Sankhyā : The Indian Journal of Statistics
1987, Volume 49, Series A, Pt. 2, pp. 141-188.

MEASURES OF RECIPROCITY IN A SOCIAL NETWORK

By A. RAMACHANDRA RAO and SURAJ BANDYOPADHYAY
Indian Statistical Institute

SUMMARY. Various measures of reciprocity in a social network are derived using graph—theoretic methods and deterministic models which can be used for comparing different populations. The limitations of the measures based upon deterministic models and the inappropriateness of the usual probabilistic measures are discussed with illustration from empirical sociological survey data.

1. Introduction

By a social network we mean the network of a sociological relation. Some aspects of social networks including models for social behaviour have been studied by several authors. To name a few, structural balance by Heider and Cartwright and Harary, clustering and hierarchy by Davis and Leinhardt, transitivity by Holland and Leinhardt, structural equivalence by Burt (1978) and Sailer (1978), cliques and connectivity by various people including Festinger, Forsyth and Katz and Luce, centrality by Bavelas and Freeman (1978), social power by Bartos and French, and strength of ties by Granovetter. Random graphs and inference from them have been studied by various authors including Frank and Capobianco. The references not explicitely given may be found in the books by Leinhardt (1977) and Holland and Leinhardt (1979).

However, there are other aspects of social networks which need to be examined in detail, specially in the context of empirical work. We propose to consider one such, namely, reciprocity in a social network. Reciprocity can be considered to be one indicator of balance or stability in social structure. It does not, however, imply cohesion in the sense of connectivity. While Heider's concept of balance deals with consistency of the attitudes in each triad in a signed network, reciprocity deals with the inter-dependence between the members of each pair in a social network.

In this paper, we study various measures of reciprocity in a social network, both existing and new, obtained by standardizing the number of reciprocal pairs in different ways. We compare them to bring out their merits and demerits so that we can make an appropriate choice. This

AMS (1985) *subject classification :* Primary 04C15, Secondary 05C20.

Key words and phrases : Reciprocity measure, Social network, directed graphs, probabilistic models, deterministic models.

problem arises in studies such as the incidence of mutual help and cooperation among the members of a population like a village, a community, or a social group. We mention that we are only considering how to measure reciprocity but are not proposing an overall model of social behaviour.

The problem of constructing a measure of reciprocity seems to have been considered first by Katz and Powell (1956). They proposed a measure and gave an application but their measure has a limitation which we shall discuss below and the application was too simple. Others like Karve and Damle (1963) collected interesting information on out-degrees and in-degrees in village social networks but did not attempt any detailed analysis. More recently, Holland and Leinhardt (1981) proposed an exponential family of distributions for a social network and gave a procedure to estimate one of their parameters, ρ, giving a measure of reciprocity. An earlier measure of Davis (1977) is a special case of this. But there seem to be some difficulties with these measures also which we shall discuss later.

Mathematically, we can take a social network to be a finite simple digraph. In an application such as referred to above, the vertices of the network may be the members of a population and an arc drawn from a vertex x to a vertex y if x takes help and cooperation from y. We shall sometimes use terminology suggested by this example though most of our discussion is applicable to the network of any sociological choice relation, indeed to the network of almost any irreflexive binary relation. Thus, we shall sometimes refer to vertices as households, arcs as ties, etc.

Reciprocity in a social network is indicated by the number s_0 of *reciprocal pairs* or *symmetric ties*, that is, unordered pairs of distinct vertices x and y such that both xy and yx are arcs of the network. However, s_0 cannot itself be used as a measure of reciprocity since it has to be standardized in order to make the values of the measure for different networks comparable. We consider the measures obtained by standardizing s_0 in different ways. The standardization is done by comparing the observed value of s_0 with other possible values in the given context. We thus have to assume something about the other possible networks, possibly even how the observed network arose. We shall first describe the various measures and then indicate the performance of each according to various criteria. The proofs will be given later on.

The methods of standardizing s_0 can be broadly divided into two categories : deterministic and probabilistic.

In the former type, we only specify which networks are possible but we do not specify how one of these is chosen. Then the most natural way of standardizing s_0 seems to be to find the minimum possible number of symmetric ties s_{min}, the maximum possible number of symmetric ties s_{max} and take the measure of reciprocity to be

$$s = 100 \times \frac{s_0 - s_{min}}{s_{max} - s_{min}}, \qquad \ldots \quad (1)$$

where we use the multiplicative factor 100 to express the measure as a percentage. Values of this measure near 0 correspond to low reciprocity and values near 100 correspond to high reciprocity.

In the latter type, we specify not only the possible networks but also the probability of occurrence of each of them. Usually a mechanism by which a network is chosen is specified. Then s_0 becomes a random variable and the best measure of reciprocity seems to be

$$s^* = 100.P(s_0 \leqslant \text{observed value}) \qquad \ldots \quad (2)$$

where the probability is obtained from the distribution of s_0 under the model. The interpretation of s^* is obvious. But usually the exact distribution of s_0 is difficult to determine. So one may have to use approximations. An alternative is to standardize s_0 in the usual statistical sense and consider

$$s^{**} = \frac{s_0 - \mu}{\sigma} \qquad \ldots \quad (3)$$

as the measure of reciprocity where μ and σ are the mean and standard deviation of s_0. We believe that the interpretation of s^{**} should still be through the distribution of s_0 (that is through s^*) but Chebychev's inequality gives us some idea even though the distribution is not known. A different way of standardizing s_0 will be used in the probabilistic models due to Katz and Powell (1956) and Holland and Leinhardt (1981).

We now briefly indicate the contents of the various sections. In Section 2, we give the deterministic measures and in Section 3 the existing probabilistic measures. In Section 4 we give a new probabilistic approach and the measure based on it by introducing the concept of a "potential set" (for each vertex) from within which it makes its choices instead of from the entire network. In Section 5 we discuss the various measures (existing and new) and compare them according to various criteria to bring out their merits and demerits. Subsection 5.1 deals with the deterministic measures while subsection 5.2 deals with the probabilistic measures. In Section 6 we mention some inequalities between the various measures under some conditions which are likely to be satisfied in empirical situations. In Section 7 we examine the applicability of each of the measures in empirical situations. For this purpose we consider the social networks of 21 villages in a central region of West Bengal and decide on the best measure in this context. We find that a deterministic measure (s_3) performs better than all the other measures. In Section 8 we show the relevance of the measure s_3 for sociological analysis. In Section 9 we present our conclusions.

The reader interested mainly in the sociological aspects may go directly to Section 8 after reading the introduction and taking a quick look at the measure s_3 (equation (13)) and the first three paragraphs of Section 7.

We now give the various methods of standardizing s_0 and the measures based on them. The first four are deterministic and the remaining probabilistic.

2. Deterministic measures

Measure I : Here we take the number of vertices N to be given. We denote the minimum and maximum possible values of s_0 by $s^{(1)}_{min}$ and $s^{(1)}_{max}$ where the superscript 1 refers to measure I. It is easy to see that in the absence of any other information

$$s^{(1)}_{min} = 0 \text{ and } s^{(1)}_{max} = \binom{N}{2} = \frac{N(N-1)}{2}. \qquad \ldots \quad (4)$$

Hence the measure of reciprocity, which we now denote by s_1, is

$$s_1 = 100 \frac{s_0}{s^{(1)}_{max}} = \frac{200 s_0}{N(N-1)}. \qquad \ldots \quad (5)$$

Measure II : Here we take the number of vertices N and the total number of ties m as given. It can be seen that

$$s^{(2)}_{min} = \max\left(0, m - \binom{N}{2}\right) \quad \text{and} \quad s^{(2)}_{max} = \left[\frac{m}{2}\right]. \qquad \ldots \quad (6)$$

where $[x]$ denotes the largest integer not exceeding the real number x. The measure s_2 is obtained by substituting $s^{(2)}_{min}$ and $s^{(2)}_{max}$ in (1) and is given by

$$s_2 = 100 \frac{s_0}{s^{(2)}_{max}} = \frac{200 s_0}{m-\varepsilon} \text{ when } m < \binom{N}{2} \qquad \ldots \quad (7)$$

where ε is 0 or 1 according as m is even or odd.

Measure III : This takes the out-degree d_i of the i-th vertex as given for all $i = 1, 2, \ldots, N$. We note that d_i is the number of choices made by the i-th vertex and it indicates its "expansiveness". The total number of ties m equals Σd_i and may be thought of as the expansiveness of the population as a whole. The values of $s^{(3)}_{\min}$ and $s^{(3)}_{\max}$ were determined by Achuthan, Rao and Rao (1984). To give these let us first arrange the d's in non-increasing order so that

$$d_1 \geq d_2 \geq \dots \geq d_N. \qquad \dots \quad (8)$$

For any t with $1 \leq t \leq N$, let

$$f(t) = \sum_{i=1}^{t} d_i - t(N-t) - \binom{t}{2}. \qquad \dots \quad (9)$$

Then

$$s^{(3)}_{\min} = \max f(t) \qquad \dots \quad (10)$$

where the maximum on the right is taken over $t = 0$, $t = N$ and all t such that $1 \leq t \leq N-1$, $d_t \geq N-t$ and $d_{t+1} < N-t$. Here we adopt the convention $f(0) = 0$. If $d_1 \leq \frac{N-1}{2}$, then $s^{(3)}_{\min} = 0$. To give $s^{(3)}_{\max}$, let

$$g(t) = \sum_{i=1}^{t} d_i - t(t-1) - \sum_{i=t+1}^{N} \min(t, d_i) \qquad \dots \quad (11)$$

for any t with $1 \leq t \leq N$. Then

$$s^{(3)}_{\max} = \left[\frac{1}{2} \sum_{i=1}^{N} d_i - \max g(t)\right] \qquad \dots \quad (12)$$

where the maximum of $g(t)$ is taken over $t = 0$ and all t such that $1 \leq t \leq N-1$, $d_t > d_{t+1}$. Again we use the convention $g(0) = 0$. If $d_1 \leq 2\sqrt{n}-2$ where n is the number of non-zero d_i's then $s^{(3)}_{\max} = [(\Sigma d_i)/2]$. Now the measure s_3 is obtained by substituting $s^{(3)}_{\min}$ and $s^{(3)}_{\max}$ from (10) and (12) in (1). Clearly

$$s_3 = \frac{s_0}{s^{(3)}_{\max}} = \frac{200 s_0}{\sum_{i=1}^{N} d_i - \varepsilon} = s_2 \text{ when } d_1 \leq 2\sqrt{n}-2 \qquad \dots \quad (13)$$

where ε is 0 or 1 according as Σd_i is even or odd.

Measure IV : The fourth method of standardizing s_0 takes the out-degree d_i and the in-degree e_i of the i-th vertex as given for all $i = 1, 2, \dots, N$. Incidentally, we note that while d_i indicates the expansiveness of the i-th vertex, e_i indicates its popularity. Clearly d_i is the number of choices made by the i-th vertex and e_i is the number of choices received by it. The suggestion that s_0 may be standardized with respect to both d's and e's appears to have been made first by Katz, Tagiuri and Wilson (1958). Now, the determination of the exact values of $s^{(4)}_{\min}$ and $s^{(4)}_{\max}$ seem to be quite difficult in the general case. However, Rao (1984) obtained some fairly good bounds : a lower bound for $s^{(4)}_{\min}$ and an upper bound for $s^{(4)}_{\max}$. Even though the values of those bounds are themselves difficult to calculate, Rao has also given some conditions under which they can be calculated very easily. He also gave an upper bound for $s^{(4)}_{\min}$ using which $s^{(4)}_{\min}$ can be determined in some cases. Having a lower bound for $s^{(4)}_{\min}$ and an upper bound for $s^{(4)}_{\max}$, we can also try to show that they are attained by actual construction. We now give these bounds. Let $V = \{1, 2, \ldots, N\}$ where we now assume that $d_i+e_i > 0$ for $i = 1, 2, \ldots, N$ (those i's for which $d_i = e_i = 0$ can be dropped). For any subset A of V, let

$$f(A) = \sum_{i\in A} e_i - \min\left\{ \sum_{i\in A} \min(|\bar{A}|, e_i), \sum_{i\notin A} \min(|A|, d_i)\right\} - \binom{|A|}{2}, \quad \ldots (14)$$

$$g(A) = \max\left\{ \sum_{i\in A} e_i - \sum_{i\in A} \min(|A|-1, d_i),\right.$$

$$\left.\sum_{i\notin A} d_i - \sum_{i\notin A} \min(|\bar{A}|-1, e_i)\right\}, \quad \ldots (15)$$

$$h(A) = \max\left\{ \sum_{i\in A} \max(e_i-d_i-|\bar{A}|, 0), \sum_{i\in A} \max(d_i-e_i-|\bar{A}|, 0)\right\} \quad \ldots (16)$$

and $$l(A) = \min\left\{ \sum_{i\in A} \min(|\bar{A}|, d_i), \sum_{i\notin A} \min(|A|, e_i)\right\} \quad \ldots (17)$$

where $\bar{A} = V-A$ and $|A|$ denotes the number of elements in A. Also let $f'(A)$ denote $f(A)$ with d's and e's interchanged. Then

$$s^{(4)}_{\min} \geqslant \max\{\max(f(A), f'(A), 0)$$
$$+\max(f(\bar{A}), f'(\bar{A}), 0)+\max(g(A)+g(\bar{A})-|A|\,.\,|\bar{A}|, 0)\}. \quad \ldots (18)$$

Some sufficient conditions for equality to hold in (18) can also be found in Rao (1984). They essentially say that the largest d_i is small compared to N, the largest e_i is not too close to N and the sum of the largest few e_i's is not too close to m. Regarding $s^{(4)}_{\max}$, we have

$$s^{(4)}_{\max} \leqslant \left[\frac{1}{2} (m-\max_A\{h(A)+h(\bar{A})+\max(g(A)-l(A), 0)+\max(g(\bar{A})-l(\bar{A}), 0)\})\right] \quad \ldots (19)$$

where $m = \Sigma d_i = \Sigma e_i$. It can be shown that

$$\text{R.H.S. of (19)} \leqslant \left[\frac{1}{2}\left(m- \sum_{i=1}^{N} \max(e_i-d_i, 0)\right)\right]$$

$$= \left[\frac{1}{2} \sum_{i=1}^{N} \min(d_i, e_i)\right]. \quad \ldots (20)$$

Thus the last expression in (20) is a simple upper bound for $s^{(4)}_{max}$. Some sufficient conditions for equality to hold in the inequality sign in (20) can be found in Rao (1984). They essentially mean that the maximum of the differences $e_i - d_i$ is not too large, the minimum of the differences $e_i - d_i$ is not too small and the differences are somewhat continuous between the minimum and the maximum. The measure s_4 is obtained by substituting the values of $s^{(4)}_{min}$ and $s^{(4)}_{max}$ in (1). Clearly

$$s_4 \doteqdot 100\, \frac{s_0}{s^{(4)}_{max}} \doteqdot \frac{200 s_0}{\Sigma \min (d_i, e_i)} \qquad \ldots \quad (21)$$

provided d's and e's satisfy some conditions indicated above. We mention one interesting fact here. Sometimes not all integer values between $s^{(4)}_{min}$ and $s^{(4)}_{max}$ are possible for s_0. But this situation occurs very rarely.

