

Sociometry is concerned with networks of relationship among groups of people. If the group is very large, the work of tracing all the relationships becomes tedious, and the task of describing the resulting net precisely becomes impossible. Here the problem of such large sociometric nets is approached with probabilistic and statistical methods.

A STUDY OF A LARGE SOCIOGRAM

by *Anatol Rapoport and William J. Horvath*

Mental Health Research Institute, The University of Michigan

THE use of mathematical models to describe social structure has certain inherent advantages. A model which describes all the essential features of a specific social organization may have a more general applicability in that it may be applied directly or with minor changes to describe the structure of similar organizations. In fact, such differences as are found by fitting similar models to a group of organizations may permit a quantitative differentiation into separate classes of what might otherwise have been considered identical structures.

It is our purpose in this study to apply a mathematical description to a body of data obtained by querying the population of a particular junior high school in Ann Arbor, Michigan, as to their friendship preferences. It is not expected that the actual numbers and detailed structural properties obtained in this study will be duplicated in any other junior high school, but certainly some of the gross structural properties which were found in this situation should apply to other similar organizations; and the particular approach described here could serve as a prototype for other similar studies.

DEFINITION OF A SOCIOGRAM

A sociogram is a description of a population in terms of relations between pairs of people in that population. This relation may be "likes" (used in this study), "dislikes," "chooses as work companion," or any of a number of others. This relation may be bivalent, e.g., "present or absent" or "positive or negative", or it may be

multivalent, e.g., "likes very much," "likes a little," "doesn't know," "dislikes," and "hates."

The set of values that the relation between A and B may take on is called the "range" of the function. The actual values are often expressed as numbers; 1 if the relation is present, 0 if it is absent, -1 if it is negative, -2 if it is strongly negative, etc. The domain of the function is the set of all ordered ($AB \neq BA$) pairs in the population. At least temporarily, A may like B while B dislikes A , so we must distinguish between the pair (A, B) and the pair (B, A).

If a relation is bivalent and it is the only one being considered, the sociogram can be represented either by a directed linear graph or as a matrix (α_{ij}), where the entry α_{ij} is the value of the relation which obtains between the ordered pair (i, j).

If the population is very large, the specification of the complete sociogram may be impractical; and even if such a description were available, the computational problems of treating linear graphs or matrices of such magnitude would be too formidable even for modern high-speed computers. The theoretician's interest, however, is seldom focused on a particular large sociogram (although particular small sociograms may sometimes be of great interest). Rather, the interesting features of large sociograms are revealed in their gross, typical properties. Thus one seeks to define classes of sociograms, or else to describe them by a few well-chosen parameters. It is perhaps natural to consider statistical parameters, since

one is interested in trends or averages, or distributions rather than particulars.

As for the taxonomy of large sociograms, this apparently involves problems of great complexity. It would seem offhand that a taxonomy of "nets" (the mathematical representations of sociograms) would arise naturally from the consideration of the statistical parameters, e.g., as a continuum of nets in the parameter space. But the statistical parameters themselves are singled out on the basis of taxonomic considerations, which have yet to be clarified. The nature of this methodological difficulty will become apparent in the course of our discussion.

The importance of an adequate theory of nets for quantitative sociology and social psychology is obvious. The social behavior of an individual is certainly strongly dependent on the behavior of others and perhaps most strongly on the behavior of *certain* others. The impacts upon an individual of the influence of other individuals follow the paths determined by the relevant sociometric net. These paths may be lines of authority, attraction, emotional involvement, etc. Also, the social behavior of the entire population is dependent on such patterns of relations, barring the trivial case when every member of the group is entirely and directly under the command of an outside authority, e.g., a platoon executing in unison the commands of a drill sergeant.

The spread of an attitude or of a piece of information through a large population is a case in point. There exist fairly sophisticated mathematical theories of contagion, in which many of the obvious parameters are included, e.g., infectivity, immunity, incubation lag, recovery or removal rates, etc. Missing from most (though not all) of these theories is a consideration of the contact structure of the population. By and large, in most theories of contagion, the population is assumed to be "well-mixed." That is to say, the probability of contagion in a given period of time between any two members of the population is taken as a parameter which may be a function of time (say the total duration of the contagion or the time since "infection" of the individuals concerned) but is the same for all pairs,

implying that the probability of contact is independent of the pair.

To whatever extent this may be a fair approximation to a realistic theory of contagion, certainly an important next step is the consideration of cases where the probability of contact is not the same for every pair. This involves the consideration of the contact structure of the net.

The assumption of equiprobability of contact or well-mixedness is equivalent to the assumption of random contact structure. A random net can be defined by the process of its construction as follows:

From each individual or "node" let there be a choices or "axones" issuing.¹ Each of these axones settles on a target node within the population. Which particular node a particular axone settles on is determined by a chance device selecting with equal probability from the entire population. The resulting net is a "random net" with an "axone density" of a .

A net which is not random will be called a "biased" net. Obviously a random net is a special case of a biased net in which all the biases are zero. It is not easy, however, to define the various biases rigorously even though their nature can be intuitively grasped. For if the choices of the terminal points of the axones are not equiprobable, there are innumerable other ways in which they can be assigned. In particular, they can be assigned "absolutely" or "relationally." An absolute assignment is one where the probability assigned to a node of being the target of an axone depends only on the former node. A relational assignment is one where this probability depends on both the potential target and the origin of the axone.

A bias would operate "absolutely" if sociometric choices depended entirely on the personal characteristics of the individuals, e.g., on their "popularity." On the other hand, if a bias were the reflection of a "social distance" between the individual choosing and the individual chosen, one would have a "relational" bias.

