BULLETIN OF
MATHEMATICAL BIOPHYSICS
VOLUME 15, 1953

SPREAD OF INFORMATION THROUGH A POPULATION WITH SOCIO-STRUCTURAL BIAS: I. ASSUMPTION OF TRANSITIVITY

ANATOL RAPOPORT
COMMITTEE ON MATHEMATICAL BIOLOGY
THE UNIVERSITY OF CHICAGO

A previously derived iteration formula for a random net was applied to some data on the ead of information through a population. It was found that if the axon density (the only e parameter in the formula) is determined by the first pair of experimental values, the dicted spread is much more rapid than the observed one. If the successive values of the pparent axon density" are calculated from the successive experimental values, it is noticed it this quantity at first suffers a sharp drop from an initial high value to its lowest value and n gradually "recovers."

An attempt is made to account for this behavior of the apparent axon density in terms of "assumption of transitivity," based on a certain socio-structural bias, namely, that the ely contacts of two individuals who themselves have been in contact are expected to be ongly overlapping. The assumption of transitivity leads to a drop in the apparent axon nsity from an arbitrary initial value to the vicinity of unity (if the actual axon density is not too small). However, the "recovery" is not accounted for, and thus the predicted spread turns out to be *slower* than the observed.

The apparent axon density. The assumption of a completely mixed population implies that contacts between all pairs of individuals are equiprobable at all times. It was shown in previous papers (e.g., Rapoport, 1951) that under this assumption the spread of state in "ordinal time" or "by removes" will be given by the following iteration formula

$$P(t+1) = [1 - x(t)]\,[1 - e^{-aP(t)}], \tag{1}$$

which can also be written

$$[1 - x(t+1)]\, e^{ax(t)} = 1 - x(0). \tag{2}$$

Here t takes discrete integral values and represents the ordinal time, that is, the number of "removes" or times the information was transmitted, $P(t)$ is the fraction of *new* knowers at the tth remove,

$$x(t) = \sum_{j=0}^{t} P(j)$$

is the total fraction of knowers at the tth remove, and a the "actual axon density," is the number of tellings per knower during the entire process, assumed independent of t. We shall continue to use the neural net term "axon density" in this sense, which suggests that a "axons" or contacts "issue" from each individual (on the average), and that this number is characteristic for a particular process.

In an experiment conducted by the Washington Public Opinion Laboratory (1952), 33 simple messages were made to diffuse through a population of 184 school children under contest conditions. It was possible to trace through how many hands (removes) each message has gone. The experiment can be thought of as one in which a single message diffuses through a population of 6072 individuals, a population large enough for the application of our probabilistic method. Accordingly equation (1) was tested by

TABLE I

t	0	1	2	3	4	5	6	7	8	9	10
$P(t)$ calc.	.030	.191	.585	.191	.002	.000					
$P(t)$ obs.	.030	.191	.191	.156	.107	.063	.035	.018	.009	.003	.002
$x(t)$ calc.	.030	.221	.806	.997	.999	.999					
$x(t)$ obs.	.030	.221	.412	.568	.675	.738	.773	.791	.802	.805	.807
$\alpha(t)$	7.2	1.47	1.61	1.81	2.03	2.28	2.30	3.06	1.67	2.99	1.91

the data of that experiment. The value of $P(0) = x(0)$ was fixed by the conditions of the experiment: each child was the starter of a message. Hence $P(0)$ was fixed at $184/6072 = .03$. The value of a was obtained from the observed value of $P(1)$. The predicted ordinal time course could thus be compared with the experimentally observed one. This comparison appears in the first four rows of Table I.