3. Existing probabilistic measures

Measure V : This is probabilistic and takes N as given. It stipulates that all the $2^{N(N-1)}$ possible networks on N vertices are equally likely. This can also be described thus : $P(i \text{ chooses } j) = \frac{1}{2}$ whenever $i \neq j$, and distinct pairs are independent. Now it is easy to see that s_0 has the binomial distribution $B\left(\binom{N}{2}, \frac{1}{4}\right)$. The mean and standard deviation, which we denote by $\mu_{(1)}$ and $\sigma_{(1)}$ respectively, are given by

$$\mu_{(1)} = \frac{N(N-1)}{8} \text{ and } \sigma_{(1)} = \sqrt{\frac{3N(N-1)}{32}}. \qquad \ldots \quad (22)$$

The measure s_1^{**} is obtained by substituting $\mu_{(1)}$ and $\sigma_{(1)}$ in (3). It is easy to show that

$$s_1^{**} \doteqdot \frac{3{\cdot}27 s_0}{N} - \frac{N}{2{\cdot}45} \qquad \ldots \quad (23)$$

when N is large ($N > 15$, say).

Measure VI : This takes N, m as given. It stipulates that all the $\binom{N(N-1)}{m}$ possible networks on N vertices with m arcs are equally likely. This amounts to saying that the m arcs are chosen at random and without replacement from the $N(N-1)$ possible arcs. Now the exact distribution of s_0 is not known but its mean and standard deviation are

$$\mu_{(2)} = \frac{m(m-1)}{2(M-1)} \text{ and } \sigma_{(2)} = \sqrt{\mu_{(2)}\left(1 - \mu_{(2)} + \frac{(m-2)(m-3)}{2(M-3)}\right)} \qquad \ldots \quad (24)$$

where $M = N(N-1)$. The measure s_2^{**} is obtained by substituting $\mu_{(2)}$ and $\sigma_{(2)}$ in (3). It can be shown that

$$\mu_{(2)} \doteq \frac{1}{2}\left(\frac{m-0\cdot5}{N-0\cdot5}\right)^2 \text{ and } \sigma_{(2)} \doteq \frac{1}{\sqrt{2}}\left(\frac{m-0\cdot5}{N-0\cdot5}\right)\left(1-\frac{m-1\cdot5}{M-2}\right) \quad \dots \ (25)$$

when N is not too small ($N > 10$, say). If, further, m/M is small ($m/M < 0\cdot1$, say) then $\sigma_{(2)}^2 \doteq \mu_{(2)}$ and

$$s_2^{**} \doteq \frac{\sqrt{2}\, s_0 N}{m} - \frac{1}{\sqrt{2}}\frac{m}{N}. \quad \dots \ (26)$$

Measure VII: This takes the out-degrees $d_1, d_2, \dots, d_N$ as given. It stipulates that all the $\prod_{i=1}^{N}\binom{N-1}{d_i}$ possible networks with the given out-degrees are equally likely. This is equivalent to saying that the i-th vertex makes its d_i choices at random and without replacement from all the vertices excluding itself ($i = 1, 2, \dots, N$) and that different vertices make their choices independently. The formulae for the mean and standard deviation of s_0 were obtained by Katz and Wilson (1956). To give these, let us write

$$t_k = \sum_{i=1}^{N} d_i^k \quad \text{for } k = 1, 2, \dots .$$

Then

$$\mu_{(3)} = \frac{1}{2(N-1)^2}(t_1^2 - t_2), \quad \dots \ (27)$$

$$\sigma_{(3)}^2 = \mu_{(3)} + \frac{1}{(N-1)^3(N-2)}(t_1^2t_2 - t_2^2 - 2t_1t_3 + 2t_4 - t_1^3 + 3t_1t_2 - 2t_3)$$

$$-\frac{1}{2(N-1)^4}(2t_1^2t_2^2 - 4t_1t_3 + 3t_4). \quad \dots \ (28)$$

The measure s_3^{**} is obtained by substituting $\mu_{(3)}$ and $\sigma_{(3)}$ in (3). It can be shown that when the d's do not differ much,

$$\mu_{(3)} \doteq \frac{N\bar{d}^2}{2(N-1)} \text{ and } \sigma_{(3)}^2 \doteq \mu_{(3)}\left(1 - \frac{\bar{d}}{N-1}\right)^2 \quad \dots \ (29)$$

where $\bar{d}$ denotes the average $\frac{1}{N}\sum_{i=1}^{N} d_i$. Since $m = N\bar{d}$, it can be seen from (25) and (29) that when N is not small and d's do not differ much, the value of s_3^{**} is close to that of s_2^{**}.

Measure VIII : This takes the out-degree d_i and the in-degree e_i of the i-th vertex as given for $i = 1, 2, \ldots, N$. It stipulates that all the possible networks with the given out-degrees and in-degrees are equally likely. However, this time even the expected value of s_0 is not known. Also there does not seem to be any easy way of generating a random network under this model. Thus one could think of the measure s_4^{**} based on this method but there is no (simple) way of calculating it.

Measure IX : The ninth method we consider, which is due to Katz and Powell (1956), is also probabilistic and takes the out-degrees $d_1, d_2, \ldots, d_N$ as given. It stipulates that there exists a parameter τ lying between -1 and 1 such that

$$E(s_0|\tau) = \frac{\sum\sum_{i<j} d_i d_j}{(N-1)^2}(1-\tau)+\frac{\Sigma d_i}{2}\tau. \quad \ldots \quad (30)$$

If τ is 0, the above expression coincides with $\mu_{(3)}$ given in (27). When all d_i's are equal to d (say), (30) can be derived from the assumption :

$$P\,(i \text{ chooses } j \text{ and } j \text{ chooses } i) = \frac{d}{N-1}\left(\frac{d}{N-1}+\tau\frac{N-1-d}{N-1}\right). \quad \ldots \quad (31)$$

Thus, basically, the method assumes that the i-th vertex makes its choices at random and without replacement from all the vertices excluding itself. But it does not make the assumption that choices of different vertices are independent. In an asymptotic sense to be explained below, τ may be regarded as an indicator of reciprocity or anti-reciprocity, the former corresponding to positive values of τ and the latter corresponding to negative values of τ. In fact, τ is the coefficient of correlation between X_{ij} and X_{ji} where X_{ij} is defined to be 1 if i chooses j and 0 otherwise.

By solving $s_0 = E(s_0|\tau)$, we obtain the following measure of reciprocity which is an unbiased estimator of 100τ :

$$\tilde{s}_3 = 100\,\frac{2(N-1)^2 s_0 - t_1^2 + t_2}{(N-1)^2 t_1 - t_1^2 + t_2} \quad \ldots \quad (32)$$

where t_1 and t_2 are as defined in Measure VII. If d_i's do not differ much then

$$\tilde{s}_3 \doteqdot 100\;\frac{2(N-1)s_0 - N\bar{d}^2}{N\bar{d}(N-1-\bar{d})}\;. \quad \ldots \quad (33)$$

If $\bar{d}$ is small compared to $N-1$ and $\bar{d}^2$ is small compared to $2s_0$ then $\tilde{s}_3 \doteqdot s_3$. This justified the statement made earlier regarding interpretation of τ. We mention that the model behind $\tilde{s}_3$ is incomplete in the sense that it specifies only the expected value of s_0 and not its distribution, not even its variance. We recall that the model behind the measure s_3 does not make any assumption about the distribution of s_0 whereas the model behind s_3^{**} specifies the distribution of s_0 completely. In this respect, $\hat{s}_3$ is mid-way between s_3 and s_3^{**}.

Measure X : Strictly speaking, there are four variants of this measure which is based on a probabilistic model introduced by Holland and Leinhardt (1981). They propose that the probability of observing any particular network is

$$\text{Const. exp} \left\{ \rho s_0 + \theta m + \sum_{i=1}^{N} \alpha_i d_i + \sum_{i=1}^{N} \beta_i e_i \right\} \qquad \ldots \quad (34)$$

where ρ, θ, α_i's and β_i's are real valued parameters and $\Sigma\alpha_i = \Sigma\beta_i = 0$. One of the important assumptions behind this distribution is that (X_{ij}, X_{ji}) and (X_{kl}, X_{lk}) are independent for distinct pairs $i < j$ and $k < l$ where X_{ij} is defined to be 1 or 0 according as i chooses j or not. It also assumes that

$$e^{\rho} = \frac{P(X_{ij} = 1 \mid X_{ji} = 1)}{P(X_{ij} = 0 \mid X_{ji} = 1)} \Bigg/ \frac{P(X_{ij} = 1 \mid X_{ji} = 0)}{P(X_{ij} = 0 \mid X_{ji} = 0)} \qquad \ldots \quad (35)$$

is independent of i, j. For the other assumptions and the derivation of the distribution, see the above referred paper. Equation (35) shows that ρ is a log-odds ratio and is an indicator of reciprocity. One can take the maximum likelihood estimator $\hat{\rho}_4$ of ρ based on the observed network as a measure of reciprocity. Though here s_0, m, d_i's and e_i's are all taken to be random variables (N is assumed given), still $\hat{\rho}_4$ eliminates the effect of expansiveness which is indicated by the parameter θ, the effect of the differences in out-degrees which is indicated by the α_i's and the effect of the differences in in-degrees which is indicated by the β_i's. Thus $\hat{\rho}_4$ is a probabilistic measure corresponding to the deterministic measure s_4. An iterative scaling algorithm for determining $\hat{\rho}_4$ is given in the above mentioned paper.

One can get probabilistic measures corresponding to s_1, s_2 and s_3 by making suitable assumptions on the parameters in (34). If we assume that $\theta = 0$ and, moreover, $\alpha_i = \beta_i = 0$ for all i, we get the measure

$$\hat{\rho}_1 = \log \left(\frac{3s_0}{\binom{N}{2} - s_0} \right)$$

corresponding to s_1. Note that $\hat{\rho}_1$ is a monotone function of s_1. Taking only $\alpha_i = \beta_i = 0$ for all i, we get the measure

$$\hat{\rho}_2 = \log\left(\frac{4s_0\left(\binom{N}{2}-m+s_0\right)}{(m-2s_0)^2}\right)$$

corresponding to s_2. A monotone function of this measure was suggested earlier by Davis (1977). The measure $\hat{\rho}_3$ is obtained by estimating ρ while assuming $\beta_i = 0$ for all i. However there are no explicit expressions for $\hat{\rho}_3$ and $\hat{\rho}_4$ which have to be computed by iterative methods. $\hat{\rho}_3$ is a function of s_0, N, m and d_i's while $\hat{\rho}_4$ is a function of those and e_i's.

4. A NEW PROBABILISTIC MEASURE

All the probabilistic approaches considered in Section 3 make (or imply) the assumption that choices are made at random from the entire population. Some of them further make an assumption of independence and some do not.

We now propose an alternate approach in which we do not make the assumption of random choices from the *entire* population (though we maintain the assumption of independence). This approach takes $d_1, d_2, \ldots, d_N$ as given and makes the following two basic assumptions :

(i) For each i, there is a non-empty set P_i of vertices (excluding i itself) from which i makes its d_i choices at random and without replacement. We may think of P_i as the set of potential choices for i.

(ii) Different vertices make their choices independently of each other.

Now s_0 can be standardised as in (3) to obtain the new measure which we will denote by $\breve{s}_3$. We note that the model behind s_3^{**} is a special case of the present model obtained by taking $P_i = \{1, 2, \ldots, N\} - \{i\}$ for all i. To give the formulae for the mean and variance of s_0 under this approach, we need some notation. Let

$$n_i = |P_i| \quad \text{and} \quad r_i = \frac{d_i}{n_i} \,(\leqslant 1). \qquad \ldots \quad (36)$$

Also let

$$S_j = \sum_{\substack{i \in P_j \\ \text{and} \\ j \in P_i}} r_i \quad \text{and} \quad T_j = \sum_{\substack{i \in P_j \\ \text{and} \\ j \in P_i}} r_i^2. \qquad \ldots \quad (37)$$

Then

$$E(s_0) = \frac{1}{2}\sum_{j=1}^{N} r_j S_j \qquad \ldots \quad (38)$$

and $$V(s_0) = \frac{1}{2}\sum_{j=1}^{N} r_j S_j - \frac{1}{2}\sum_{j=1}^{N} r_j^2 T_j - \sum_{j=1}^{N} \frac{r_j(1-r_j)}{(n_j-1)}(S_j^2 - T_j). \quad \ldots \quad (39)$$

Now $\check{s}_3$ can be calculated from (3) by using (38) and (39).

Though the measure $\check{s}_3$ requires the exact specification of P_i for each i, yet we can calculate $E(s_0)$ and $V(s_0)$ easily when some conditions are satisfied.

Assuming $r_j = r$ for all $j = 1, \ldots, N$, (38) and (39) reduce to

$$E(s_0) = \frac{1}{2} r^2 \left(\sum_{j=1}^{N} m_j \right) \quad \ldots \quad (40)$$

$$V(s_0) = \frac{r^2(1-r^2)}{2} \left(\sum_{j=1}^{N} m_j \right) - r^3(1-r) \sum_{j=1}^{N} \frac{m_j(m_j-1)}{(n_j-1)} \quad \ldots \quad (41)$$

where m_j denotes the number of i's such that $i \in P_j$ and $j \in P_i$.

If we further assume that

$$\frac{m_j-1}{n_j-1} = s$$

for all $j = 1, \ldots, N$, then the above equations simplify to

$$E(s_0) = \frac{rs}{2} m + \frac{r^2(1-s)}{2} N \quad \ldots \quad (42)$$

and $$V(s_0) = E(s_0)\{1-r^2-2r(1-r)s\}. \quad \ldots \quad (43)$$

If $s = 1$, (42) and (43) simplify to

$$E(s_0) = \frac{mr}{2} \quad \text{and} \quad V(s_0) = \frac{mr}{2}(1-r)^2. \quad \ldots \quad (44)$$

5. Discussion of the measures

Having described the various measures of reciprocity, we now give a summary of some of their properties in Table 1.