¹ We will assume that a is equal for every member of our nets, but the concept of random nets may be generalized by allowing a to be random also.

Several different forms of bias have been investigated (Rapoport, 1951, 1953, 1956, 1957; Solomonoff & Rapoport, 1951) and one previous experiment dealing with the determination of bias (Dodd, Rainboth, & Nehnevajsa, unpublished) has been reported.

THE DATA COLLECTION FOR THE PRESENT STUDY

To develop further the techniques used in the previous investigations, data for the sociogram of a junior high school in the Ann Arbor area were collected shortly after the beginning of the 1960-61 school year. Each pupil in both schools was asked to write his name, age, grade, and home room number on a card and to fill in the blanks in the statements:

- "1. My best friend in (name of school) Junior High School is _____
2. My second best friend. . . _____
- ...
8. My eighth best friend. . . is _____."

Naturally some pupils were absent the first day and an attempt was made to cover them by returning a week later. Even so, there remained some absentees and some children who, though present, did not fill out their cards. All of our analyses so far show that these "absentees" are a random sample from the population and so no bias was introduced by this less-than-perfect data gathering.

Of the many quantitative properties which may be studied in such a large sociogram, we will report in this paper on just two, the popularity distribution, and the "connectivity" of the net. The first of these is a distributional property and the second a structural one. Other measures were applied as well and some of these may be reported in a later paper.

DISTRIBUTION OF POPULARITY

Tabulations were made of the number of individuals who received a given number of votes as first, second, third, etc., friend. This is shown in Table 1. According to this tabulation 380 persons received no votes as first friend, 248 received 1 vote as first

TABLE 1
NUMBER OF PERSONS RECEIVING GIVEN NUMBER OF VOTES

Number of votes received	First friend	Second friend	Third friend	Fourth friend	Fifth friend	Sixth friend	Seventh friend	Eighth friend
0	380	399	413	413	407	417	418	439
1	284	257	263	265	257	252	264	244
2	136	145	112	116	133	122	118	115
3	42	33	38	35	41	50	35	44
4	11	19	22	18	11	12	17	8
5	5	3	10	8	10	2	3	6
6	0	1	0	3	0	2	4	3
7	0	1	1	1	0	2	0	0
8	0	1						
9	0							
10	1							
Unused ballots	98	101	113	120	119	129	147	173

friend, etc., 1 person receiving 10 votes as first friend.

The only consistent difference among the votes received for first, second, third, etc., friend as revealed by this table is the gradual increase in number of persons receiving zero votes in going from the first friend to eighth friend choices. This is largely accounted for by the increase in unmarked, missing, and mistaken ballots as the voting progressed from the first friend to the eighth friend. There were 98 such ballots in the first friend balloting and 173 in the eighth friend. Thus the increased number of persons receiving one and two votes as the first friend are transferred to the zero vote category in the eighth friend balloting by approximately the increased number of unused ballots.

The nature of the distribution function

A study of the mathematical form of the distribution of votes among the school population can give some insight into the process of choice used by the students in picking friends. For example, if everyone in school were equally popular and all students submitted ballots, then we would expect an average of one vote per person for the n th friend and the distribution of the number of zero, one, two, three, etc., votes would be given by the Poisson distribution with expected value 1; Ne^{-1} , $\frac{Ne^{-1}}{1!}$, $\frac{Ne^{-1}}{2!}$,

$\frac{Ne^{-1}}{3!}$, etc., where N is the total school population and e is 2.718. In Table 2 we have fitted Poisson distributions to the observed vote distributions for second and fourth friends using the expected values found from the data of votes per person of 0.882 and 0.860. Comparisons were also made for ballots for the other friends but these two sets of data typify the results. In all cases the Poisson distribution predicts a higher than observed value of one vote and lower values of zero and four, five, six and more votes. A Chi-square test shows that discrepancies of this magnitude are beyond the .05% level of chance fluctuations.

Greenwood and Yule (1920), in a study of accident statistics, have suggested a modification of the Poisson distribution to take into account the fact that not every person has the same expectation (of receiving a vote). The probability of receiving x votes is then given by

$$(1) \quad p(x) = \int_0^{\infty} p(x, z)p(z) dz.$$

$p(x, z)$ is the Poisson distribution with expectation z , and $p(z)$ is the probability that a person chosen at random has an expectation lying between z and $z + dz$. For the function $p(z)$ they chose the Pearson Type III function

$$(2) \quad p(z) = K e^{-\gamma z} z^{\alpha-1}$$

where α and γ are constants, and the nor-

malization constant, $K = \frac{\gamma^{\alpha}}{\Gamma(\alpha)}$. Evaluation of the function $p(x)$ with the distribution function given in (2) leads to the negative binomial distribution in (3) below:

$$(3) \quad p(x) = \binom{-\alpha}{x} \left(\frac{\gamma}{\gamma+1} \right)^{\alpha} \left(\frac{-1}{\gamma+1} \right)^x$$

The distribution (3) is sometimes called the Compound Poisson but since the Poisson distribution can be combined with other types of weighting functions, we shall call it the Greenwood-Yule distribution.