Obviously equation (1) predicts an ordinal time course of a spread which is much more rapid than the observed one. Attempts to account for this discrepancy led us to examine the so-called "apparent axon density" of the experimental process. The apparent axon density of any real ordinal time course of the spread of information is defined as the function $\alpha(t)$, which, when substituted for the actual axon density as in equation (1), will account for all the points in the ordinal time course. In other words, we shall describe any ordinal time course by the equation

$$P(t+1) = [1 - x(t)]\ [1 - e^{-\alpha(t)P(t)}], \tag{3}$$

where $\alpha(t)$ is a function to be determined. Obviously if all the $P(t)$, and therefore the $x(t)$, of the process are known, $\alpha(t)$ can be determined by solving for it in equation (3). We have, in fact,

$$\alpha(t) = \frac{1}{P(t)} \log \frac{1 - x(t)}{1 - x(t+1)}. \tag{4}$$

The fundamental assumption of the completely random process is equivalent to the assumption that the expression on the right side of (4) is a constant (i.e., the actual axon density a). Therefore, given any actual process, the constancy of the expression on the right of (4) is a test of the random net hypothesis.

The determination of $\alpha(t)$ in the experiment referred to gave a set of values which appears in the last row of Table I. We observe that $\alpha(t)$ starts at a rather high value (7.2) at $t = 0$, drops sharply for $t = 1$ to its lowest value, then rises steadily for several successive values. For $t > 7$, $\alpha(t)$ becomes erratic. However, as is seen from Table I, $P(t)$ becomes extremely small for $t > 7$, and, as can be deduced from (4), $\alpha(t)$ becomes excessively sensitive to small fluctuations in $P(t)$ when $x(t)$ is close to unity. It follows that the values of $\alpha(t)$ in that range are not reliable. At any rate, it appears reasonable to assume that the behavior of $\alpha(t)$ in the experiment at hand is characterized by a large initial drop and a subsequent steady "recovery" in its significant range.

One could attempt to explain this behavior of $\alpha(t)$ by psychological considerations. For example, one could attribute the initial high value of α to a "start effect," that is, the enthusiasm of initiating the process. If, as is the case in some message diffusion experiments, only a portion of the population are the "starters," they can be expected to be more strongly motivated to spread the information than those who get the information "second hand." This can conceivably explain the initial drop in $\alpha(t)$. In the case considered, however, this explanation is not convincing, inasmuch as every one of the 184 children started one of the 33 messages. The same individuals, therefore, were involved in all the removes. It is hard to believe that the motivation of the same individuals fell so sharply immediately after the start that the axon density was reduced by a factor of 5.

In addition, we are still faced with the opposite effect, namely, the unmistakable steady rise of the apparent axon density. One could, of course, also explain this effect by psychological considerations The experiment was conducted under contest conditions, where prizes were offered both for giving and for receiving as much information as possible. It is conceivable that as the population became more and more saturated with knowers, and, therefore, as it became more and more difficult to find nonknowers to give information to, the total number of contacts tended to increase with each remove.

In this paper we will not make any psychological hypotheses. We will try instead to explain at least the initial drop in $\alpha(t)$ by some assumptions concerning the *structure* of the population, which places certain constraints on the possible contacts, so that they are no longer equiprobable.

The finite acquaintance circle. The first constraint will be contained in the assumption that the contacts of each individual take place only within his acquaintance circle. We will assume first that an individual has on the average q acquaintances randomly chosen from the population. We will assume, moreover, that q is large compared to a. Following the reasoning in the derivation of equation (1) (Rapoport, *loc. cit.*) it is easy to see that under the modified assumption of the finite acquaintance circle we shall have

$$P(t+1) = [1 - x(t)]\left[1 - \left(1 - \frac{1}{q}\right)^{aP(t)q}\right]. \tag{5}$$

If q is fairly large, so that the power in the second bracket on the right of (5) can be approximated by an exponential, (5) reduces to (1), and no essential modification has been introduced. If, however, q, although large compared to a, is small compared to N, we can reason along a different line.