Range gives the possible values of the measure. It is continuous subject to the condition that s_0 is an integer, except in the case of s_4 and s_4^{**}. The neutrality point for a measure may be thought of as the value corresponding to "no tendency towards either reciprocation or anti-reciprocation". However since there is no specific neutrality point for the deterministic measures while it is 0 for all the probabilistic measures considered, we do not include it in the table. Sensitivity refers to the change in the measure per unit change in s_0. This is inversely related to the stability of the measure. We are not considering sensitivity to changes in the parameters since different

TABLE 1. PROPERTIES OF THE VARIOUS MEASURES OF RECIPROCITY

measure	specifications	range
(1)	(2)	(3)
	1.1 Deterministic measures	
s_0	Nil	0 to ∞
s_1	N	0 to 100
s_2	N, m	0 to 100
s_3	$N, m, d_1, \ldots, d_N$	0 to 100
s_4	$N, m, d_1, \ldots, d_N$ and $e_1, \ldots, e_N$	0 to 100 (may not be continuous)
	1.2 Probabilistic measures	
s_1^{**}	N given, P (i chooses j) $= \frac{1}{2}$. Distinct pairs independent	Depends on N (can vary from $-\infty$ to ∞, not symmetric about 0)
s_2^{**}	N, m given, m arcs chosen at random from $N(N-1)$ possible pairs	Depends on N, m (can vary from $-\infty$ to ∞, not symmetric about 0)
s_3^{**}	$d_1, \ldots, d_N$ given, $P(i \text{ chooses } j) = \frac{d_i}{N-1}$, Different vertices choose independently	Depends on d's (can vary from $-\infty$ to ∞, not symmetric about 0)
s_4^{**}	$d_1, \ldots, d_N$ and $e_1, \ldots, e_N$ given, all possible digraphs equally likely	Depends on d's and e's (can vary from $-\infty$ to ∞, not symmetric about 0)
$\tilde{s}_3$	$d_1, \ldots, d_N$ given, $P(i \text{ chooses } j) = \frac{d_i}{N-1}$, constant correlation between X_{ij} and X_{ji}	Depends on d's (can vary from -100 to 100, not symmetric about 0)
$\hat{\rho}_1$	N given, exp (ρs_0)	$(-\infty, \infty)$
$\hat{\rho}_2$	N given, exp $(\rho s_0 + \theta m)$	$(-\infty, \infty)$
$\hat{\rho}_3$	N given, exp $(\rho s_0 + \theta m + \Sigma \alpha_i d_i)$	$(-\infty, \infty)$
$\hat{\rho}_4$	N given, $\exp(\rho s_0 + \theta m + \Sigma \alpha_i d_i + \Sigma \beta_i e_i)$	$(-\infty, \infty)$
$\breve{s}_3$	d_i's and P_i's given, i makes d_i choices from P_i, different vertices choose independently	Depends on d_is and P_is, (can vary from $-\infty$ to ∞)

TABLE 1 (Contd.)

measure	interpretation	sensitivity	effect of adding isolated vertices
(1)	(4)	(5)	(6)
		Deterministic	
s_0	Direct	Not applicable	Nil
s_1	Direct	Very low	Decreases
s_2	Direct	Low	Cannot decrease $\left(\text{remains same when } m < \binom{N}{2}\right)$
s_3	Direct	Moderate	Cannot decrease $\left(\text{remains same when } d_{\max} < \frac{N-1}{2}\right)$
s_4	Direct	High	Nil
		Probabilistic	
s_1^{**}	Through cdf	Very low	Decreases
s_2^{**}	Through cdf	Low	Increases except when $s_0 \doteqdot \binom{N}{2}$
s_3^{**}	Through cdf	Moderate	Increases except when $s_0 \doteqdot \binom{N}{2}$
s_4^{**}	Through cdf	High	Nil
$\bar{r}_3$	100 times an intra-class correlation coefficient	Moderate	Increases
$\hat{\rho}_1$	Log-odds ratio	Low	Decreases
$\hat{\rho}_2$	Log-odds ratio	Low unless $s_0 \doteqdot 0$ or $\frac{m}{2}$.	Increases
$\hat{\rho}_3$	Log-odds ratio	?	Does not change
$\hat{\rho}_4$	Log-odds ratio	?	Does not change
$\breve{s}_5$	Through cdf	?	Does not change

TABLE 1. (Contd.)

measure	amenable to pooling when	nature of calculation	remarks
(1)	(7)	(8)	(9)
		Deterministic	
s_0	No symmetric ties between populations	Counting	Raw measure
s_1	Never	Simple	Combined measure of reciprocity and expansiveness
s_2	$m < \binom{N}{2}$ in each population and no ties between them	Simple	Eliminates the effect of global expansiveness
s_3	$d_{max} < 2\sqrt{n}-2$ in each population and no ties between them	Somewhat complicated (easy when $d_{max} < 2\sqrt{n}-2$)	Eliminates the effect of individual expansiveness
s_4	d's small and e's not too concentrated in each population, no ties between them	Complicated (approx. known when d's and e's are small)	Eliminates the effect of individual expansiveness and popularity
		Probabilistic	
s_1^{**}	Never	Simple	Same as for s_1
s_2^{**}	Never	Somewhat simple	Same as for s_2
s_3^{**}	Never	Complicated (approx. known when d's do not differ much)	Same as for s_3
s_4^{**}	Never	No method is known	Same as for s_4
$\tilde{s}_3$	Never	Somewhat simple	Same as for s_3
$\hat{\rho}_1$	Never	Simple	Eliminates the effect of size
$\hat{\rho}_2$	Never	Simple	Eliminates the effects of size and global expansiveness
$\hat{\rho}_3$	Never	Complicated (iterative)	Eliminates the effect of individual expansiveness
$\hat{\rho}_4$	Never	Complicated (iterative)	Eliminates the effect of individual expansiveness and popularity
$\breve{s}_3$	Always	Complicated (easy when $d_i \propto n_i$ and $m_i-1 \propto n_i-1$)	Eliminates the effect of individual expansiveness and the scope of individual choices

measures are based on different sets of parameters and so cannot be compared. By saying that a measure is amenable to pooling we refer to the following: Suppose we consider the disjoint union of some populations and we want to get the value of the measure for the union. Assume that ties occur only within the populations considered though, a priori, they could also occur between them. If the measure is a ratio, then under some mild conditions, the value of the measure for the union may be obtainable simply as the ratio of the sum of the numerators and the sum of the denominators for the populations. This implies that the value for the union is a weighted average of the values for the populations. We refer to this situation by saying that the measure is amenable to pooling.

We will now justify some of the statements made regarding the entries in Table 1. We will discuss the measures one by one.

5.1 *Discussion of the deterministic measures.* The measure s_1 corrects s_0 for the size of the population. However it assumes that s_0 can actually go up to $\binom{N}{2}$ and therefore will be good when the number of ties is of the order of $\frac{1}{2}\binom{N}{2}$. One may think of s_1 as an estimate of the probability that a pair chosen at random is connected by a reciprocal tie. The measure s_1 usually has very low sensitivity since $\binom{N}{2}$ is usually large. Clearly s_1 is not amenable to pooling since $\binom{N_1+N_2}{2}$ is larger than $\binom{N_1}{2}+\binom{N_2}{2}$.

The measure s_2 corrects s_0 for both the size of the population and the total number of ties. Thus it eliminates the effect of the expansiveness of the population as a whole. It measures reciprocity *among the observed ties.* Clearly s_2 has low sensitivity unless m is very close to 0 or very close to $N(N-1)$. If some isolated vertices are added, m and so $s^{(2)}_{\max}$ do not change and $s^{(2)}_{\min}$ remains the same or decreases, so s_2 either remains the same or increases. It is easy to see that s_2 is amenable to pooling across populations provided $m < \binom{N}{2}$ in each of them and there are no ties between any two of them. It may be noted however that $s^{(2)}_{\max}$ for a group of k populations can differ from the sum of the component $s^{(2)}_{\max}$'s by $k/2$ since, for the group, the integer part should be taken after adding the m's. But the effect of this is usually negligible. It is perhaps worth noting also that the pooled value of s_2 is in fact the weighted average of the values of s_2 for the populations, the weights being the corresponding $s_{\max}$'s, that is, the m's.

The measure s_3 corrects s_0 for the out-degrees of the various vertices. Thus s_3 eliminates the effect of the individual expansiveness on s_0. The measure s_3 will be appropriate if the d_i's can be considered to be characteristics of the respective vertices (or if d_i's are specified otherwise like when one is asked to name the three best friends). Whether or not d_i's are characteristics of the vertices can perhaps be studied by observing them over a period of time. Also for s_3 to be applicable, a tie xy must be the result of x choosing to go to y (that is, x taking the initiative). The measure s_3 is slightly more sensitive than s_2 in general but the sensitivity is low if the d's are small compared to the number of non-zero d's and Σd_i is not close to 0. If some isolated vertices are added, $s^{(3)}_{\max}$ does not change and $s^{(3)}_{\min}$ either remains the same or decreases, so s_3 either remains the same or increases. The calculation of $s^{(3)}_{\min}$ and $s^{(3)}_{\max}$ and so of s_3 may seem to be somewhat complicated in general but even in using equations (10) and (12), very few t's have to be considered when d's are small. When $d_1 \leqslant 2\sqrt{n}-2$, we have the simple formula (13). Using (10) and (12) it can also be shown that if any two d_i's differ by at most 1 then $s^{(3)}_{\min} = s^{(2)}_{\min}$ and $s^{(3)}_{\max} = s^{(2)}_{\max}$ and so $s_3 = s_2$. This would be true even if d's are large. For example, if $N = 6$ and $m = 20$ the range of s_0 is 5 through 10. If now the d's are given to be 4, 4, 3, 3, 3, 3 then also the range is 5 through 10. But if the d's are given to be 5, 5, 5, 2, 2, 1 then the range is 5 through 8. If the d's are given to be 5, 5, 5, 5, 0, 0 then s_0 can take only the value 6.

The measure s_4 corrects s_0 for the out-degree and in-degree of each vertex. Thus it eliminates the effect of not only each individual's expansiveness but also the effect of each individual's popularity, on s_0. The measure s_4 will be appropriate when both the d_i and e_i can be considered to be characteristics of the i-th vertex. The logic behind the measure s_4 assumes that both x and y have an active role with respect to each other in the formation of a tie $x \rightarrow y$, x by choosing to go to y and y by choosing to receive x.

The measure s_4 can be highly sensitive to small changes in s_0 since, when both in-degrees and out-degrees are fixed, the range of s_0 may be quite small. Also the sensitivity of s_4 can change much more than that of s_3 if the parameters are changed slightly (see Appendix for examples). We incidentally mention that out-degrees and in-degrees cannot be specified arbitrarily. Necessary and sufficient conditions for the existence of a simple directed graph with specified out-degrees and in-degrees can be found in Berge (1973), Chen (1976) and Harary, Norman and Cartwright (1965).

Another interesting phenomenon can occur sometimes when both the out-degrees and in-degrees are fixed. This is that *not* all integer values between

$s^{(4)}_{min}$ and $s^{(4)}_{max}$ may be attained by s_0. Suppose for example that the out-degree sequence and the in-degree sequence of a group of six households are both equal to

$$(2, 2, 2, 2, 1, 1). \qquad \ldots \quad (45)$$

Then it can be seen that all values from 0 to 5 are possible for s_0 except 4. We will only prove that 4 cannot be attained. Since the out-degree of each vertex equals its in-degree it follows that the number of unreciprocated arcs (entering or leaving) at each vertex is either 0 or at least 2. It follows that the total number of unreciprocated arcs is either 0 or at least 3, hence s_0 cannot take the value 4. However, the occurrence of a phenomenon like this may be very rare even in small populations.

It is obvious that the measure s_4 is not affected by the addition of isolated vertices. Also when d's are small, the in-degrees are not concentrated at too few vertices and $(e_i - d_i)$'s are somewhat continuous, then s_4 is amenable to pooling as seen from (21).

Regarding the calculation of the measure s_4, as already mentioned, the values of $s^{(4)}_{min}$ and $s^{(4)}_{max}$ are not known in general but some good bounds (namely, (18) and (19)) are available. However those bounds are also very difficult to calculate. But when the d's and e's satisfy some conditions, the lower bound for $s^{(4)}_{min}$ reduces to 0 and the upper bound for $s^{(4)}_{max}$ simplifies to the right hand side of (20). Then s_4 is approximately $200 s_0 / (\Sigma \min(d_i, e_i) - \epsilon)$ where ϵ is 0 or 1 according as $\Sigma \min(d_i, e_i)$ is even or odd. This will be the exact value of s_4 if

$$s^{(4)}_{min} = 0 \quad \text{and} \quad s^{(4)}_{max} = \left[\frac{1}{2} \Sigma \min(d_i, e_i)\right]. \qquad \ldots \quad (46)$$

One may verify the equalities in (46), when they are true, by what we call the *method of interchanges*. We now describe this method briefly.

A *simple interchange* in a digraph is the following: let u, v, x, y be four distinct vertices such that the arcs $u \to v$, $x \to y$ are present and the arcs $u \to y$ and $x \to v$ are not present in the digraph as shown in the left side of Figure 1. (In this figure we have used broken lines to denote the absence of an arc.) Then we delete the first two arcs and introduce the last two as shown in the figure. Clearly this simple interchange does not alter the out-degree and the in-degree of any vertex. Conversely it can be proved that if D_1 and D_2 are two digraphs with the same set of vertices and if the out-degree (resp. in-degree) of each vertex in D_1 is equal to its out-degree (resp.

in-degree) in D_2, then D_2 can be obtained from D_1 by a sequence of simple interchanges, see Chen (1976, pp. 420).

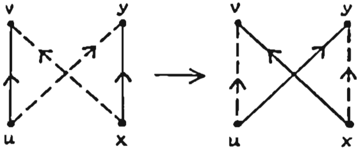


Fig. 1.