An entirely different derivation of (3) may be obtained from the theory of stochastic processes, where one assumes a time-dependent process such that the probability that at any given time, t , an event will occur in the interval $(t, t + dt)$ increases linearly with the number, x , of the events which have occurred in $(0, t)$, and for a given value of x the probability decreases with increasing value of t . This particular contagious distribution is called a Polya process and has been used to describe the morbidity in epidemics. It is interesting that the identical distribution can result from two very different processes; Greenwood and Yule assume that the events are mutually independent, and that the intensities vary from individual to individual, while Polya assumes that the events are stochastically dependent, the occurrence of an event increasing the probability that further events will occur. Thus, a good agreement

TABLE 2
COMPARISON OF OBSERVED AND FITTED VALUES

Number of votes per person	Second Friend			Fourth Friend		
	Observed value	Poisson distribution	Greenwood- Yule distri- bution	Observed value	Poisson distribution	Greenwood- Yule distri- bution
0	399	356	399	413	363	420
1	257	314	267	265	313	253
2	145	138	122	116	134	114
3	33	40.5	46.6	35	38.5	45.9
4	19	8.90	16.2	18	8.25	17.2
5	3	1.54	5.3	8	1.45	6.2
6	1	.23	1.6	3	.18	2.18
7	1	.029	.49	1	.02	.75
8	1	.0032	.15	0	—	—

TABLE 3
CONSTANTS FOR THE GREENWOOD-YULE
DISTRIBUTION

Friend	γ	α
1	4.45	3.95
2	3.18	2.80
3	2.41	2.09
4	2.35	2.02
5	3.59	3.09
6	2.91	2.47
7	2.99	2.48
8	2.73	2.18
Average all friends	3.03	2.64

between an observed distribution and the distribution (3) may be interpreted in two ways, and further analysis will be needed to determine which model offers the best explanation for the generation of the observed values.

The distribution (3) was tested against the same set of data shown in Table 2 for which the Poisson distribution gave a poor fit. The constants γ and α were evaluated using the observed means, \bar{x} , and variances, s^2 , according to the expressions²

$$(4) \quad \gamma = \frac{\bar{x}}{s^2 - \bar{x}} \quad \alpha = \gamma \bar{x}$$

The fitted values were then calculated and the results are presented in Table 2. The fit to the observed value is rather close. The Chi-square test shows agreement with the observed value at the 20% level for the second friend data and 60% level for the fourth friend. The fitted values for the other friends were all within this range of closeness.

The values of α and γ calculated from the relations (4) above have been tabulated in Table 3. Except for the fact that the values for the first friend are slightly higher than for the other choices, there is no noticeable trend in the values. The probability density function, $p(z)$, of equation (2) has been plotted in Fig. 1 for the first, third, fifth, and seventh friend choices. The curves are seen to be extremely close

together with no significant trend in going from the higher to the lower order of friends. The close agreement in the "Popularity Intensity Functions" among the different orders of friends suggests that an average curve derived from the data can be used to describe the situation for any order of friend. One might also infer from this that if a certain person has a high intrinsic "popularity intensity" for first friend, he would have this same value for second, third, etc., friend.

If we make the assumption that the same intensity function applies to all order of friends, then it is possible to calculate the joint probabilities of being chosen x_1 times as first friend and x_2 times as second friend. This will be given by

$$(5) \quad p(x_2/x_1) = \int_0^\infty p(x_2, z)p(x_1, z)p(z) dz$$

substituting the Poisson functions and the

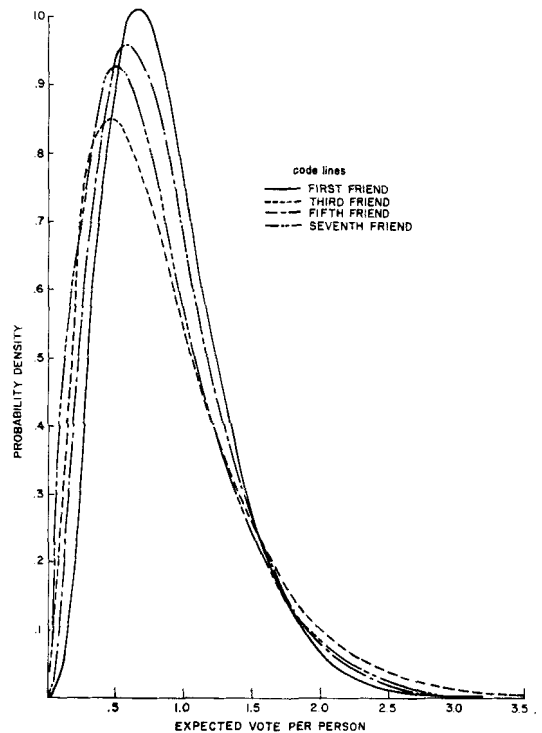


FIG. 1. Popularity intensity functions obtained from best fit of Greenwood-Yule function to observed popularity distribution.

² For approximate methods of fitting the negative binomial see Hald (1955, p. 727-731).

TABLE 4
COMPARISON OF OBSERVED AND CALCULATED NUMBER OF PERSONS RECEIVING GIVEN
NUMBER OF VOTES AS SECOND FRIEND AFTER RECEIVING A STATED NUMBER OF
VOTES AS FIRST FRIEND

First friend	0		1		2		3		4		5	
Second friend	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.
0	193	197	131	121	53	46.5	16	14.3	4	3.9	1	1.0
1	114	121	91	93	37	43.0	12	15.5	2	4.8	1	1.3
2	56	46.5	46	43.0	30	23.1	8	9.5	3	3.3	2	1.0
3	10	14.3	11	15.5	7	9.5	4	6.6	1	3.4	0	1.4
4	5	3.9	5	4.8	7	3.3	0	3.4	1	2.9	1	1.8
5	2	1.0	0	1.3	0	1.0	1	1.4	0	1.8	0	1.8

Intensity function from (2) we get

$$(6) \quad p(x_2/x_1) = \frac{\gamma^\alpha}{(2 + \gamma)^{\alpha+x_1+x_2}} \cdot \frac{\Gamma(\alpha + x_1 + x_2)}{\Gamma(\alpha)x_1!x_2!}$$