Let us fix our attention on an arbitrary individual A at the tracing of the $(t+1)$th remove. We seek the probability that on that remove A does *not* receive the information from an arbitrarily chosen individual B among the q individuals in his acquaintance circle. This can happen in either of two mutually exclusive ways: either B did not become a knower on the tth remove or he did become a knower on the tth remove, but his contacts among his own acquaintances did not include A. The probability we seek is the sum of the probabilities of these two events, that is,

$$1 - P(t) + P(t)\left(1 - \frac{1}{q}\right)^{a}. \tag{6}$$

Now if all the states of the acquaintances of A are independent of each other, and if q is small compared to the total population, so that sampling with replacement can be assumed for any sample of individuals not greater than q, then the probability that A did not receive the information on the $(t+1)$th remove will be given by

$$\left\{1 - P(t) + P(t)\left(1 - \frac{1}{q}\right)^{a}\right\}^{q}. \tag{7}$$

Then the probability that A did receive the information on the $(t+1)$th remove will be

$$1-\left\{1-P(t)+P(t)\left(1-\frac{1}{q}\right)^{a}\right\}^{q}=1-\left\{1-P(t)\,m\right\}^{q}, \tag{8}$$

where

$$m=1-\left(1-\frac{1}{q}\right)^{a}. \tag{9}$$

Expression (8) corresponds to the expression $1-e^{-aP(t)}$ in the completely random case. We therefore write for our modified equation representing the spread of information by removes

$$P(t+1)=[1-x(t)]\,[1-\{1-P(t)\,m\}^{q}]. \tag{10}$$

Solving for $\alpha(t)$, as defined by (4), we have

$$\alpha(t)=\frac{-q}{P(t)}\log[1-P(t)\,m]. \tag{11}$$

Since $P(t)<1$, and $m<1$, we can expand the right side of (11) to obtain the series

$$qm+\tfrac{1}{2}qm^{2}P(t)+\tfrac{1}{3}qm^{3}P(t)^{2}+\ldots. \tag{12}$$

We note further that if q is large compared to a

$$m=1-\left(1-\frac{1}{q}\right)^{a}\sim\frac{a}{q}\ll 1, \tag{13}$$

so that

$$\alpha(t)=a\left[1+\frac{aP(t)}{2q}+\frac{\{aP(t)\}^{2}}{3q^{2}}+\ldots\right]. \tag{14}$$

Since $P(t)<1$, and $a/q\ll 1$, the series converge rapidly. Thus $\alpha(t)$ depends (but very weakly) upon $P(t)$, rising and falling with it. As q becomes very large $\alpha(t)$ tends to a, as, of course, should be the case and as is clear from (5).

Although this approach to the finite acquaintance circle case still gives no substantially new result, it does lend itself rather readily to the imposition of a socio-structural bias, which we will now discuss.

The dependence of probabilities. So far the assumption underlying the whole argument was that the probabilities, $1-P(t)m$, that each of the q individuals in A's acquaintance circle did not inform A on the $(t+1)$th remove were all equal, that is, the associated events were all independent. Another way of saying this is that our knowledge that the first acquaintance did not inform A did not affect our assumption about the state of the second acquaintance, etc. If we drop this assumption of independence, the compound probability that none of A's q acquaintances informed A can no longer be represented by (7). We must write instead of the qth power a q-fold product

$$\sum_{k=0}^{q-1} [1 - P_k(t)\, m] , \tag{15}$$

where the $P_k(t)$ are conditional probabilities to be determined below. Using expression (15) instead of (7) in equation (10), and solving for $\alpha(t)$, defined by (4), we now have

$$\alpha(t) = \frac{-1}{P(t)} \sum_{k=0}^{q-1} \log [1 - P_k(t)\, m] . \tag{16}$$

For $m \ll 1$ (and *a forteriori* $P_k(t)m \ll 1$), the logarithm in (16) can be well approximated by $-P_k(t)m$, and we have the simplified form of $\alpha(t)$, namely,

$$\alpha(t) = \frac{m}{P(t)} \sum_{k=0}^{q-1} P_k(t) . \tag{17}$$

For the special case of the completely mixed population all the P_k are equal, and

$$\alpha(t) = qm \cong a , \tag{18}$$

as, of course, should be the case. It now remains to determine the P_k on the basis of certain assumptions which we wish to make about our socio-structural bias.