Given a lower bound δ for $s^{(4)}_{\min}$ we can try to show that it is attained as follows. Start with the observed social network and go on applying simple interchanges until we reach a network with δ symmetric ties. We usually consider only those simple interchanges which decrease the number of symmetric ties since then there cannot be any cycling. Occasionally we may have to use interchanges which do not decrease the number of symmetric ties. Sometimes, more general types of interchanges or shortcuts may be useful. For example, if there are two disjoint sets of vertices V_1 and V_2 with equal size such that every pair of vertices within V_i is symmetrically tied up ($i = 1, 2$) and there is no tie between V_1 and V_2, then the arcs within V_1 and the arcs within V_2 can be dropped and new unreciprocated arcs used to connect the vertices in $V_1 \bigcup V_2$ so that out-degrees and in-degrees are not altered. In a similar way, if we have an upper bound θ for $s^{(4)}_{\max}$ we can try to show that it is attained, by the method of interchanges. For this we have found the following shortcut very useful: find six vertices u, v, w, x, y, z such that $u \rightarrow v$, $v \rightarrow w$, $x \rightarrow y$, $y \rightarrow z$ are arcs and $u \rightarrow w$, $w \rightarrow v$, $v \rightarrow u$, $x \rightarrow y$, $y \rightarrow x$, $v \rightarrow y$, $y \rightarrow v$ are not arcs. Then delete the arcs $u \rightarrow v$, $v \rightarrow w$, $x \rightarrow y$ and $y \rightarrow z$ and introduce the arcs $u \rightarrow w$, $x \rightarrow z$, $v \rightarrow y$ and $y \rightarrow v$. This increases the number of symmetric ties.

Having discussed s_3 and s_4 in some detail, suppose we are asked which of them is to be preferred. We think there is no answer which is *universally* valid. For, suppose the observed social network is the second one in Figure 2. Then using the measure s_3 one could argue that there is low reciprocity (the value of s_3 is 20·0) and the fact that this is due to the differential popularity of the vertices is a different matter. Again, using s_4 one might retort that what is intuitively felt need not be correct and could insist that there is high reciprocity ($s_4 = 100{\cdot}0$). See Appendix for derivation of the values of s_3 and s_4. In the figures we have indicated symmetric ties by lines without arrow marks.

5.2 *Discussion of the probabilistic measures.* We will now go on to measures based on the probabilistic methods. It is easy to see that the range of each of the measures s_i^{**}, $\hat{\rho}_i$ $(i = 1, 2, 3, 4)$ and $\check{s}_3$ is contained in $(-\infty, \infty)$, while the range of $\bar{s}_3$ is contained in $[-100, 100]$. It is also clear that all these have a specific neutrality point, namely 0. For s_i^{**} and $\check{s}_3$ this corresponds to the observed s_0 being equal to its expected value under the model. For $\bar{s}_3$ and $\hat{\rho}_i$ the neutrality point corresponds to the parameter (τ or ρ) being 0. The neutrality point is usually interpreted as 'no tendency toward either reciprocation or anti-reciprocation'. This conclusion would be valid when it is assumed that choices are made at random from the entire

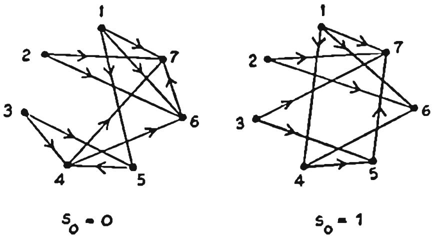


Fig. 2.

population (or from the sets P_i in the case of $\check{s}_3$). We will return to this later. The interpretation of any s_i^{**} should be through the tail probability s_i^{*}. The same holds for $\check{s}_3$ but $\bar{s}_3/100$ is to be interpreted like a correlation coefficient, rather like an intra-class correlation coefficient and the interpretation of $\hat{\rho}_i$ is that of a log-odds ratio. Clearly none of the measures based on the probabilistic models considered above, with the exception of $\check{s}_3$, is amenable to pooling since for any population such measures assume that choices are made at random from the entire population. This can also be seen from the formula of each of these measures since the mean (resp. standard deviation) for a union of populations cannot be obtained by adding the means (resp. standard deviations) for the populations. However, the measure $\check{s}_3$ is amenable to pooling whenever the potential set P_i for each vertex i is contained in the population to which i belongs.

The formula (22) and the approximation (23) for s_1^{**} are easy to prove. When isolated vertices are added, $s_0/\sigma_{(1)}$ decreases and $\mu_{(1)}/\sigma_{(1)}$ increases, hence s_1^{**} decreases. The measure s_1^{**}, like s_1, will be good only when the number of ties is of the order of $\frac{1}{2}\binom{N}{2}$.

The formulae (24) used for s_2^{**} can be derived as follows. We can write

$$s_0 = \sum_{1 \leqslant i < j \leqslant N} Z_{ij} \quad \ldots \quad (47)$$

where Z_{ij} is defined to be 1 if the i-th and j-th vertices are symmetrically tied up and 0 otherwise. Since m arcs are chosen at random and without replacement from the $M = N(N-1)$ possible arcs, it is easy to see that

$$P(Z_{ij} = 1) = \frac{m(m-1)}{M(M-1)}$$

and

$$P(Z_{ij} = 1 \text{ and } Z_{kl} = 1) = \frac{m(m-1)(m-2)(m-3)}{M(M-1)(M-2)(M-3)}$$

for any two pairs i, j and k, l even if they have a common element. Now using simple results on expectation and variance, (24) can be proved. From (24) it can be shown that $\sigma^2_{(2)} \leqslant \mu_{(2)}$. The approximation (25) is obtained essentially by replacing $m(m-1)$ by $(m-0{\cdot}5)^2$, etc. When m/M is small, it is easy to see that $\sigma^2_{(2)} \doteqdot \mu_{(2)}$ and perhaps the distribution of s_0 is approximately Poisson, see Katz, Tagiuri and Wilson (1958). It can be shown (by considering the derivative of s_2^{**} with respect to N) that s_2^{**} increases when isolated vertices are added unless $s_0 \doteqdot \frac{m}{2} \doteqdot \frac{N(N-1)}{2}$ before the isolated vertices are added.

The formulae (27) and (28) used for finding s_3^{**} were obtained by Katz and Wilson (1956) as already mentioned. They also showed that equality holds in (29) when all d_i's are equal. If the d's do not differ much and are small compared to N then s_3^{**} is close to s_2^{**} since $\mu_{(3)}$ and $\mu_{(2)}$ are both close to $\frac{1}{2}\bar{d}^2$ and $\sigma^2_{(3)}$ and $\sigma^2_{(2)}$ are both close to $\frac{1}{2}\bar{d}^2\left(1-\frac{\bar{d}}{N-1}\right)^2$. But we have noticed that s_3^{**} is close to s_2^{**} provided only that d's do not differ much, even when d's are large compared to N. The situation here is similar to that between s_2 and s_3 discussed earlier.

Regarding the measure s_4^{**}, practically nothing is known about it. Even $\mu_{(4)}$ seems to be very difficult to find. The basic difficulty here is that for any two vertices x and y, the probability that x chooses y is a complicated function of the out-degrees and in-degrees of not only x and y but of all the vertices in the network. Simulation studies also seem to be difficult here since there does not seem to be any way of generating a random digraph with given out-degrees and in-degrees except by first listing all possible digraphs. No formula is known even for the number of such digraphs though Katz and Powell (1954) gave a complicated way of calculating this number.

The model behind the measure $\tilde{s}_3$ essentially assumes that the i-th vertex makes its d_i choices at random and without replacement from all other vertices. But it does not assume that different vertices choose independently. Instead it assumes that there is a common correlation coefficient τ between X_{ij} and X_{ji} whenever $i \neq j$, where

$$X_{ij} = \begin{cases} 1, & \text{if } i \text{ chooses } j, \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that

$$E(X_{ij}) = \frac{d_i}{N-1}, \quad V(X_{ij}) = \frac{d_i}{N-1} \cdot \frac{N-1-d_i}{N-1}$$

and

$$P(X_{ij} = 1 \text{ and } X_{ji} = 1) = \frac{d_i d_j}{(N-1)^2} + \tau \sqrt{\frac{d_i d_j}{(N-1)^2} \frac{(N-1-d_i)}{N-1} \frac{(N-1-d_j)}{N-1}}. \quad \ldots \quad (48)$$

Clearly (48) reduces to (31) when all d_i's are equal to d. Also then

$$\begin{aligned} E(s_0 \mid \tau) &= \sum_{i<j}\sum P(X_{ij} = X_{ji} = 1) \\ &= \frac{Nd^2}{2(N-1)} + \tau \frac{Nd(N-1-d)}{2(N-1)} \\ &= \frac{Nd^2}{2(N-1)} (1-\tau) + \frac{Nd}{2} \tau. \end{aligned} \quad \ldots \quad (49)$$

Equation (30) is a generalization of (49) to the case when d_i's are not necessarily equal and (32) can be obtained easily by solving $E(s_0 \mid \tau) = s_0$ for τ. It is easy to establish (33) when d_i's are nearly equal. It is also easy to see that $\tilde{s}_3 \leqslant s_3$ at all times. As already mentioned, the model does not specify the distribution of s_0, not even the variance of s_0 which involves $P(X_{ij} = X_{ji} = X_{kl} = X_{lk} = 1)$ for any i, j, k, l. It can easily be seen that when all d_i's are equal to d, the maximum value of $\tilde{s}_3$ is 100 corresponding to $s_0 = Nd/2$ but the minimum depends on d. It is $-100d/(N-1-d)$, attained when $s_0 = 0$, if $d < (N-1)/2$. If $d > (N-1)/2$ then the minimum value of $\tilde{s}_3$ is $-100(N-1-d)/d$ attained when $s_0 = Nd - \binom{N}{2}$. When d_i's are not all equal, both the maximum and minimum values of $\tilde{s}_3$ depend on d's. The maximum would be 100 if $s^{(3)}_{\max} = \Sigma d_i/2$ and the minimum would be -100

if all d_i's are $(N-1)/2$. In any case, $\tilde{s}_3$ lies between -100 and 100. The interpretation of $\tilde{s}_3$ is clearly like that of a correlation coefficient, rather 100 times an intraclass correlation coefficient. It can be shown by differentiating the expression of $\tilde{s}_3$ with respect to N that the value of $\tilde{s}_3$ increases when isolated vertices are added.

Each of the measures $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \hat{\rho}_4$ makes the important assumption that i choosing j is statistically independent of i choosing k. Further they assume that the probability that i chooses j depends only on ρ, θ, α_i, α_j, β_i and β_j. Hence if all α_i's are 0 and all β_i's are 0 (corresponding to all out-degrees being equal and all in-degrees being equal) then the probability that i chooses j is independent of i and j. Thus, in a way, these models assume random choices from the entire population though taking care of the overall expansiveness, the differences in the out-degrees and the differences in in-degrees. The interpretation of $\hat{\rho}_i$ is that it is a log-odds ratio. One can actually get another measure comparable to $s_i^*/100$ by using the log likelihood ratio for testing the hypothesis $\rho = 0$ against the appropriate alternative and getting the tail probability under its null distribution. The null distribution is approximately χ_1^2 as pointed out by Holland and Leinhardt (1981). But we shall not pursue this.

Clearly $\hat{\rho}_1$ decreases when isolated vertices are added while $\hat{\rho}_2$ increases. The measure $\hat{\rho}_2$ treats the null pairs also as reciprocal. However, our idea of reciprocity is i and j going to each other, not each of i and j *not going* to the other. The measures $\hat{\rho}_3$ and $\hat{\rho}_4$ do not change when isolates are added because α_i and β_i are taken to be $-\infty$ for isolates and thus isolates are left out of consideration while fitting the distribution (34). We now give a simple illustration to demonstrate that none of $\hat{\rho}_1, \ldots, \hat{\rho}_4$ is amenable to pooling. Consider a network with $N = 10$ vertices, $s_0 = 2$ and $d_i = e_i = 2$ for all i. For such a population $\exp \hat{\rho}_1 = 0{\cdot}14$ and $\exp \hat{\rho}_i = 0{\cdot}84$ for $i = 2, 3, 4$. If we now consider a disjoint union of two such populations, then for the union, $\exp \hat{\rho}_1 = 0{\cdot}065$ and $\exp \hat{\rho}_i = 2{\cdot}4$ for $i = 2, 3, 4$. Thus the value of any one of these measures, for the union, is not an average of the values for the populations. So none of these measures is amenable to pooling. In fact, as we take the union of more copies, $\hat{\rho}_1$ goes on decreasing and each of $\hat{\rho}_2, \hat{\rho}_3$ and $\hat{\rho}_4$ goes on increasing.

We now give the proof of (38) and (39) used in the definition of $\tilde{s}_3$. For this we define

$$X_{ij} = \begin{cases} 1, & \text{if } i \text{ chooses } j \\ 0, & \text{otherwise} \end{cases}$$

as before and

$$Z_{ij} = X_{ij}X_{ji}.$$

Thus Z_{ij} is 1 if i chooses j and j chooses i and 0 otherwise. Clearly then

$$s_0 = \frac{1}{2} \sum\sum_{i \neq j} Z_{ij}.$$

Also,

$$E(Z_{ij}) = P(Z_{ij} = 1) = \begin{cases} r_i r_j, & \text{if } j \in P_i \text{ and } i \in P_j \\ 0, & \text{otherwise.} \end{cases}$$

Hence

$$E(s_0) = \frac{1}{2} \sum\sum_{\substack{j \in P_i \\ i \in P_j}} r_i r_j = \frac{1}{2} \sum_{j=1}^{N} r_j S_j$$

which is (38). Also

$$V(Z_{ij}) = \begin{cases} r_i r_j (1 - r_i r_j), & \text{if } j \in P_i \text{ and } i \in P_j \\ 0, & \text{otherwise.} \end{cases}$$

To find the covariances, we note that if i, j, k, l are all distinct then $\text{cov}\,(Z_{ij}, Z_{kl}) = 0$. If i, j, k are distinct, then

$$E(Z_{ij}Z_{jk}) = P(Z_{ij} = 1 \text{ and } Z_{jk} = 1)$$

$$= \begin{cases} r_i r_j r_k \dfrac{(d_j - 1)}{(n_j - 1)}, & \text{if } i,\ k \in P_j,\ j \in P_i \text{ and } j \in P_k \\ 0, & \text{otherwise.} \end{cases}$$

Hence

$$\text{cov}(Z_{ij}, Z_{jk}) = \begin{cases} -\dfrac{r_i r_j r_k (1 - r_j)}{n_j - 1}, & \text{if } i,\ k \in P_j,\ j \in P_i \text{ and } j \in P_k \\ 0, & \text{otherwise.} \end{cases}$$

Now

$$s_0 = \sum\sum_{\substack{i < j \\ i \in P_j,\, j \in P_i}} Z_{ij}.$$

So

$$V(s_0) = \frac{1}{2} \sum\sum_{\substack{i \in P_j, \\ j \in P_i}} r_i r_j (1 - r_i r_j) - 2 \sum\sum\sum_{\substack{i < j < k \\ i, k \in P_j, j \in P_i, j \in P_k}} \frac{r_i r_j r_k (1 - r_j)}{n_j - 1}$$

$$-2 \sum\sum\sum_{\substack{i < j < k \\ j, k \in P_i, i \in P_j, i \in P_k}} \frac{r_i r_j r_k (1 - r_i)}{n_i - 1} - 2 \sum\sum\sum_{\substack{i < j < k \\ i, j \in P_k, k \in P_i, k \in P_j}} \frac{r_i r_j r_k (1 - r_k)}{n_k - 1}$$

$$= \frac{1}{2} \sum\sum_{\substack{i \in P_j, \\ j \in P_i}} (r_i r_j - r_i^2 r_j^2) - \sum_{j=1}^{N} \frac{r_j (1 - r_j)}{n_j - 1} \sum\sum_{\substack{i \neq k \\ i, k \in P_j, j \in P_i, j \in P_k}} r_i r_k.$$

Hence (39) follows.