Table 4 compares the values calculated from this formula for the number of persons receiving a given number of votes as second friend after having received 0, 1, 2, 3, 4, or 5 votes as first friend. The values of α and γ were obtained by fitting the Greenwood-Yule function to the first friend data. The observed values are shown side by side in the same table. Although the observed and calculated values are quite close, a model assuming complete independence between the choice of first and second friends gives almost as good agreement. For example, 380 persons received no votes as first friend and 399 no votes as second friend. Assuming complete independence and random choice we would expect $\frac{380 \times 399}{859} = 179$ to receive zero votes for

first and second friend. This is not quite as good a fit to 193 as the 197 calculated by formula (6) but is actually fairly close. The number expected to receive 1 vote for both first and second friend on the assumption of random choice is 88 compared with 93 for the correlated choice model. Again the observed value, 91, falls between the two calculated values.

Apparently the observed Intensity Function produces only a minor perturbation in

TABLE 5
NUMBER OF PERSONS RECEIVING GIVEN NUMBER
OF VOTES FOR ALL EIGHT FRIENDS

Number of votes	Observed number of persons receiving votes	Calculated number
0	26	29.7
1	43	56.5
2	61	74.5
3	82	83.2
4	95	85.0
5	83	81.8
6	93	75.4
7	68	67.5
8	49	58.9
9	63	50.5
10	42	42.6
11	32	35.5
12	30	29.2
13	17	23.8
14	12	19.3
15	19	15.5
16	9	12.4
17	5	9.9
18	6	7.8
19	7	6.2
20	1	4.8
21	5	3.8
22	2	2.9
23	2	2.3
24	2	1.8
25	1	1.4
26	1	1.0
27	1	.8
28	1	.6
29	1	.5

the random choice model if only the first and second conditional choice probabilities are considered. However, we can also calculate the distribution of the number of votes

received for all eight friends on the assumption of an average Intensity Function which is identical for all friends. The convolution integral in this case can be reduced to the following rather simple expression:

$$(7) \quad P_s(n) = \left(\frac{\gamma}{8 + \gamma}\right)^\alpha \left(\frac{8}{8 + \gamma}\right)^n \cdot \frac{\Gamma(\alpha + n)}{n! \Gamma(\alpha)}$$

where n is the number of votes received for all eight friends, α and γ are the average values from Table 3, and $P_s(n)$ gives the probability that a person received n votes for all eight friends.

The observed and calculated values have been tabulated in Table 5. The agreement between formula (7) and the observed values is excellent. A Chi-square test gives a probability of 0.55 that the observed variations are due to chance. The independent choice model would predict only two persons to receive zero votes for all eight friends compared with the observed value of 26 and the calculated value of 30 from formula (7).

The conclusion to be drawn from this analysis is that the observed distribution of votes among the school population can be satisfactorily explained by the Compound Poisson distribution of Greenwood and Yule or alternatively as a Polya "Contagion Process." Whatever the mechanism leading to this result, it may be viewed as though each person had a characteristic popularity intensity which is the same for all ballots from the first to eighth friend.

CONNECTIVITY

Consider a net (random or biased) with axone density a (small) and population N (large). Arbitrarily select a small fraction of nodes p_0 . From these nodes ap_0N axones issue, which terminate on some set of nodes. Among these target nodes, some may be members of the starting set. Others will be newly contacted. Call the fraction of the population represented by the newly contacted nodes p_1 . Continue the process, calling each newly contacted fraction p_2 , p_3 , \dots , p_t , \dots . Thus the nodes represented

by p_t are those t times removed from the starters. Eventually some p_t will be zero because all of the nodes contacted on the previous remove will be among the p_j ($j = 0, 1 \dots t - 1$), and no new ones will have been contacted. At this point our tracing procedure is ended.

We shall be interested in the expected values of p_t and of the derived variables. $X = \sum_{j=0}^t p_j$ and its asymptotic value, X_∞ , as functions of a and of the bias parameters.

The expected values of p_t in a random net have been previously derived by one of us (Rapoport, 1951). They are given by the following iteration formula

$$(8) \quad p_{t+1} = (1 - X_t) (1 - e^{-ap^t}).$$

It has also been shown (Solomonoff & Rapoport, 1951) that the expected cumulated fraction of the population so traced satisfies the transcendental equation

$$(9) \quad X_\infty = 1 - (1 - p_0)e^{-aX_\infty}.$$

If the number of possible targets for each axone is reduced from N (the total population) to q (the size of an "acquaintance circle"), formulas (8) and (9) still hold even if $q \ll N$, so long as $q \gg 1$ and provided the intersection of the acquaintance circles of two individuals who are themselves acquainted has the same expected number of individuals in common as the intersection of two arbitrarily selected acquaintance circles.

On the other hand, if we assume that the acquaintance circles of two acquainted individuals are identical, this amounts to the assumption that the entire population "falls apart" into a set of mutually exclusive cliques of size q . If *within* the cliques, targets are arbitrarily chosen, we have several small random nets instead of a single large one. In a previous treatment of the problem (Rapoport, 1953), bias was defined in such a way that the "overlap" between the acquaintance circles of acquainted individuals was denoted by a parameter θ , $0 \leq \theta \leq 1$. The case $\theta = 0$ corresponded to the random net, while $\theta = 1$ was a reflection of a very tight bias

in which friends of friends were very likely (but not certain) to be friends. The iteration formula for the biased case will then be given by an expression identical to (8) except that an "apparent axone density" α replaces the a in (8). After the first remove α is given by the approximate expression

$$(10) \quad \alpha \cong 1 - e^{-a\theta} + (1 - \theta)a.$$

It is substantially independent of t and varies from $1 - e^{-a}$ for $\theta = 1$ (strongest bias) to a for $\theta = 0$ (random net).

