTON RANDOM
1
2
3
4
5
6
7
8
9
10
PATH LENGTH

Figure 8a.

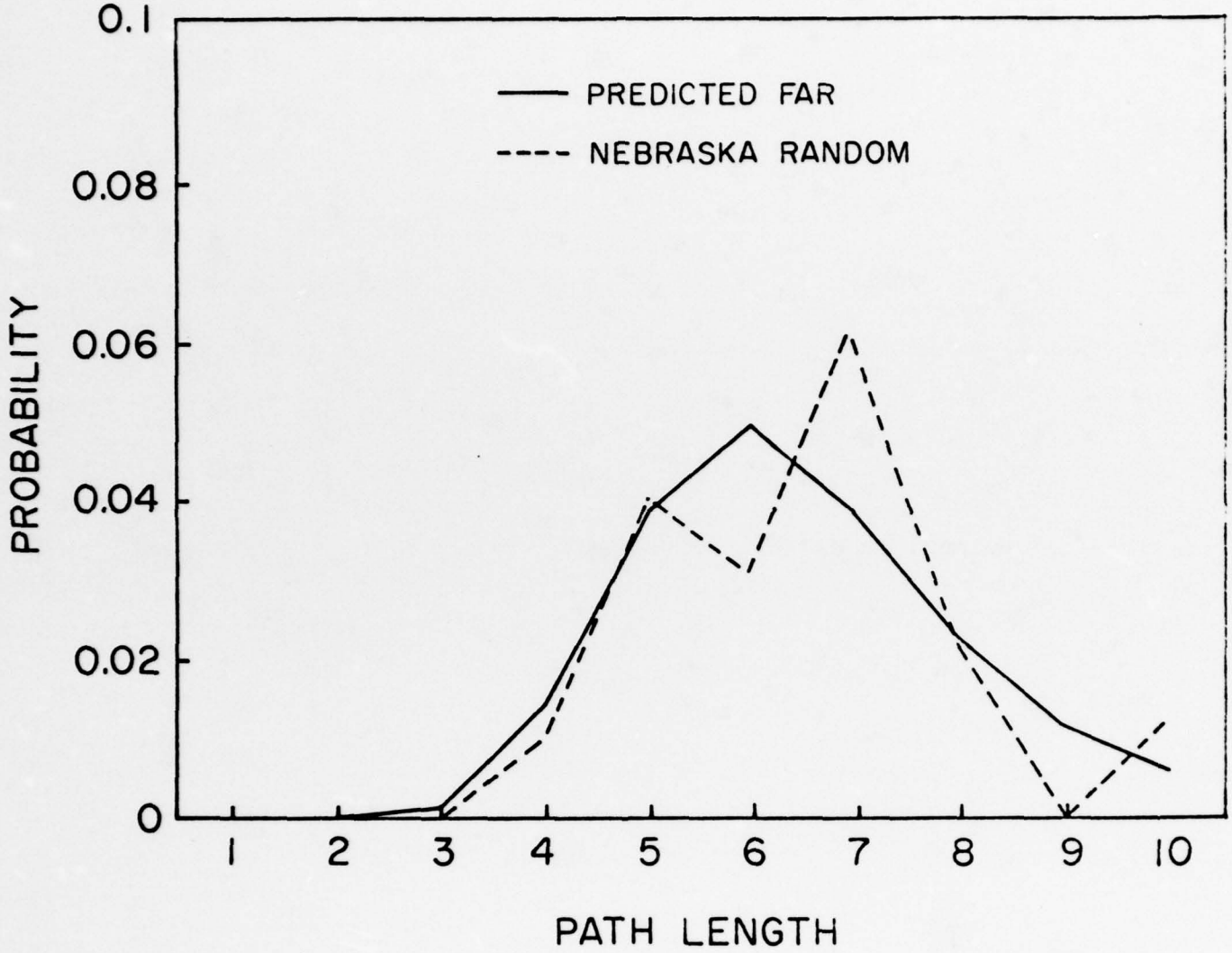

0.1
0.08
0.06
0.04
0.02
0
PROBABILITY
PREDICTED FAR
NEBRASKA RANDOM
1
2
3
4
5
6
7
8
9
10
PATH LENGTH

Figure 8b.