Equations (40) and (41) are proved easily when $r_j = r$ for all j. If we further assume $(m_j-1)/(n_j-1) = s$, we have

$$\sum_{j=1}^{N} m_j = \sum_{j=1}^{N} (1+s(n_j-1)) = s\left(\sum_{j=1}^{N} n_j\right)+N(1-s)$$

$$= s\left(\sum_{j=1}^{N} \frac{d_j}{r}\right)+N(1-s) = \frac{sm}{r}+N(1-s)$$

and so (42) and (43) follow.

From the formulae (38) and (39) it is clear that if $d_j = 0$ for some j then the corresponding set P_j is of no consequence. This shows that $\breve{s}_3$ is not altered when isolated vertices are added.

The exact distribution of s_0 under the model behind $\breve{s}_3$ is difficult to determine but we think normal approximation will be fairly good when N is large and d_i's are not too close to n_i's on the whole. In any case, we can get some idea of the tail probabilities from Chebychev's inequality.

The main difficulty one would face in using the measure $\breve{s}_3$ would be in deciding the sets P_i. We mention some possibilities. The sets P_i may be ascertained at the time of the survey. Or, they may be specified by the model being tested like : independent random choices from within one's own kins and caste members. They may also be specified by other theoretical considerations based on the network itself.

Fortunately, under the assumptions that $r_i = r$ for all i (this means that each vertex makes choices in a fixed proportion to the number of its potential choices) and $(m_j-1)/(n_j-1) = s$ (this means that roughly, a fixed proportion of the potential choices of any vertex can reciprocate with it), $E(s_0)$ and $V(s_0)$ can be calculated from (42) and (43) even without exact knowledge of the sets P_i. Since we think these two conditions may be assumed to hold at least approximately, we may use (42) and (43) to find the value of $\breve{s}_3$. We still have to specify the values of r and s. In the absence of prior information about r and s, an r between 0·3 and 0·8 and an s between 0·6 and 1·0 could be a reasonable choice in most situations.

6. Inequalities between the various measures

Comparing pair-wise the various measures considered above, the following string of inequalities can be proved under some mild conditions :

$$s_1^{**} < s_1 < \breve{s}_3 < s_3 = s_2 < s_4 \qquad \ldots \quad (50)$$

Also under similar assumptions,

$$s_3^{**} \doteqdot s_2^{**} \doteqdot \frac{N}{141}\, s_3. \qquad \cdots \quad (51)$$

We briefly indicate the proofs of (50) and (51). If $N > 15$ and $s_0 < N^2/8$ then $s_1^{**} < 0$ from (23) and so $s_1^{**} < s_1$. If $15 < N < 62$ then $s_1^{**} < \frac{3 \cdot 27 s_0}{N} < s_1$ from (5) and (23). If d's are small compared to N and $s_0 > \frac{1}{2}\bar{d}^2$ then $s_1 < \tilde{s}_3$ from (5) and (33). The inequality $\tilde{s}_3 < s_3$ can be proved from (13) and (32) if $d_1 \leqslant 2\sqrt{n}-2$. Under the same assumption, $s_3 = s_2$ by (13). If the d's and e's are small compared to N and $(e_i - d_i')$'s are somewhat continuous then $s_3 < s_4$ follows from (13) and (21). When d's are small compared to N, (51) follows from (26), (29) and (33).

7. Applicability of the measures in an empirical study

We shall now examine the applicability of the various measures of reciprocity considered above, using survey data. For this purpose we refer to the data collected in course of a study of the relation of help and cooperation between households in 21 villages in a central region of West Bengal, India, (see Bandyopadhyay and von Eschen, 1980). We may mention that a household was chosen as the unit of study and the head of each household was asked to name the households whom he could depend upon and request for help or support in case an emergency arose or a crisis developed. The help or support could be of any kind : physical, material or financial. The point to be noted is that the main thing in this sort of transaction is to be considered as part of voluntary obligation between the giver and the receiver.

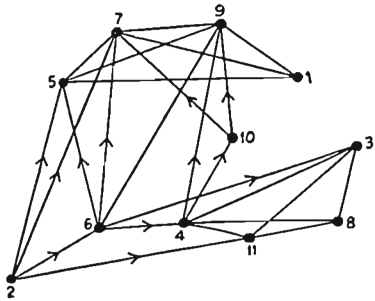


Fig. 3.

Any head of household thus referred to by a respondent could be a relative or friend or some government official, i.e., anybody whom he could really depend upon for support at the time of his need. The head of household referred to could be in the same village or elsewhere outside the village. It is noteworthy that in the entire survey altogether only eight villagers and that too scattered in three villages referred at all to someone outside the village.

Henceforth, the villages will be the populations and the households the vertices in respective populations. We provide an idea of the actual social networks thus obtained by giving them for two of the smaller villages in Figures 3 and 4, though these are not typical of all the villages.

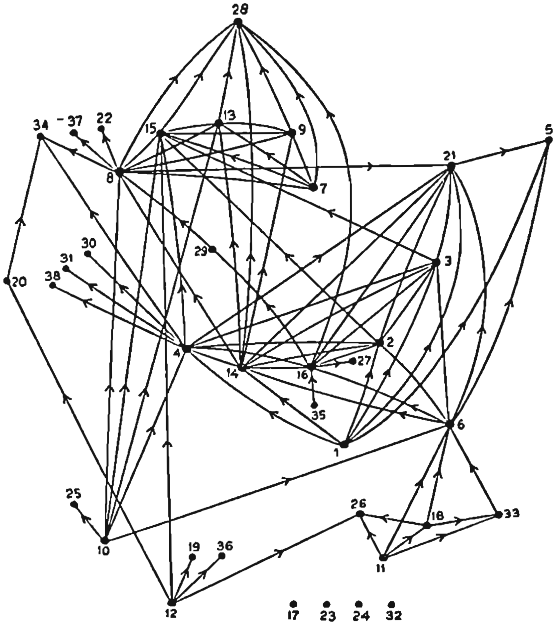


Fig. 4.

Table 2 gives the values of the measures for each of the 21 villages. The number of households N is given in column 3 for ready reference. The measure s_4^{**} is not given since its value is not known. We incidentally note that the inequalities (50) and the approximations (51) hold good for each of the villages. The value of $\hat{\rho}_1$ is not given since it is a monotone function of s_1. The values of $\hat{\rho}_2$ are given in column 11 of Table 2. We did not find the values of $\hat{\rho}_3$ and $\hat{\rho}_4$ since their computation for the larger villages takes lot of time even on a computer. We have calculated the values of $\breve{s}_3$, given in column 12 of Table 2, under the following assumptions : Since the value of r can be between 0 and 1, we assume a value somewhere in the middle of the range, say 0·5. Regarding s, the most natural value would be 1 which means that whenever i can go to j, j can go to i. However, in several villages there are some vertices with large in-degrees. If $s = 1$ then the size of the potential set for such a vertex has to be very large and, thus, not of the order of d_i/r. Hence we choose a value for s which is slightly smaller than 1, say 0·8. With the choice $r = 0·5$ and $s = 0·8$, the formulae (42) and (43) become

$$E(s_0) = \frac{m}{5}+\frac{N}{40}$$

and

$$V(s_0) = 0·35\ E(s_0).$$

TABLE 2. THE VARIOUS MEASURES OF RECIPROCITY

sl. no.	name of the village	N	s_0	s_1	$s_2 = s_3$	s_4	s_1^{**}	$s_2^{**} \doteqdot s_3^{**}$	$\bar{s}_3$	$\hat{\rho}_2$	$\breve{s}_3$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	Bamandiha	11	1	1.82	33.3	100.0	—4.0	2.4	29.8	2.5	—0.7
2	Bonachapar	11	13	23.64	68.4	92.8	—0.2	3.9	51.9	2.4	3.1
3	Kashtatari	13	6	7.09	37.5	85.7	—3.5	1.8	19.6	1.0	—0.6
4	Latapukur	38	24	3.41	45.3	68.6	—13.1	10.8	40.5	2.8	9.6
5	Sankarpur	67	0	0.00	0.0	0.0	—27.2	—0.6	—1.25	—∞	—6.1
6	K. Gopalnagar	71	50	2.01	43.5	60.2	—26.5	20.2	40.5	3.3	0.5
7	Hinglo	91	114	2.78	99.1	100.0	—32.8	63.3	98.7	12.2	15.9
8	Sukna	100	9	0.18	5.5	22.0	—40.3	1.6	2.3	0.0	—12.1
9	Raspur	101	55	1.09	30.9	51.4	—39.2	20.2	28.4	2.8	—3.7
10	Amjola	116	38	0.57	27.0	45.8	—46.1	20.7	25.4	3.1	—4.7
11	Seorakuri	130	61	0.73	28.5	58.1	—51.3	24.4	26.6	3.0	—5.0
12	Maladang	150	81	0.72	66.4	88.0	—50.3	69.9	66.0	6.3	6.6
13	Doucha	152	188	1.64	95.0	98.4	—57.8	102.0	94.9	10.0	19.5
14	K. Nimdaspur	152	100	0.87	43.9	76.9	—59.7	45.9	42.7	4.2	0.9
15	Palan	100	249	1.96	96.5	98.4	—60.0	108.8	96.2	10.0	23.1
16	Md. Bazar	166	25	0.18	19.2	45.5	—67.1	21.6	18.5	3.4	—7.0
17	Baidyanathpur	167	22	0.16	12.2	21.8	—67.5	12.9	11.0	2.5	—10.5
18	Khairakuri	218	39	0.16	21.8	33.3	—88.2	32.7	21.2	3.8	—7.3
19	Kobilpur	239	388	1.36	88.6	98.7	—92.1	149.5	88.4	8.4	26.0
20	Angargaria	267	40	0.15	23.0	44.6	—104.1	42.6	23.4	4.2	—7.1
21	Haridaspur	286	18	0.04	3.6	17.0	—116.3	4.9	2.4	1.1	—22.2

Note : The serial numbers used in this table have been followed in subsequent tables also.

The data used in the calculation of the various measures is given in Table 3 where the serial numbers of the villages are same as those in Table 2. We note that since only the frequency distribution of the out-degrees is relevant for the measures s_2, s_3 and $\bar{s}_3$ we have given the out-degree sequence in the form $0^{n_0} 1^{n_1} 2^{n_2} \ldots$ where n_k denotes the frequency of k as an out-degree. The values of $s^{(1)}_{min}$, $s^{(2)}_{min}$, $s^{(3)}_{min}$ and $s^{(4)}_{min}$ are 0 for all the villages. The (out-degree, in-degree)-pairs used in calculating s_4 are not given because it would take too much space. Note that it would not be enough to give the frequency distributions of the out-degrees and the in-degrees separately. We just mention that there are several large in-degrees in some villages (70, 106 and 161 in three villages). At the other extreme we have some villages where the in-degrees and out-degrees almost coincide.

TABLE 3. DATA USED IN THE COMPUTATION OF VARIOUS MEASURES

village serial no.	$d_1, d_2, \ldots, d_N$	$s^{(1)}_{max}$	$s^{(2)}_{max} = s^{(3)}_{max}$	$s^{(4)}_{max}$	$\mu_{(1)}$	$\sigma_{(1)}$	$\mu_{(2)} \doteqdot \mu_{(3)}$	$\sigma_{(2)} \doteqdot \sigma_{(3)}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	$0^{5}\ 1^{6}$	55	3	1	13.8	3.2	0.14	0.36
2	$2^{1}\ 3^{6}\ 4^{2}\ 5^{2}$	55	19	14	13.8	3.2	6.45	1.68
3	$1^{3}\ 2^{9}\ 3^{4}\ 4^{9}$	78	16	7	19.5	3.8	3.41	1.40
4	$0^{17}\ 1^{4}\ 3^{1}\ 4^{8}\ 5^{4}\ 6^{8}\ 7^{2}\ 8^{1}\ 9^{2}\ (10)^{1}$	703	53	35	176	11.5	4.04	1.85
5	$0^{23}\ 1^{38}\ 2^{9}\ 3^{2}$	2211	28	1	553	20.4	0.36	0.59
6	$0^{6}\ 1^{1}\ 2^{11}\ 3^{24}\ 4^{13}\ 5^{6}\ 6^{7}$	2485	115	83	621	21.6	5.35	2.21
7	$1^{18}\ 2^{29}\ 3^{36}\ 4^{13}$	4095	115	114	1024	27.7	3.24	1.75
8	$2^{6}\ 3^{69}\ 4^{30}\ 5^{1}$	4950	163	41	1238	30.5	5.30	2.28
9	$0^{1}\ 1^{1}\ 2^{9}\ 3^{41}\ 4^{37}\ 5^{11}\ 6^{1}\ 7^{1}$	5050	178	107	1263	30.8	6.26	2.41
10	$0^{15}\ 1^{20}\ 2^{26}\ 3^{28}\ 4^{20}\ 5^{7}\ 6^{2}\ 8^{1}$	6670	141	83	1668	35.4	2.97	1.69
11	$0^{1}\ 1^{12}\ 2^{32}\ 3^{30}\ 4^{33}\ 5^{18}\ 6^{7}\ (11)^{1}$	8385	214	105	2096	39.7	5.48	2.28
12	$0^{64}\ 1^{4}\ 2^{6}\ 3^{5}\ 4^{10}\ 5^{7}\ 6^{9}\ 7^{4}\ 8^{9}$	11175	122	92	2794	45.8	1.33	1.14
13	$0^{8}\ 1^{27}\ 2^{34}\ 3^{46}\ 4^{28}\ 5^{12}\ 6^{1}$	11476	198	191	2869	46.4	3.41	1.81
14	$0^{1}\ 1^{11}\ 2^{30}\ 3^{66}\ 4^{39}\ 5^{6}\ 6^{2}\ 7^{1}$	11476	228	130	2869	46.4	4.52	2.08
15	$1^{12}\ 2^{93}\ 3^{41}\ 4^{56}\ 5^{16}\ 6^{2}$	12720	258	253	3180	48.8	5.24	2.24
16	$0^{20}\ 1^{41}\ 2^{99}\ 3^{26}\ 4^{2}$	13695	130	55	3424	50.7	1.23	1.10
17	$0^{20}\ 1^{35}\ 2^{49}\ 3^{30}\ 4^{19}\ 5^{6}\ 6^{3}\ 7^{1}$	13861	181	101	3465	51.0	2.30	1.51
18	$0^{49}\ 1^{46}\ 2^{64}\ 3^{63}\ 4^{6}\ 5^{1}$	23653	170	117	5913	66.6	1.35	1.15
19	$0^{100}\ 1^{30}\ 2^{22}\ 3^{19}\ 4^{11}\ 5^{8}\ 6^{11}\ 7^{4}\ 8^{1}\ (13)^{8}\ (14)^{6}\ (17)^{1}\ (18)^{4}\ (19)^{15}$	28441	438	393	7110	73.0	6.74	2.55
20	$0^{43}\ 1^{64}\ 2^{98}\ 3^{84}\ 4^{8}\ 5^{4}\ 6^{1}$	32896	205	110	8224	78.5	1.27	1.12
21	$0^{18}\ 1^{9}\ 2^{17}\ 3^{84}\ 4^{117}\ 5^{46}\ 6^{1}\ 7^{1}$	40755	497	106	10189	87.4	6.07	2.43