"TRACING" THE SOCIOGRAM

The information obtained from the ballots returned by the students was coded and tabulated on punched paper tape so that it could be stored on the drum of an LGP 30 computer. A number of different programs were prepared for the analysis of the net structure. The particular data reported on in this paper were obtained by starting with a random sample of nine ballots and tracing two choices at a time, for example the n th and $(n + 1)$ th friends.

The number of new n th and $(n + 1)$ th friends contacted on each remove was determined until the chain ended either on a person previously contacted or a person who did not return a ballot. The whole procedure was then repeated on another random sample of nine. A total of at least 30 such random samplings were made and the chain of n th and $(n + 1)$ th friends was traced forward each time until no further new contacts were made and the chain terminated.

Since no behavioral process is involved in such a tracing, obviously explanations based on motivation and the like should not enter into the theory. Therefore we expect that if the theory of biased nets leading to equation (10) is correct, then the equation involving θ as a constant should adequately describe the tracing.

If we confined ourselves to a tracing over the first and second friends only, we obtained an average a of about 1.75 instead of 2.0, because of the axones "lost" in the blanks. (Absentees appeared in our population as cards with no blanks filled.) This reduction

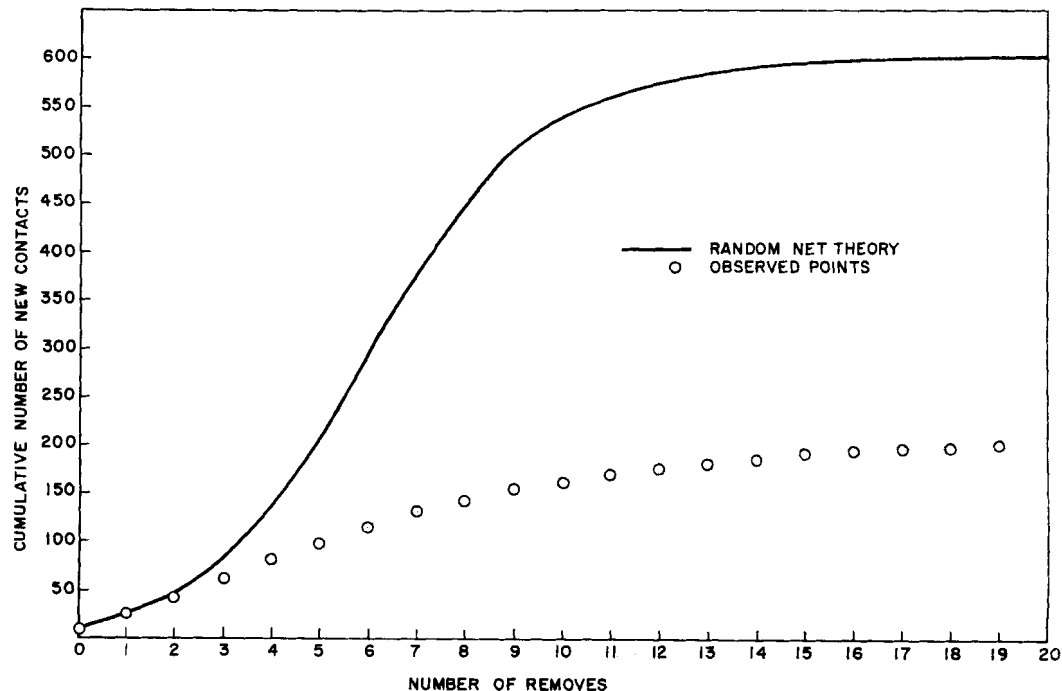


FIG. 2. Cumulative Number of New Contacts as a Function of the Number of Removes Compared with Random Net Theory. First and Second Friends.

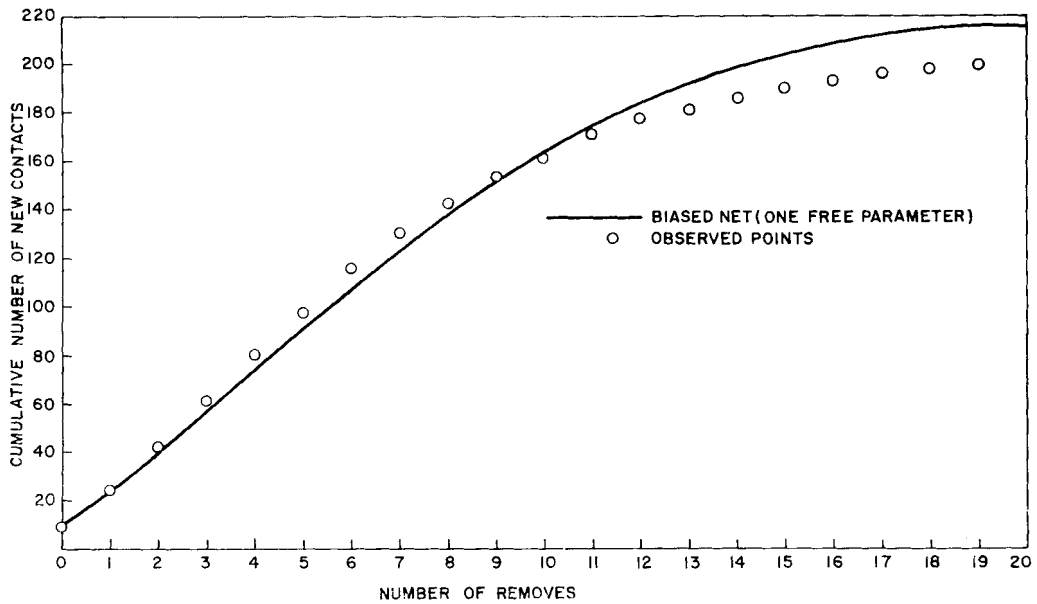


FIG. 3. Cumulative Number of New Contacts Compared with Biased Net Theory $\theta = 0.8$. First and Second Friends.