The conditional probability P_1. Let us compute the following conditional probability: given that an arbitrary selected acquaintance of A did *not* inform him on the $(t+1)$th remove, what is the probability that this acquaintance was not a new knower on the tth remove?

We apply Bayes' Rule

$$p(H_k \mid E) = \frac{p(E \mid H_k)\, p(H_k)}{\sum p(E \mid H_j)\, p(H_j)} , \tag{19}$$

which gives the probability of a hypothesis H_k, given the occurrence of the event E, in terms of the probabilities of the event E, given each of the possible hypotheses H_j, and the probabilities of these hypotheses. In our case we have two hypotheses, namely,

H_1: he is a new knower on the tth remove, and

H_2: he is not a new knower on the tth remove.

Therefore

$$p(H_1) = P(t) ; \qquad p(H_2) = 1 - P(t) . \tag{20}$$

Our event E stands for "he did not inform A." Hence

$$p(E|H_1) = \left(1 - \frac{1}{q}\right)^a = 1 - m ; \qquad p(E|H_2) = 1 . \tag{21}$$

The probability we seek is

$$p(H_2|E) = \frac{1 - P(t)}{1 - P(t)\, m}, \tag{22}$$

whence the complementary probability, namely, that the acquaintance of A who did not inform him is a new knower on the tth remove, is

$$P_1(t) = 1 - p(H_2|E) = \frac{P(t)\,(1 - m)}{1 - P(t)\, m}. \tag{23}$$

The assumption of transitivity. Let us now think of our population of N individuals as composed of N subsets or neighborhoods, each composed of a "representative" member of the population and his q acquaintances. Of course, the neighborhoods are overlapping. The over-all density of new knowers in the population on the tth remove is by definition $P(t)$. However, if we select a *subset* of the neighborhoods according to some criterion, we can expect that the density of new knowers in that subset will be different from $P(t)$, depending on the criterion we use in the selection of the subset. Let the criterion now be the following: we select all those neighborhoods in which the representative individuals were not informed on the $(t + 1)$th remove by an arbitrarily selected individual in the corresponding neighborhoods. According to our argument above, the fraction of new knowers among those non-informing individuals is not $P(t)$ but slightly less, namely, $P_1(t)$, given by equation (23).

The "assumption of transitivity" which we now make is the following: $P_1(t)$ can be taken as the density of new knowers in the subset of neighborhoods characterized by the criterion above (i.e., those neighborhoods in which none of the arbitrarily selected first acquaintances of the representative individuals informed the representative individuals on the tth remove). This is tantamount to making the following two assumptions:

1. The density of new knowers in a subset of neighborhoods can be taken as the density of new knowers in a set of individuals, each arbitrarily selected from each neighborhood.

2. If instead of choosing a given set of individuals, to define a subset of neighborhoods, we substitute for each of the individuals an acquaintance of his, we get a subset of neighborhoods practically identical with the original one.

The validity of the first assumption depends on the size of the subset. If the number of individuals in it is sufficiently large, then the density of new knowers in a sample $(1/q)$th the size of it can be taken to be the density of new knowers in the whole subset. The second assumption implies that the neighborhoods (acquaintance circles) of two individuals who are acquainted are very strongly interlocking. If the population were divided into mutually exclusive cliques, within which all the individuals were acquainted with each other, then the second assumption would be satisfied exactly, because in that case any individual in a clique would represent the same clique. Our model, however, differs from the exclusive clique model, as will appear in the discussion below.

We see that our assumptions imply that on the tth remove the density of new knowers in neighborhoods, characterized by the non-telling by the first arbitrarily selected acquaintances to the representative individual, is $P_1(t)$. By similar reasoning, the density of new knowers in neighborhoods characterized by the non-telling of the first two arbitrarily selected acquaintances will be

$$P_2(t) = \frac{P_1(t)\,(1-m)}{1-P_1(t)\,m}, \tag{24}$$

and, in general,

$$P_k(t) = \frac{P_{k-1}(t)\,(1-m)}{1-P_{k-1}(t)\,m}. \tag{25}$$

Hence the probability that the representative individual is not informed by *any* of his acquaintances is

$$\prod_{k=0}^{q-1}[1-P_k(t)\,m], \tag{26}$$

and our quantities P_k are identical with the $P_k(t)$ of expression (15).