From column 5 of Table 2 we observe that the values of the measure s_1 are extremely small except for two small villages. This is because all the villages except the four smallest have low expansiveness. We might say the measure s_1 is low not because ties are not reciprocated but because the number of ties is small.

The values of s_2 cover the entire range from 0 to 100. We mention that $m < \binom{N}{2}$ for all the villages, so s_2 is calculated from (7).

One may consider the measure s_3 to be more appropriate in the present context. This is because we find from column 2 of Table 3 that the out-degrees do not differ much in most villages. Thus instead of taking only m as given, it looks more reasonable to assume that $d_1, d_2, \ldots, d_N$ are given. This also has the operational justification that each household can choose the number as well as the set of households whose help it takes when in need. Thus we are led to the measure s_3.

We make a few observations about the out-degree sequences observed in the villages. Firstly we note that the average number $\bar{d}$ of choices made by a household does not vary much over the villages. To see this, it is convenient to use $\left[\frac{1}{2}\Sigma d_i\right] = s^{(3)}_{\max}$ tabulated in column 4 of Table 3. The value of $\bar{d}$ is between 2.1 and 3.7 for 15 of the 21 villages and does not seem to depend on the size of the village. The effective d_{max} (i.e., the maximum of d's taken after omitting a few large values, say the largest 5% of the d's) is 4, 5 or 6 for 15 of the villages. Finally, we note that the empirical probability distribution of d's, that is, the frequency distribution with the frequencies divided by N, does not vary much over the villages.

For all the villages except the second and fourth, the inequality $d_1 \leqslant 2\sqrt{n}-2$ holds. Even for these two villages it can be shown directly from (10) and (12) that $s^{(3)}_{\min} = 0$ and $s^{(3)}_{\max} = [\Sigma d_i/2]$, hence $s_3 = s_2$ for all the 21 villages.

We make one observation about the measure s_3. Apparently this measure assumes that all other households in the village are potential choices for the i-th household ($i = 1, \ldots, N$). However as we have already noted, d_i's are small compared to N. So even with potential sets much smaller than the entire village, from which the households make their respective choices, $s^{(3)}_{\min}$ remains 0 and $s^{(3)}_{\max}$ remains $[\Sigma d_i/2]$. Thus s_3 is quite robust with respect to changes in the potential sets unlike the measures based on probabilistic models.

The values of the measure s_4 for the 21 villages are given in column 7 of Table 2. We note that the values of s_4 also cover the range from 0 to 100. The measure s_4 seems to be fairly good (the rank correlation between s_3 and s_4 is 0·91). We have been able to show that $s^{(4)}_{\min} = 0$ for 11 of the 21 villages by using the bounds of Rao (1984). Moreover, we found using the method of interchanges that $s^{(4)}_{\min} = 0$ and $s^{(4)}_{\max} = \left[\frac{1}{2}\Sigma \min(d_i, e_i)\right]$ for all the 21 villages. Also, every integer value between the minimum and the maximum can be assumed by s_0.

We found that in some villages s_4 is highly sensitive to small changes in s_0 and, thus, unstable. This is true especially in the first and fifth villages where $s^{(4)}_{\max} = 1$. For the fifth village we have $s_4 = s_3 = 0$ since $s_0 = 0$. If s_0 were 1 then s_4 would be 100 though s_3 would only be 3·6. Since $s_0 = 1$ in the first village, we have $s_4 = 100$ for this village though s_3 is only 33·3.

From column 8 of Table 2 we see that except for the three smallest villages the value of s_1^{**} is less than -13 and decreases almost monotonically with increasing N, reaching the value $-116{\cdot}3$. The value of s_1^{*} is practically 0 for all villages except one. Similarly, s_2^{*} and s_3^{*} are practically 100 for all but four of the villages. Thus none of s_1^{**}, s_2^{**} and s_3^{**} is good in the present context.

We do not give the values of s_2^{**} and s_3^{**} separately in Table 2 since we found that they are very close. There is less than 3% difference between $\mu_{(2)}$ and $\mu_{(3)}$ and less than 2% difference between $\sigma_{(2)}$ and $\sigma_{(3)}$ for all the villages. We mention another interesting fact. Since the empirical probability distribution of d's does not vary much over the villages, $\mu_{(3)}$ and $\sigma_{(3)}$ are fairly constant over the villages and so s_3^{**} is of the form $a+bs_0$ where a, b are constants. Thus there is a high correlation between s_3^{**} and s_0.

Columns 6 and 10 of Table 2 show that the value of $\bar{s}_3$ is quite close to that of s_3 except for three small villages. This is because for all the other villages, $\bar{d}$ is small compared to N and $\bar{d}^2$ is small compared to $2s_0$. However the interpretation of the values of $\bar{s}_3$ is different from that of s_3. According to the values of $\bar{s}_3$ there is a tendency towards reciprocation in most of the villages. But we feel it is more appropriate to say that several villages (e.g. those with serial numbers 17 and 21) have, in fact, very low reciprocity. In all the villages the maximum possible value of $\bar{s}_3$ is 100 but the minimum varies slightly from village to village. The minimum is close to 0, so $\bar{s}_3$ is practically never negative.

From column 11 of Table 2, we see that the value of $\hat{\rho}_2$ is negative for only one village. According to this measure also there is tendency towards reciprocation in almost all villages. As already mentioned, we do not think this reasonable.

We finally come to the measure $\check{s}_3$. From column 12 of Table 2, we see that $\check{s}_3$ is negative for 12 of the 21 villages and positive for the remaining nine. This measure seems to be fairly good. We may also note that the correlation coefficient between $\check{s}_3$ and s_3 for the 21 villages is 0·94 while the Spearman rank correlation is 0·95. Thus $\check{s}_3$ and s_3 give more or less the same ordering of the villages according to reciprocity. Even in terms of the interpretation of the values, $\check{s}_3$ is somewhat close to s_3 in the sense that both give the same classification of a village as having low reciprocity, middle order reciprocity or high reciprocity. Thus among the the probabilistic measures considered, $\check{s}_3$ comes closest to the deterministic measure s_3. However in terms of tail probabilities, even $\check{s}_3$ can not differentiate the bottom ten villages or the top five villages.

From the above discussion, it is clear that for networks such as those of the 21 villages under consideration, if we have to choose one measure of reciprocity, it has to be one of s_3, s_4 and $\check{s}_3$. Among these, s_4 is difficult to calculate and may be unstable for some networks. The measure $\check{s}_3$ fails to differentiate the villages with low reciprocity and also those with high reciprocity (in terms of the tail probability). Incidentally, it may be noted that the probabilistic measures s_1^{**}, s_2^{**}, s_3^{**} also suffer from the above defect. These may be useful for testing the null hypothesis (viz., the model assumed) but not for measuring reciprocity. Hence we propose s_3 as the most suitable measure of reciprocity for networks such as those of the 21 villages.

8. Relevance of s_3 : a sociological illustration

We have already observed that there is a wide variation in reciprocity, as measured by s_3, among the twenty-one villages of our sample (col. 6 of Table 2). The over-all pooled value of s_3 for these twenty-one villages considered together is 43·9 which seems to be of the middle order. We find comparatively even much lower values of s_3 for a few villages and extremely high values of s_3 in a few other villages. In this section we shall not propose any strictly causal explanation of this variation in social behaviour, but suggest that this is not just a spurious phenomenon by indicating with which aspect of the society the pattern of variation in reciprocity tends to be highly associated.

In Table 4 we give the values of usually considered demographic, economic and social variables for each of the 21 villages. In the last row, we give the Spearman rank correlation (r_s) between s_3 and the variable tabulated in each column. We find that s_3 does not show any significant association with the demographic factors.

TABLE 4. SOCIO-ECONOMIC DATA OF THE VILLAGES

village sl. no.	increase(+)/decrease(−) at the time of survey as percent of what it was 20 years back			density per acre at the time of survey	percent of total hh having cultvn. as the principal occupn.	percent of total workers engaged in non-agrl. occupation		Numbers of			migrant hhs. as % of the total no. of hhs
	households (hhs)	population	density per acre			at all	mainly	castes	kin-groups	hhs. per kin-group	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	—23.1	—15.0	—19.0	0.17	3.3	06.7	0.0	1	5	2.2	27.3
2	72.7	54.9	97.2	0.71	18.6	81.4	0.0	1	3	3.7	18.2
3	0.0	—70.0	—78.7	0.27	0.0	100.0	0.0	2	5	2.6	30.8
4	60.7	63.6	65.8	0.68	0.0	100.0	1.3	1	6	6.3	18.4
5	35.1	—41.7	—20.4	0.65	34.2	64.9	0.0	8	8	8.4	59.7
6	69.4	79.8	75.9	0.51	19.6	71.3	12.9	3	24	3.0	26.8
7	10.3	43.5	42.3	1.11	33.3	57.7	10.5	11	23	4.0	7.7
8	21.5	40.5	40.0	1.05	10.5	83.0	7.4	8	11	9.1	64.0
9	9.7	51.6	53.6	0.43	38.7	61.3	2.4	10	35	2.0	23.8
10	41.6	72.5	73.6	1.58	36.4	53.1	25.3	10	30	2.9	32.8
11	73.8	—29.4	7.3	1.00	23.7	69.3	14.0	14	66	2.0	30.0
12	43.7	17.6	17.6	2.20	25.4	74.4	3.6	13	33	4.6	30.7
13	18.7	62.6	62.4	1.51	33.7	50.8	50.2	15	31	4.9	11.8
14	48.2	67.4	60.7	0.80	35.9	52.5	25.1	14	70	2.2	27.0
15	6.4	20.0	29.8	1.09	35.7	56.1	20.9	12	21	7.6	8.8
16	50.0	98.2	97.7	5.24	33.0	53.6	26.8	24	17	9.8	59.0
17	39.2	59.7	61.3	1.00	25.4	71.2	26.6	16	32	5.2	40.1
18	10.9	— 0.9	— 1.0	1.91	42.1	47.9	12.7	15	44	5.0	44.0
19	35.0	56.4	55.4	1.43	30.4	65.0	18.8	11	66	3.6	27.2
20	67.4	96.4	95.8	2.31	23.7	52.8	40.9	15	57	4.5	52.1
21	20.0	62.6	63.0	1.94	36.5	59.7	6.9	15	26	11.0	47.9
r_s	—0.10	0.01	0.05	—0.14	—0.12	0.03	0.05	—0.25	0.0	—0.35	—0.00

The impact of the economic factors can possibly depend on the proximity of the village to the centre of influence. We consider three types of influence, viz., accessibility through metalled roads and the influences of the nearest trading centre and the wholesale market or district headquarters. The relevant data are given in Table 5. We find again that none of the types of distance considered has any *consistent* impact upon reciprocity. One may argue that starting from a somewhat low value, the average value of s_3 appears to rise to a peak somewhere in the middle of the range of the distance and then falls. But we should be cautious even in accepting this pattern. In most of the distance categories the dispersion of the individual values of s_3 is quite high, so these do not show the same pattern as the averages. Thus proximity does not appear to be the main factor behind reciprocity. However, it may have some subsidiary role to play about which we shall come back later.

TABLE 5. VALUES OF s_3 BY DISTANCE CATEGORIES

distance (in miles)	no. of villages	values of s_3	average values of s_3	
			pooled	ordinary
(1)	(2)	(3)	(4)	(5)
all weather walking distance from metalled road				
0	7	5.5, 19.2, 23.9, 28.5, 43.5, 43.9, 95.0	38.4	33.7
1–2	6	0, 21.8, 27.0, 37.5, 66.4, 88.6	59.6	40.2
3–6	6	30.9, 33.3, 45.3, 68.4, 96.5, 99.1	73.0	62.2
7–10	2	3.6, 12.2	5.9	7.9
distance from the nearest trading centre				
0	6	12.2, 19.2, 21.8, 23.9, 28.5, 95.0	34.6	33.4
1– 2	7	27.0, 37.5, 43.5, 43.9, 66.4, 88.6, 96.5	69.2	57.6
3– 6	3	0, 5.5, 99.1	40.2	34.9
7–10	5	3.6, 30.9, 33.3, 45.3, 68.4	14.7	36.3
distance from wholesale market/district headquarter				
1– 2	2	21.8, 27.0	23.9	24.4
3– 6	6	19.2, 23.9, 28.5, 66.4, 88.6, 95.0	60.6	53.6
7–10	5	0, 5.5, 37.5, 96.5, 99.1	65.3	47.7
11–14	8	3.6, 12.2, 30.9, 33.3, 43.5, 43.9, 45.3, 68.4	22.2	35.1
distance from river/canal				
< 1	10	21.8, 33.3, 37.5, 45.3, 66.4, 68.4, 88.6, 95.0, 96.5, 99.1	78.7	65.2
1 <	11	0.0, 3.6, 5.5, 12.2, 19.2, 23.9, 27.0, 28.5, 30.9, 43.5, 43.9	20.5	12.9

Next we look at direct indicators of economic activities given in columns 6, 7 and 8 of Table 4. Again, we find that none of these shows any noteworthy correlation with s_3.