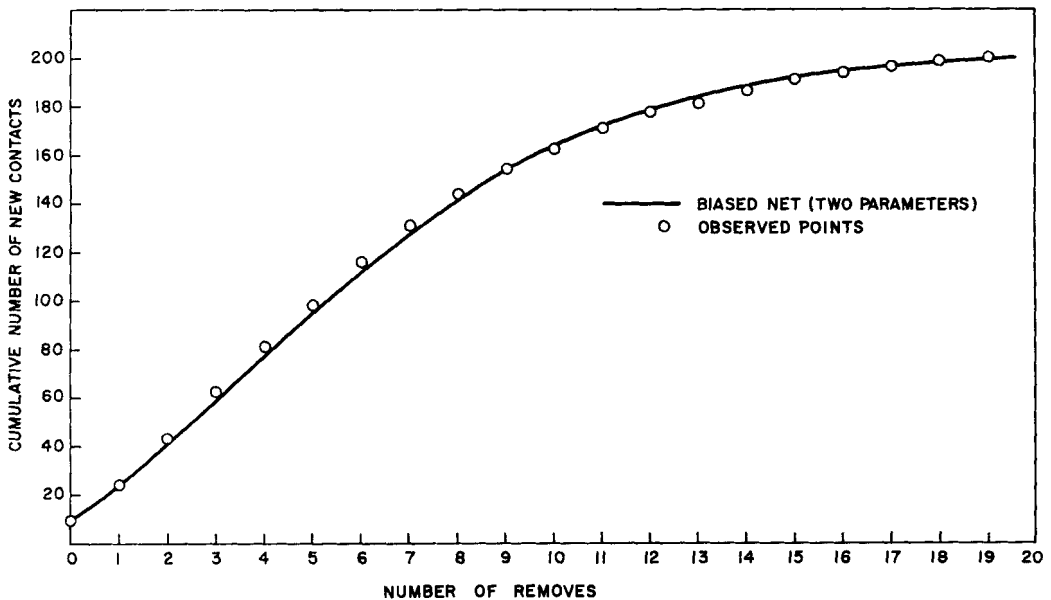


FIG. 4. Cumulative Number of New Contacts Compared with Biased Net Theory $\theta = 0.77$ and $N^* = 601$. First and Second Friends.

of the "intended" a to an "effective" a makes for no theoretical difficulty since, as we have noted, only the expected number of axones appears as a variable in the successive steps of our tracing.

Thus we have to deal with the following

parameters: N , the total population; a , the effective axone density, to be estimated by comparing p_0 and p_1 via equation (8); and θ , to be estimated so as to give the best fit to the remaining curve.

Fig. 2 shows the comparison of the data

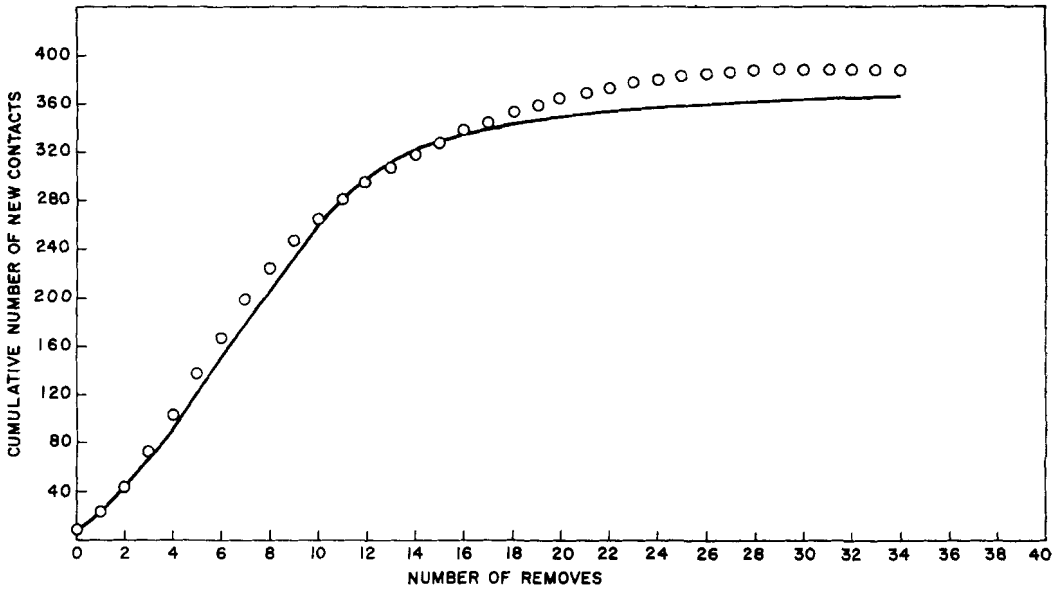


FIG. 5. Cumulative Number of New Contacts Compared with Biased Net Theory $\theta = .67$ and $N^* = 861$. Second and Third Friends.