We can now show by induction that

$$P_k(t) = \frac{P(t)\,s^k}{1-P(t)+P(t)\,s^k}, \tag{27}$$

where $s = 1 - m$. For suppose that (27) holds for k. Then, according to (25),

$$\left.\begin{aligned} P_{k+1} &= \frac{P(t)\,s^k}{1-P(t)+P(t)\,s^k}s\left[1-\frac{P(t)\,s^k}{1-P(t)+P(t)\,s^k}(1-s)\right]^{-1} \\ &= \frac{P(t)\,s^{k+1}}{1-P(t)+P(t)\,s^{k+1}}, \end{aligned}\right\} \tag{28}$$

which establishes the induction.

To obtain the expression $\alpha(t)$ as given by (17), we seek

$$\sum_{k=0}^{q-1}P_k(t).$$

Since $m \ll 1$, and $s = 1 - m$, s is nearly unity, and its successive powers differ slightly from each other. We can therefore approximate our sum by the integral

$$\int_0^q \frac{P(t)\, s^k}{1 - P(t) + P(t)\, s^k}\, dk = \frac{1}{\log s} \log\left[1 - P(t)\,(1 - s^q)\right]. \qquad (29)$$

Recalling that $s = 1 - m \cong 1 - a/q$, we can further approximate the right side of (29) by

$$-\frac{1}{m} \log\left[1 - P(t)\,(1 - e^{-a})\right], \qquad (30)$$

thus obtaining the desired expression, namely,

$$\alpha(t) = \frac{-1}{P(t)} \log\left[1 - P(t)\,(1 - e^{-a})\right]. \qquad (31)$$

Equation (31) holds for $t \geqq 1$. For $t = 0$, we must take $\alpha(t)$ as a in equation (1), because the *initial* knowers, being selected entirely at random from the population, determine a set of entirely randomly selected neighborhoods, so that the very first step of the process must be supposed to be governed by randomness. We thus have the apparent axon density given by two equations, namely,

$$\left.\begin{aligned} \alpha(0) &= \frac{1}{P(0)} \log \frac{1 - x(0)}{1 - x(1)} \\ \alpha(t) &= \frac{1}{P(t)} \log\left[1 - P(t)\,(1 - e^{-a})\right], \qquad t > 0\,. \end{aligned}\right\} \qquad (32)$$

If $P(0)$ is sufficiently small and a not too large, $P(1)$ will also be a small fraction. In that case $\alpha(1)$ will be approximately equal to $1 - e^{-a}$, or, for $a > 2$, quite close to unity, where it will remain during the remainder of the process, since the successive P's must then become even smaller. In part III of this series, some indication will be given on the closeness of approximation by (31) to (17).

The large drop from $\alpha(0)$ to $\alpha(1)$ is thus accounted for by the assumption of transitivity, but not the subsequent rise of $\alpha(t)$. A modification of the assumptions toward that end will be undertaken in part II.

The spread of state by removes. Equation (4) gives the ordinal time course of the process if an explicit expression is substituted for $\alpha(t)$. In our case, this is given by (31). Thus

$$-\frac{1}{P(t)} \log\left[1 - P\,(1 - e^{-a})\right] = \frac{1}{P(t)} \log \frac{1 - x(t)}{1 - x(t+1)}\,.$$