We finally come to the social composition of a village as measured by, for example, the number of castes, number of kingroups or the average size of a kingroup. Of these, the average size of a kingroup shows considerable correlation with s_3 (the rank correlation is $-0{\cdot}35$ and the ordinary correlation coefficient is $-0{\cdot}31$). But even here the correlation is not large enough for the variable to provide, *by itself*, a satisfactory interpretation of the values of s_3.

8.1 *The main factor affecting reciprocity.* Even though the number of castes, number of kingroups and the average size of a kingroup do not show high association with s_3, we found that another aspect of social composition, viz., the percentage of migrants, x, in the village has high (negative) correlation with s_3. The rank correlation is $-0{\cdot}90$ and the ordinary correlation coefficient is $-0{\cdot}83$. Thus a substantial part of the variation in reciprocity across the villages is accounted for by the incidence of migration, identified as the percent of those who do not consider the village as the place where they have permanently settled. Incidentally, we found that for all such households the village was not their paternal ancestral home either. We may also mention here that the partial correlation coefficient between s_3 and x eliminating the effect of any of the other variables in Table 4 does not differ much from ρ_{s_3x}. It remains greater than 0·828 in magnitude, so the correlation between s_3 and x is not spurious.

8.2 *Non-linear regression.* The values of x, the percentage of migrants, are given in column 12 of Table 4 and we present the graph of s_3 versus x along with a free hand drawing of the general trend in Diagram 1. From this diagram it is clear that though the value of s_3 tends to decrease with increasing value of x, the trend itself is not linear. Perhaps this is why $\rho^2_{s_3x}$ is only 0·69. Our aim is to show that the value of s_3 is determined by the value of the explanatory variable (migration) but not necessarily as a linear function. Thus we are interested more in correlation ratio than in correlation coefficient. However as we have data for only 21 villages, we cannot hope to evaluate accurately the correlation ratio. So we settle for fitting a polynomial of small degree in the explanatory variable to the data.

If we use a quadratic in x to predict s_3 we find the best predictor, obtained by the method of least squares, to be

$$\hat{s}_3 = 126{\cdot}04 - 3{\cdot}817x + 0{\cdot}03105x^2.$$

The predicted values are given in column 2 of Table 6. The square of the correlation coefficient between s_3 and the best predictor is $\rho^2_{s_3.x.x^2} = 0{\cdot}77$ which is considerably larger than $\rho^2_{s_3.x}$. Thus the second degree term seems to be important (it is significant at 5% level). Also, a very substantial part of the variations in s_3 can be explained by using a quadratic in x. However, it may still be worthwhile to try to explain any significant deviation of the individual villages from the overall trend in Diagram 1. Because, we shall

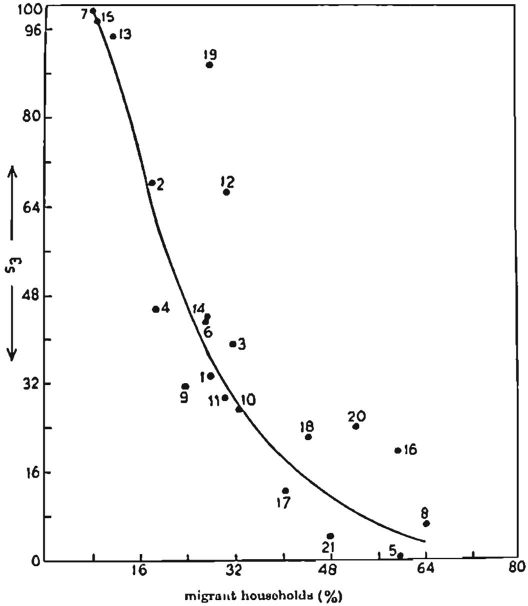


Diagram 1

then know if we need any further change in our hypothesis concerning the variations in s_3. We note that the villages with serial numbers 12, 16, 19 and 20 show particularly high reciprocity in relation to the percentage of migrant households while the villages with serial numbers 4, 9 and 21 show somewhat low reciprocity.

TABLE 6. MODIFIED AND ESTIMATED VALUES OF s_3

sl. no.	$\hat{s}_3$	s_3^0	s_3	$\hat{\hat{s}}_3$
(1)	(2)	(3)	(4)	(5)
1	45.0	33.3	33.3	51.8
2	66.9	34.2	68.4	76.7
3	38.0	37.5	37.5	27.9
4	66.3	34.0	45.3	62.0
5	9.0	0.0	0.0	5.9
6	46.1	23.4	43.5	45.5
7	98.5	99.1	99.1	93.0
8	9.1	3.1	5.5	11.7
9	52.8	22.6	30.9	50.1
10	34.3	27.0	27.0	20.5
11	39.5	15.8	28.5	28.9
12	38.2	27.4	24.5	66.8
13	85.3	79.2	95.0	86.4
14	45.7	26.6	43.9	47.9
15	94.0	96.5	96.5	89.2
16	9.1	5.9	2.7	19.9
17	23.0	12.2	12.2	16.4
18	18.3	4.1	8.9	9.2
19	45.2	23.2	39.5	64.6
20	11.6	6.4	3.4	9.6
21	14.6	2.3	2.9	4.9

We may consider the deviations from the trend in one of the following ways. The specific features of the villages may be thought as responsible for these deviations. Or, we may modify the measure s_3 in some way. Alternatively we may modify the explanatory variable, namely, migration, in some way. Lastly we may also use a multi-causal model involving migration along with another variable.

8.3 *Village specificities.* The first approach leads to village specificities. We may observe special circumstances in one or two villages. For example, in the village with serial number 19, we noticed the presence of three large cliques which led to strong intra-group solidarity and high reciprocity. How-over, village specificities should be used as a last resort after exhausting the possibility of explaining the deviations in a systematic manner through the other approaches. Besides, special features of a village can at best explain the direction of its deviation, positive or negative, but not the magnitude. So we consider the other approaches below.

8.4 *Modifying the dependent variable.* It is reasonable to suppose that a household normally goes to the available kins in the village for help and support. In fact, we found that 71% of all the requests for help and support made by the households in the 21 villages taken together are restricted to their kins.

We will now consider modifying s_3 using the idea of kinship. For this purpose, we tabulate the number N_{nk} of households with no kin in the village, in column 2 of Table 7. We also give the values of $N_{mg,nk}$, the number of

TABLE 7. DATA USED IN MODIFYING s_3 AND z

sl. no.	N_{nk}	$N_{mg,nk}$	N_{ts}	$N_{mg,ts}$	z	u
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	3	0	3	0	37.5	37.5
2	3	1	0	0	18.2	12.5
3	1	1	0	0	30.8	25.0
4	4	1	4	0	20.6	17.6
5	3	1	8	0	67.8	60.9
6	13	6	1	1	25.7	22.4
7	1	0	0	0	7.7	7.8
8	7	3	0	0	64.0	65.6
9	15	4	1	0	24.0	23.3
10	4	0	5	0	34.2	33.9
11	9	4	1	1	29.5	28.9
12	81	27	82	29	25.0	27.5
13	19	3	7	0	12.4	11.3
14	33	13	0	0	27.0	23.5
15	3	0	0	0	8.8	8.0
16	36	25	21	18	55.2	50.2
17	19	4	6	0	41.6	42.6
18	16	13	32	19	41.4	41.1
19	101	48	89	36	19.3	12.3
20	30	22	28	24	48.0	49.3
21	14	5	15	3	49.4	48.5

households who are both migrants and without kins, in column 3. From this table we find that the villages showing particularly high reciprocity (relative to x) in Diagram 1, namely those with serial numbers 12, 16, 18, 19 and 20, have a large number of households without kins. More precisely, they are the villages with large values of $N_{mg,nk}$. Thus the presence of a large number of households who are both migrants and have no kins seems to somehow inflate the value of s_3 relative to the percentage of migrants. One plausible explanation for this might be that while migrants tend to develop more one-way ties, migrants who have no kins in the village may develop symmetric ties out of necessity. Another could be that the presence of a large number of migrants without kins induces more solidarity among the rest of the population. We are, of course, not sure of the veracity of either.

Consequently, if the value of s_3 is deflated by using a function of $N_{mg,nk}$ we may get a better fit. To do this, we note that s_3 is a ratio and so it is perhaps reasonable to reduce it by the proportion of households who are both migrants and without kins. We get the best fit by taking the base to be $\min(N_{mg}, N_{nk})$ which is the maximum possible value for $N_{mg,nk}$. We thus consider

$$s_3^0 = s_3\left(1-\frac{N_{mg,nk}}{\min(N_{mg}, N_{nk})}\right).$$

The values of s_3^0 are given in column 3 of Table 6. In this table, we have taken s_3^0 to be s_3 when $N_{mg} \leqslant 1$ or $N_{nk} \leqslant 1$. In Diagram 2 we draw the graph of s_3^0 versus x. We find that all but two of the points lie close to the trend shown by a free hand curve. Though the value of $\rho^2_{s_3^0 . x}$ is only 0·69, that of $\rho^2_{s_3^0 . x, x^2}$ is 0·90. Thus the percentage of migrants together with the proportion of households who are both migrants and without kins almost determines the value of s_3.

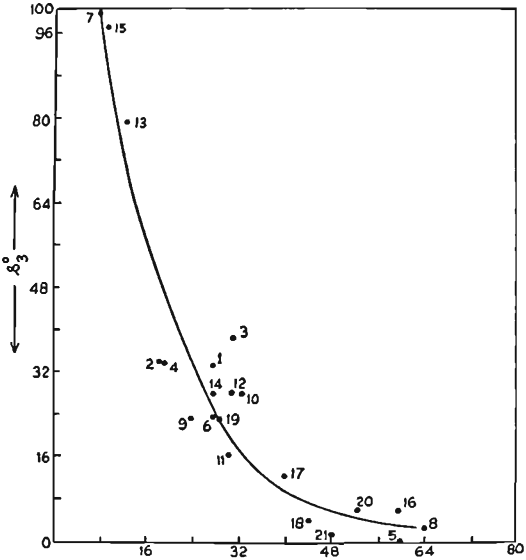
If we use the proportion $N_{mg,nk}/N_{nk}$ to deflate s_3, then also we get a fairly good fit. Other proportions like $N_{mg,nk}/N_{mg}$ and N_{nk}/N also give a better fit than in Diagram 1 but not one as good as the above two.

Another phenomenon, which is closely related to the lack of kins in the village is isolatedness. We label a household an isolate if it neither makes any request to others for help and support nor receives any such request.

If we use isolatedness instead of lack of kins to deflate the value of s_3 we may form the following measure

$$s_3' = s_3\left(1-\frac{N_{mg,is}}{\min(N_{mg}, N_{is})}\right)$$

where N_{is} is the number of isolates and $N_{mg,is}$ is the number of isolated migrants. The values of N_{is}, $N_{mg,is}$ and s_3' are given in columns 4 and 5 of Table 7 and column 4 of Table 6. We note that the values of s_3' are somewhat close to those of s_3^0. However, the graph of s_3' versus x given in Diagram 3 shows a better, and, near-perfect fit except for the villages with serial number 4 and 9. The value of $\rho^2_{s_3',x}$ is 0·81 and that of $\rho^2_{s_3',x\cdot x^2}$ is 0·96.



migrant households (%)

Diagram 2

Thus we can reconcile most of the deviations in Diagram 1 by modifying s_2 to s_3', s_3^0 or to one of the others mentioned above.

However, one may also note that in each of the above cases we find some villages which are away from the general trend. Besides, s_3' and s_3^0 are highly unstable when N_{is} and N_{nk} are small. Moreover, these measures are derived heuristically. Whatever good fit we have observed can, therefore, be data-specific. These measures as such may not, therefore, be sufficient to infer about the pattern of relationships. One, rather, needs to be extremely cautious in this regard. So we look at the other approaches below, which, we hope, will not be so much data-specific and hence have wider applicability.

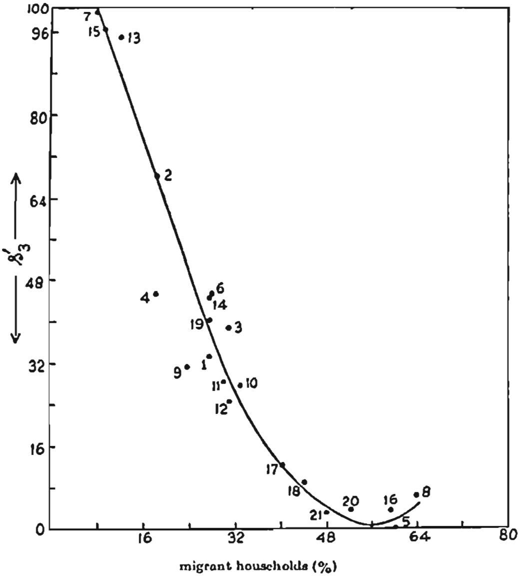
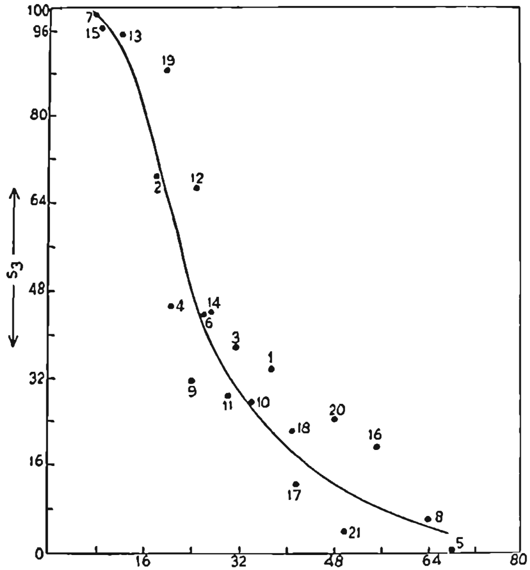


Diagram 3

8.5 *Modifying the explanatory variable.* We notice that the way we have conceptualised the measure s_3 of reciprocity, the isolates are not taken into account while measuring reciprocity. So we can think of modifying the percentage of migrants also by disregarding the isolates. We may thus consider the new explanatory variable

$$z = 100\left(\frac{N_{mg} - N_{mg,is}}{N - N_{is}}\right)$$

The values of z are tabulated in column 6 of Table 7. In Diagram 4 we plot the values of s_3 against z. We see that the fit is better than in Diagram 1 and $\rho_{s_3, z} = -0{\cdot}88$. However, many villages show considerable deviation from the trend.



migrant households excluding the isolates

Diagram 4

Since the incidence of isolatedness which is used to modify migration has to be measured from the observed network itself, we may try lack of kins instead of isolatedness. The corresponding modification of x is

$$u = 100 \left(\frac{N_{mg} - N_{mg,nk}}{N - N_{nk}} \right).$$

The values of u are tabulated in column 7 of Table 7. Modifying x to u also improves the fit but, again, not enough.