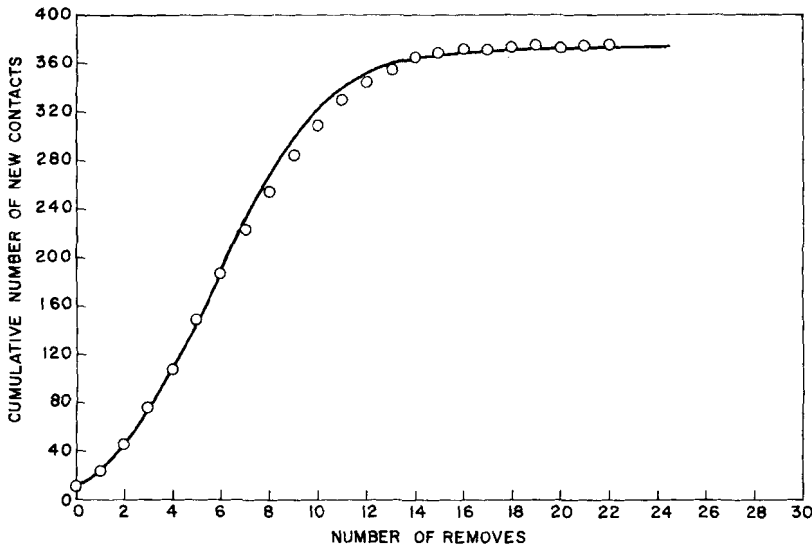


FIG. 6. Cumulative Number of New Contacts Compared with Biased Net Theory $\theta = .50$ and $N^* = 620$. Third and Fourth Friends.

with the null hypothesis, namely that the sociometric choice pattern constitutes a random net. There is little doubt that the hypothesis should be rejected.

Fig. 3 shows the same comparison under the assumption of a bias given by $\theta = 0.8$ (hence $\alpha(t) = 1.15$ for $t \geq 1$). There is a

marked improvement in the fit, but still a sizeable discrepancy between the theoretically predicted and the observed curves for X_t .

It appears, therefore, that "cliquishness," as measured by a single parameter, does not entirely account for the discrepancy between

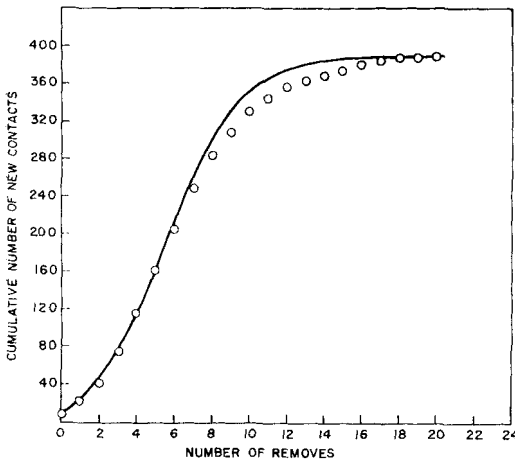


FIG. 7. Cumulative number of new contacts compared with biased net theory $\theta = .40$ and $N^* = 631$. Fourth and fifth friends.

implied by our acquaintance circle overlap parameter θ . We can, however, introduce a bias governing the *absolute* assignment of axones to targets, i.e., a "popularity" bias, which, one feels intuitively, ought to be independent of relational biases.

The question before us is how such a bias would be reflected in the *tracing* formula. Again a lack of rigorous theory compels us to venture a guess, namely, that the introduction of a popularity bias is tantamount to a reduction of the "effective" population, i.e., the value of N which is implicit in the calculation of the p_j and x_j . This interpretation is obvious in the special extreme case where the entire population consists of two classes, the "popular" and the "unpopular," meaning that the unpopular

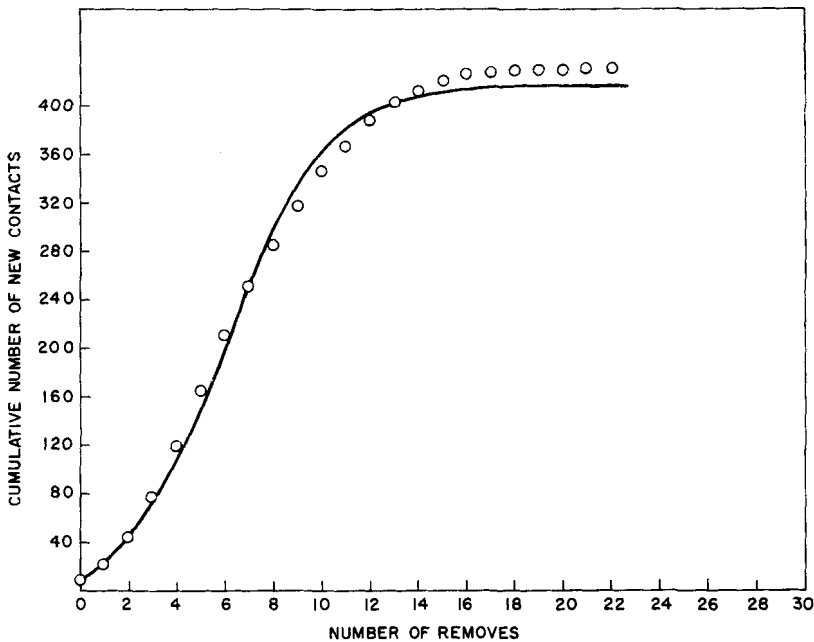


FIG. 8. Cumulative Number of New Contacts Compared with Biased Net Theory $\theta = .30$ and $N^* = 700$. Seventh and Eighth Friends.

the data obtained from average tracings and the expected tracing curve for a random net. The search for a second parameter seems indicated. But we cannot introduce a second *relational* parameter without risking a contradiction, for we have no way of knowing what relational biases are already

receive no sociometric choices, so that they all go to the popular.