Recalling that $P(t+1) = x(t+1) - x(t)$, and rearranging, we obtain

$$P(t+1) = [P(t)]\,[1 - x(t)]\,[1 - e^{-a}]\,, \qquad t > 0\,. \tag{33}$$

The "physical meaning" of equation (33) is the following. Suppose all the new knowers were "bunched" in clusters and normally talked only among themselves. If, however, a non-knower happened to get into one of these clusters, he could also be contacted by one of its members with the same probability as any other individual in it. Let the average size of these clusters be r individuals. Then, if once one gets into a cluster, the probability of receiving the information will be

$$1 - \left(1 - \frac{1}{r}\right)^{ar} \cong 1 - e^{-a}\,, \tag{34}$$

that is, this probability will be practically independent of the size of the cluster if the cluster is sufficiently large, so that the approximation (34) holds. On the other hand, the probability of getting inside the cluster equals the probability of meeting a new knower on the tth remove, that is, $P(t)$. Finally, $1 - x(t)$ is the probability of being a non-knower at t. This is the meaning of the three factors in (33).

Discussion. Although one of our assumptions above is equivalent to assuming mutually exclusive cliques, our model is not equivalent to the closed clique model. In the latter, the probabilities of becoming a knower cannot be taken independently for each remove. If, for example, an individual is not a knower by the tth remove, the knowledge of this fact should be reflected in assumptions about the state of affairs in his clique. For example, the probability that it had no *initial* knowers is increased thereby. This dependence of the distribution of knower densities on t does not appear in our equations.

The kind of situation reflected in equation (33) is, where there is a limited mixing in the population in the sense that the "acquaintance circle" does not stay fixed but changes with t, rather as if individuals moved through the population, and their "circles" were defined as simply their geographical vicinities. The mixing is limited, however, in that there is a certain "inertia," i.e., new knowers tend to talk almost exclusively to new knowers. This happens because a group of new knowers, told by individuals close together are also close together. It is as if the contacts "lingered" for a while, so that groups of new knowers being in the immediate vicinity of their informants, continued in the vicinity of each other and passed their information only to those who happened to wander close by.

The conclusion that under these conditions the apparent axon density becomes approximately unity indicates that the assumption of transitivity is too strong. In practically all the information diffusion experiments performed, $a(0)$ dropped from various values (ranging from 3 to 12) to a value of $a(1)$ ranging from 1 to 2, with only very occasional values greater than 2 or smaller than 1 recorded. It appears, therefore, that a weaker restriction on contacts than the one implied by our assumptions should be chosen.

On the other hand, a situation in which $a(t)$ drops from an arbitrary initial value to unity and remains there is easy to construct. Suppose all individuals are arranged linearly on a line N units long and are denoted by positive and negative integers. Suppose the individual denoted by 0 is the initial knower. Then $P(0) = 1/N$. Let each new knower now tell n individuals on his right and n on his left. Then, because of the complete overlap, the individuals contacted by the extremes on either side will be the only new knowers. Hence $P(t)$ will be $2n/N$ for all $t > 0$, and

$$x(t) = \frac{2nt+1}{N}. \tag{35}$$

Although a, the actual axon density, is approximately $2n$, the apparent axon density is given by

$$a(t) = \frac{N}{2n} \log \frac{1 - \dfrac{(2nt+1)}{N}}{1 - \dfrac{[2n(t+1)+1]}{N}} \cong 1 \tag{36}$$

for very large N and moderate t. In fact, under these conditions $a(t)$ will tend to rise slowly with t.

If we construct a similar model in two dimensions, $P(t)$ will grow linearly (as the perimeter of the area of knowers) and $a(t)$ will exhibit both an initial drop and a steady subsequent rise.

It seems, therefore, that a model combining some features of randomness and some of a spreading "wave" of knowers from a "focus of infection" should be able to account for the facts. Some such "mixed" models will be offered in part II.

This investigation is part of the work done under Contract No. AF 19(122)-161 between the U.S. Air Force Cambridge Research Laboratories and the University of Chicago.

LITERATURE

Rapoport, A. 1951. "Nets with Distance Bias." *Bull. Math. Biophysics*, **13**, 107–17.

Washington Public Opinion Laboratory. 1952. *Project Revere*. (Mimeographed report.) Volume III.

RECEIVED 6-18-53