8.6 *Multifactor approach.* We now try a multifactor approach instead of modifying s_3 or x. We shall first use a linear function of two explanatory variables to predict the values of s_3. Let i denote the variable in the i-th column of Table 6 so that 12 corresponds to x. If none of i and j is 12 then $\rho^2_{s_3.i,j}$ is less than 0·69 which is the value of $\rho^2_{s_3,x}$. Even if $i = 12$ we find the maximum value of $\rho^2_{s_3.i,j}$ to be 0·75 attained when $j = 5$. This is true even if we include the percentage w of households without kins in the village as one of the variables. But we have already seen that $\rho^2_{s_3.x,x^2} = 0{\cdot}77$. Thus the single variable migration, when properly used, explains the variations in s_3 better than any linear function of any two of the variables considered. Since, however, there are considerable deviations from the quadratic trend in Diagram 1, we next see if we can get a better fit by combining a quadratic in x with a linear function of another variable. The best function of the type $a+bx+cx^2+dw$ obtained by the method of least squares is

$$\hat{s}_3 = 126{\cdot}71-4{\cdot}840x+0{\cdot}0462x^2+0{\cdot}838w.$$

The value of $\rho^2_{s_3,\hat{s}_3}$ is 0·88. We incidentally note that each of the terms containing x^2 and w is significant at 1% level. The values of $\hat{s}_3$ are given in column 5 of Table 6. Though the values of $\hat{s}_3$ are generally close to those of s_3, still, several villages (those with serial numbers 1, 4, 9, 18, 19 and 20) show considerable difference. We can obtain a better fit, but not one as good as in Diagram 3, by using $N_{mg,nk}/N_{nk}$ or $N_{mg,nk}/\min(N_{mg}, N_{nk})$ instead of w. However, as these ratios may be unstable, we omit the details.

8.7 *A longitudinal explanation for "deviant" villages.* We notice that in Diagram 3 as well as in various types of our attempts to explain the deviations in it, a few villages (specially those with serial numbers 19 and 20) consistently lie above the average curve. Similarly there are some (specially those with serial numbers 4 and 9) which consistently lie below.

In this context, we may note some additional features of these "deviant villages", so to say. The qualitative information we have gathered from various sources regarding soil type, crops grown and their yields as well as how these villages were established and became what they are now, indicate the following. Such villages like those with serial nos. 19 and 20 are situated close to the river Mayurakshi or its channels. Also the soil here is rich in quality, being mainly alluvial and thus, better suited for agricultural purposes. Moreover, they are near to metalled roads and markets. On the other hand, villages like those with serial nos. 4 and 9 lack all such facilities. These features have led to keen competition for land in the former villages, particularly between two castes, viz., the Sadgopes and the Brahmins. We think that these details may be relevant to understand what we are empirically observing. A situation like this may have induced the landed castes to be more cohesive internally. High reciprocity could have been a manifestation of this cohesion.

We hope that we have been able to show what we had intended to do in this part of our paper. Namely: obtaining the values of s_3 for different populations (i.e., villages in the present case) is not a futile mathematical exercise. Rather, these values indicate a pattern which subsequently can lead us to important processes within the society. Thus, s_3 is relevant for sociological analysis.

9. Conclusions

We have classified the various measures of reciprocity as deterministic and probabilistic. We have shown that the measures based on the usual probabilistic models are not appropriate in some empirical studies. They assume that the choices are made at random from the entire population, say, a village. Obviously, this type of assumption ignores demographic, economic and social barriers. In reality, however, such barriers exist and they restrict the boundary within which one actually makes his choices. We have argued that if one wants to work with a probabilistic model, perhaps, one should think in terms of such sub-populations, or, "potential sets", so to say, out of which the choices are made by each one. One can develop even more complicated models.

The deterministic measures, however, do not suffer from the above limitation. Among the various deterministic measures of reciprocity considered for the social network of a population, we find s_3 (equation 13) and s_4 (equation 21) to be the two best measures. Since s_4 may be highly sensitive and hence unstable as well as difficult to calculate, s_3 may be preferred.

We have also indicated the relevance of s_3 using sociological survey data. Using s_3 as the measure of reciprocity we have found that out of the 21 villages considered, not many show high reciprocity. In fact, reciprocity appears to cover the entire spectrum of possible values 0 to 100. Thus, the image of a rural community being a normatively reciprocative social entity does not appear to be tenable.

Though there is a considerable variation in reciprocity across the villages as indicated by the values of s_3, it is not a random variation and can be explained. Quite interestingly, the demographic characteristics of villages, occupational composition or admixtures of castes or kin-groups do not seem to have much impact upon reciprocity in these villages. It is also noteworthy that agriculture *per se* is not necessarily more conducive to reciprocity, nor for that matter, is urbanisation to anti-reciprocity.

We have found that migration, by itself, accounts for most of the variation in reciprocity in these villages. The greater the proportion of migrants in a village the less is the reciprocity. We find lack of kins as the next contributing factor, though of much less importance. It may also be added that proximity to the road, market and river, differences in natural fertility of land for cultivation as well as competition for ownership of good quality land among caste groups may have played some role in one way or another in some of the villages. On the whole, what seems to be crucial is the division of the villagers into the two broad groups of permanent settlers and the others as well as into those who do not have any kin in the village and those who have.

Thus, using the measure s_3 we have shown that a strong regularity underlies the pattern of reciprocity in the social networks of rural communities. We realise, however, that as our conclusions are based upon a rather small number of them, one may perhaps need to conduct a larger survey to make the conclusions more reliable. One may also take into consideration the strength of relationships between the households, perhaps by counting the number of times the head of one approached the other in a given period of time. The measure s_3 can be used, with a slight modification, to measure reciprocity in such a weighted network. One may, in addition, study reciprocity in networks of other meaningful relations besides the one we have considered to get an overall picture of the society.

Appendix

As s_3 and s_4 seem to be two of the best measures in many contexts, we will now discuss some examples to bring out the basic differences between the two. Consider the social network of a small population of, say, seven vertices with out-degrees sequence

$$(3, 2, 2, 2, 1, 1, 0). \qquad \dots \quad (52)$$

Since $\Sigma d_i = 11$, it follows that s_0 cannot exceed 5. In Figure 5 we give one social network for each value of s_0 from 0 to 5, all with the out-degree sequence (52).

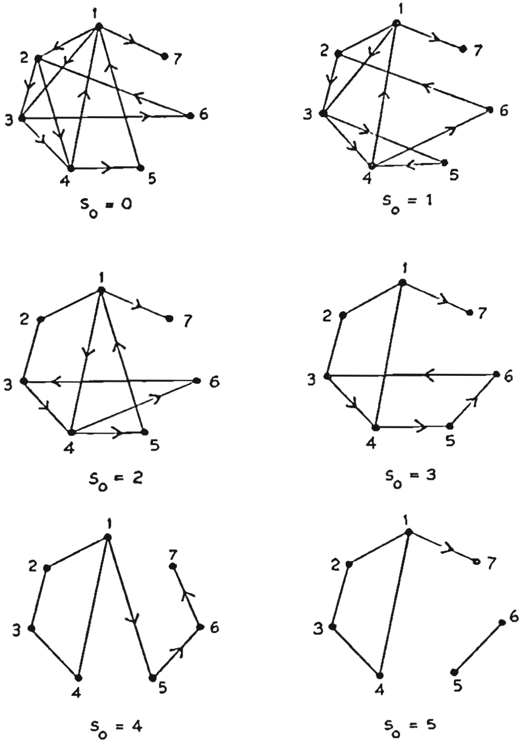


Fig. 5.

Suppose now that the in-degree sequence is given to be

$$(2, 2, 2, 2, 1, 1, 1) \quad \ldots \quad (53)$$

where the same labelling of the vertices is used for both the out-degree sequence and the in-degree sequence. Then the range of s_0 does not change since the in-degree sequence of each of the six social networks in Figure 5 is in fact (53). In general, when the out-degrees are small compared to N and the in-degree of each vertex is close to its out-degree, the range of s_0 given both d's and e's is not much different from the range of s_0 given d's alone.

Suppose next that the in-degree sequence of the population of seven vertices with out-degree sequence (52) is given to be

$$(0, 0, 0, 2, 2, 3, 4). \qquad \ldots \quad (54)$$

Now, the number of symmetric ties at the i-th vertex cannot exceed the minimum of d_i and e_i. Hence we have

$$s^{(4)}_{\max} \leqslant \frac{1}{2} \Sigma \min(d_i, e_i) = \frac{1}{2} (0+0+0+2+1+1+0) = 2$$

Further, if $s_0 = 2$ then the fourth vertex must be symmetrically tied up with the fifth and sixth vertices, hence the arcs $4 \rightarrow 7$, $5 \rightarrow 7$, and $6 \rightarrow 7$ cannot be present and the seventh vertex cannot have in-degree 4, a contradiction. Thus $s^{(4)}_{\max} \leqslant 1$. The two social networks given in Figure 2 show that s_0 can take the values 0 and 1. Thus when the in-degree sequence is given to be (54), the range of s_0 shrinks from $\{0, 1, 2, 3, 4, 5\}$ to $\{0, 1\}$. It is also possible that s_0 may get completely fixed either at 0 or at some other value. For example, if the in-degree sequence is given to be

$$(0, 0, 0, 0, 2, 4, 5) \qquad \ldots \quad (55)$$

it can be shown that s_0 can take only the value 0.

The above examples show that the sensitivity of s_4 can be high. We will now demonstrate that the sensitivity of s_4 can change much more than that of s_3 if the parameters are changed a little. Suppose that the out-degree sequence of the population of seven vertices is altered from (52) to

$$(4, 3, 3, 1, 0, 0, 0) \qquad \ldots \quad (56)$$

and the in-degree sequence is altered from (53) to

$$(1, 1, 2, 2, 2, 1, 1). \qquad \ldots \quad (57)$$

Then it can be seen easily that the range of s_0 shrinks from $\{0, 1, 2, 3, 4, 5\}$ to $\{0, 1, 2\}$. We note that in going from (52) to (56) each d_i is altered by at most 1 and in going from (53) to (57) each e_i is altered by atmost 1 and Σd_i is altered by only 1. If only the out-degrees are considered, such an alteration cannot change $s^{(3)}_{\max}$ by more than 1 and $s^{(3)}_{\min}$ remains 0.

Acknowledgements

The authors are grateful to Professors B. P. Adhikari and J. K. Ghosh of Indian Statistical Institute and Donald von Eschen of McGill University for sustained interest in the present work and their suggestions. They are also indebted to Professors Bikas K. Sinha and Y. R. Sarma for many valuable comments. Finally, thanks are due to Mr. Bhola Nath Ghosh for technical assistance.

References

Achuthan, N., Rao, S. B. and Ramachandra Rao A. (1984): The number of symmetric edges in a digraph with prescribed out-degrees. In Vijayan, K. S. and Singhi, N. M. (Eds.): *Combinatorics and Applications Proceedings of the Seminar* held in honour of Professor S. S. Shrikhande at Indian Statistical Institute, Calcutta.

Bandyopadhyay, Suraj and Donald Von Eschen (1980): *The Conditions of Rural Progress in India.* Submitted as a report to the Canadian International Development Agency.

Berge, C. (1973): *Graphs and Hypergraphs*, Amsterdam: North-Holland.

Burt, Ronald S. (1978): Cohesion versus structural equivalence as a basis for network subgroups. *Sociological Methods and Research*, 7, 189-212.

Chen, W. K. (1976): *Applied Graph Theory*, Amsterdam. North Holland.

Davis, J. A. (1977): Sociometric triads as multi-variate systems. *J. Math. Sociology*, 5, 41-69.

Freeman, Linton C. (1978): Centrality in social networks conceptual clarification. *Social Networks*, 1, 215-239.

Harary, F., Norman, R. Z. and Cartwright, D. (1965): *Structural Models: An Introduction to the Theory of Directed Graphs*, New York: Wiley.

Holland, P. W. and Leinhardt S. (1979a): *Perspectives on Social Network research*, Academic Press, New York.

——— (1979b): Structural sociometry in Holland, P. W. and S. Leinhardt: 1979a, 63-83.

Holland, P. W. and Leinhardt, S. (1981): An exponential family of probability distributions for directed graphs. *J. Amer. Stat. Asso.*, 76, 33-60.

Karve, I. and Damle, Y. B. (1963): *Group Relations in Village Community*, Deccan College, Poona.

Katz, L. and Powell, J. H. (1954): The number of locally restricted graphs. *Proceedings of the American Statistical Association*, 5, 621-626.

——— (1956): Measurement of the tendency toward reciprocation of choice in Moreno, J. L. (ed.) *Sociometry and the Science of Man*, New York, Beacon House, 403-409.

Katz, L., Tagiuri R. and Wilson, T. R. (1958): A note on estimating the statistical significance of mutuality. *The Journal of General Psychology*, 58, 97-103.

Katz, L. and Wilson, T. R. (1956): The variance of the number of mutual choices in sociometry. *Psychometrika*, 21, 299-304.

Leinhardt, S. (1977): *Social Networks: A Developing Paradigm*, Academic Press, New York.

Rao, A. Ramachandra (1984): Bounds for the number of symmetric edges in a digraph with specified out-degrees and in-degrees. Submitted for publication.

Sailer, Lee D. (1978): Structural equivalence: meaning and definition, computation and application. *Social Networks*, 1, 73-90.

Paper received: May, 1986.
Revised: October, 1986.