Less extreme forms of the popularity bias should have a similar effect, namely a reduction of the "effective" population as it is reflected in the tracings. Introducing now $N^* = 601$ instead of $N = 861$ to calculate

TABLE 6
MEASURED VALUES OF CONSTANTS IN BIASED
NET THEORY

Rank order of friends	a	α	θ	N^*
1 and 2	1.78	1.15	.77	601
2 and 3	1.75	1.26	.67	861 ^a
3 and 4	1.72	1.44	.50	620
4 and 5	1.73	1.53	.40	631
7 and 8	1.62	1.50	.30	700

^a This value seems anomalous. It appears to indicate that popularity bias does not operate in the tracings of second and third friends. The value 861 was chosen as giving the best fit under the restriction $N^* \leq N$ and this fit is still the worst among the tracings. The remaining values of N^* increase with numerical rank order suggesting a weakening of the popularity bias. This apparently is not the case since direct examination of the distributions of sociometric choice show no such effect.

the p_j from the observed P_j , we obtain the theoretical curve for the average tracing through friends 1 and 2 as shown in Fig. 4.

We have now obtained the best possible fit to the data on an average tracing through first and second friends by adjusting two free parameters, α and N^* . There are, however, some restraints on both parameters. The parameter α is determined by θ and a (cf. equation 10). Of these, θ lies between 0 and 1, and a can be empirically determined by counting failures to name friends of the respective rank orders. It can also be computed from

$$(11) \quad a = \frac{1}{p_0} \log \frac{1 - x_0}{1 - x_1}.$$

The two quantities agree closely throughout.

The "effective" population N^* is subject to the restriction $N^* < N$ ($N = 861$), by our argument above concerning the nature of the popularity bias. Aside from these restrictions, θ and N^* (hence α and N^*) are free parameters. Curves as smooth as the cumulative tracing curves can probably be comfortably fitted by two almost free parameters regardless of the underlying model, so long as the mathematical function to be fitted has the required properties (i.e.,

almost a constant slope initially and horizontal asymptote). Thus the fits in themselves do not imply a strong corroboration of our theory. Intuitively we would expect, however, that θ should monotonically decrease as the numerical rank order of the friends through which tracings are made increases. This is because one would expect the friendship relations, and therefore the overlap bias of the acquaintance circles, to become less tight with increasing numerical rank order.

Since we have interpreted N^* as an "effective" population from which choices are actually made, we would not expect it to vary greatly with increasing rank order of the friendship, since, as we have previously shown, the popularity intensity function does not change significantly in going from the first to the eighth friend.

The data which are exhibited in Figs. 5-8 show comparisons between theoretical curves and data in tracings through successive consecutive pairs of friends. Table 6 summarizes the behavior of the parameters a , α , θ , and N^* .

We see that a [calculated from (11)] remains almost exactly constant in the first four tracings and drops somewhat in the last. This drop is exactly accounted for by a direct count of "failures to name." The parameter α increases in the first four tracings but remains about the same in the 7-8 tracing as in 4-5. However, since a decreases, θ also decreases. The monotone decrease of θ with rank order makes reasonable our interpretation of that parameter as a measure of tightness of acquaintance circles and corroborates our hypothesis concerning the relaxation of this tightness with increasing rank order. The effective population N^* , on the other hand, tends to increase steadily somewhat with rank order of the friendship choice, except for one anomalous value in the 2-3 tracing. Since we would expect no increase because of the constancy of the popularity bias, the significance, if any, of this variation cannot be explained in terms of our present theoretical framework.

REFERENCES

- Dodd, S. C., Rainboth, E. D., & Nehnevajsa, J. *Revere studies on interaction*. Washington Public Opinion Laboratory, unpublished.
- Greenwood, M., & Yule, G. U. An inquiry into the nature of frequency distributions representative of multiple happenings. *J. Roy. Stat. Soc.*, 1920, 83, 255-279.
- Hald, A. *Statistical theory with engineering applications*. New York: John Wiley, 1955.
- Rapoport, A. Nets with distance bias. *Bull. Math. Biophysics*, 1951, 13, 85-91.
- Rapoport, A. Spread of information through a population with sociostructural bias. *Bull. Math. Biophysics*, 1953, 15, 523-543.
- Rapoport, A. The diffusion problem in mass behavior. *General Systems*, 1956, 1, 48-55.
- Rapoport, A. Contribution to the theory of random and biased nets. *Bull. Math. Biophysics*, 1957, 19, 257-277.
- Solomonoff, R., & Rapoport, A. Connectivity of random nets. *Bull. Math. Biophysics*, 1951, 13, 107-117.

(Manuscript received June 30, 1961)



There are two equal and eternal ways of looking at this twilight world of ours: we may see it as the twilight of evening or the twilight of morning; we may think of anything down to a fallen acorn as a descendant or as an ancestor. There are times when we are almost crushed, not so much with the load of evil as with the load of the goodness of humanity, when we feel that we are nothing but the inheritors of an ancient splendor. But there are other times when everything seems primitive, when the ancient stars are only sparks blown from a boy's bonfire, when the whole earth seems so young and experimental that even the white hair of the aged, in the fine Biblical phrase, is like almond trees that blossom, like the white hawthorn grown in May. That it is good for a man to realize that he is "the heir of all the ages" is pretty commonly admitted; it is a less popular but equally important point that it is good for him sometimes to realize that he is not only an ancestor, but an ancestor of primal antiquity; it is good for him to wonder whether he is not a hero, and to experience ennobling doubts as to whether he is not a solar myth.

G. K. CHESTERTON, *A Defence of Nonsense*