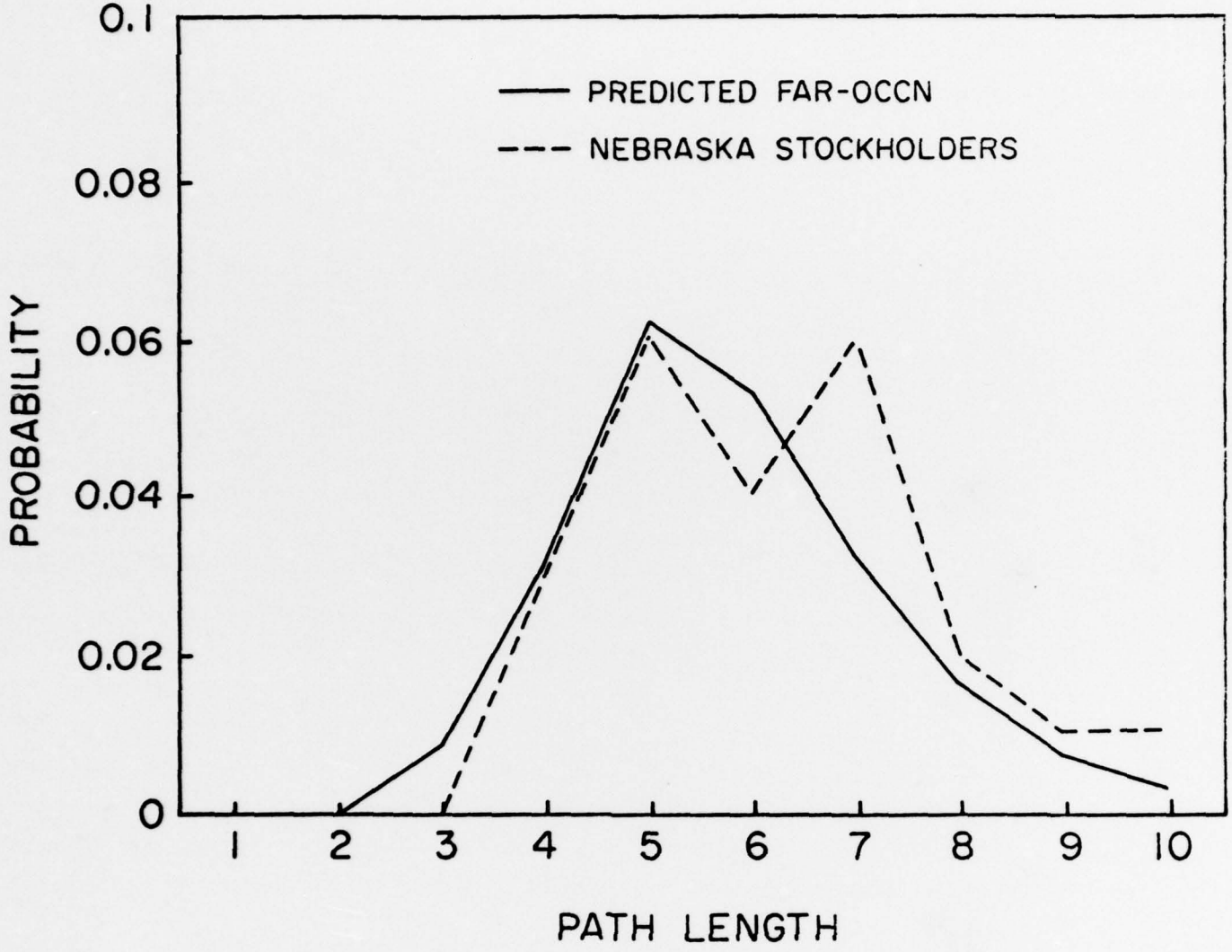

0.1
0.08
0.06
0.04
0.02
0
PROBABILITY
PREDICTED FAR-OCCN
NEBRASKA STOCKHOLDERS
1
2
3
4
5
6
7
8
9
10
PATH LENGTH

Figure 8c.