[Lecture Notes for the Workshop on Social Networks jointly organized by IMSc, Chennai and IIT, Madras during 20-24 February, 2012]

On Social Networks: formation, data and few analytic techniques

by

Rabindranath Jana
Sociological Research Unit
Indian Statistical Institute, Kolkata
rabindranathjana65@gmail.com
rjana@isical.ac.in

[This lecture is dedicated to the unforgettable memory of both late Prof. Suraj Bandyopadhyay and late Prof. A. R. Rao]

[Lecture Notes for the Workshop on Social Networks jointly organized by IMSc, Chennai and IIT, Madras during 20-24 February, 2012]

On Social Networks: formation, data and few analytic techniques

by
Rabindranath Jana,
Sociological Research Unit,
Indian Statistical Institute, Kolkata

The concept of social network and methodology of social network analysis (SNA) have created much interest in the minds of the researchers not only in social and behavioral sciences but also in statistical and mathematical sciences in recent decades. It has become an ideal meeting ground of all these disciplines.

What is network ?

A network is generally defined as specific type of relation linking a set of persons, objects or events. Relations are the building blocks of network analysis. Different types of relations identify different networks, even when imposed on the same set of objects. e.g., social networks, transport network, economic network, electrical network, and so on. Here we are concerned with social network only.

What is social network (SN) ?

A social network is defined as an articulation of ties of a social relationship among social units, like persons, households, organizations, etc. which are called *actors* and the *ties* are links between individuals in the network. e.g., kinship network, marriage network, friendship network, etc.

How social network is built up ?

In a society,

involuntary ties of relationship: relationship by kinship, caste and religious or community affiliation, etc.; and

voluntary ties of relationship: ties of friendship or individuals spending leisure time together, making social visits, attending invitations on the occasion of a family ceremony or household festival, extending help and support to a household at the time of an urgent need and so on.

In reality, social units interact with one another in different fields of activities in many ways. In the process, they build up networks of ties of regular interaction which are also discerned as social networks. Thus, generally, the units and the ties constituting a social network can be different according to the relationship under consideration.

What's different about social network data ?

In "conventional" sociological data,

(i) a rectangular array of measurements
(ii) rows: cases or, objects or, subjects or, observations or, households which are generally unir of study
(iii) columns: scores (quantitative or qualitative) on attributes or variables or measures
(iv) each cell: score of some actor on some attribute
(v) focus: on actors and attributes

Hh. srl. no.	Head's name	Age (in yrs.)	Sex	Caste/ Comm.	Out-degree	In-degree
1	Paresh	32	Male	SC		
2	Sita	27	Female	ST		
3	Rahim	29	Male	Muslim		
4	Minati	55	Female	General		
5	Sk. Jalal	65	Male	Muslim		

In "social network" data,

(i) a square array of measurements
(ii) rows: cases or, objects or, subjects or, observations or, households and so on.
(iii) columns: same set of cases or, objects or, subjects or, observations or, households and so on.

(iv) each cell: relationship between the concerned pair of actors and this is generally unit of study

(v) focus: on actors and relations

	Paresh	Sita	Rahim	Minati	Sk. Jalal
Paresh	x				
Sita		x			
Rahim			x		
Minati				x	
Sk. Jalal					x

How social network data are collected ?

Before going to discuss about collection of SN data, we have to be familiar with the following terms:

- **Structural and composite variables:** Structural variables are measured on ties of specific kind between pairs of actors and are the cornerstone of social network data (e.g., help relation between households, friendships between people, trade relations between nations, and so on), whereas composition variables are the actors' attributes as measured in the standard social and behavioural sciences (e.g., gender, caste/community, religion, principal occupation, education, etc.)
- **Mode:** By mode we refer to a distinct set of entities on which structural variables are measured. One-mode, two-mode, and so on.
  → *One-mode*: when actors in the network from one set and most common type network
  → *Two-mode*: two sets of actors (e.g., flow of donation from corporations to non-profit organization---Galaskiewicz and Wasserman study), one set of actors as first mode and one set of events as second mode or *affiliation network* (e.g., attendance of a set of women to a variety of social functions)
- **Network as a whole or full network:** Distribution of the ties on a specific kind among all the actors considered under study.
- **Ego-centric network/personal network/local network:** When full network is not possible:
  → *Ego-centric network (with alter connections)*: focal actor (ego) with other actors

(alters) to which ego is connected and connections among the alters.
* Can get idea about overall network density, prevailence of reciprocal ties, cliques and the like, but not about distance, reachability/connectedness, hierarchy etc.
→ *Ego-centric network (ego only)*: focal actor (ego) with other actors (alters) to which ego is connected.
* can understand something about the differences in the actors' locations in social structure and make some predictions about how these locations constrain their behavior.

Here I like to restrict myself to one-mode full networks studied at ISI. For collection of such network data, the attention on the following items is necessary:

- **Relation:** Interaction for seeking help, friendship, liking respect, transactions/transfer (of material or non-material resources), kinship, and so on.
- **Unit of observation (or, Actor):** Households, People, organizations, subgroups, community, nations/states, etc.
- **Boundary specification of the actors:** Village/Mouza, metropolitan city, class-room, etc.
- **Respondent:** Head of a household, actor itself, Chief Executive of a organization, etc.
- **Reference time period:** Neither too low nor too high
- **Survey design:**
  → Collection technique: Questionnaires, Interviews, Participatory-observations.
  → Survey technique: complete enumeration, sample survey, snow ball sampling.
- **Modelling unit:** Actor, dyad, triad, subgroup, full network.

In case of our study at ISI,

Relation: Interaction for seeking help.
Unit of observation (or, Actor): Households
Boundary specification of the actors: Village/Mouza
Respondent: Head of a household
Reference time period: within last 5 yrs.
Survey design: Interview with open-ended check-list questions and complete enumeration.
Modelling unit: Full network.

Different steps of data collection

Step-1 (listing phase): It is not necessarily Hh. to Hh. survey.
Undertake a suitable way to make a complete list of all households with few compositions variables (Head's name, Head's father's name, Para name, Caste/community, etc.)

Step -2 (main survey): It is strictly Hh. to Hh. survey.
Composition variables as required by the researcher are considered (like, Head's sex, Head's education, Head's education, family main source of livelihood, family size, Cultivable land (own. Leased in, leased out), etc.

Structural variables are as given below:

For getting help (financial, material, physical),

→ to whom you approached within last year and help received ? Is he inside or outside village ? If inside village, in which para ? What relation to him ? What helps and how many times for each help ? If help is financial/material, what amount/quantity ? What purposes served by each help ?

Methods and models studied at ISI for analyzing SN data

Relational quantification:

→ Directional and dichotomous (DD)
→ Directional and valued (DV)
→ Non-directional and dichotomous (ND)
→ Non-directional and valued (NV)

At ISI, the first two have been studied.

What is social network analysis (SNA) ?

SNA means an analysis of various characteristics of the pattern of relational ties and drawing inferences about the network as a whole or about those belonging to it considered individually or in groups.

Use of SNA techniques :

SNA techniques are used to study

(i) the pattern of social relationship;

(ii) demographic process like migration from one region to another;

(iii) economic process (say, postal money order or trade exchange between regions; volume of flow of goods between countries, flow of traffic between different places, etc).

Graphical representation of SN and its use

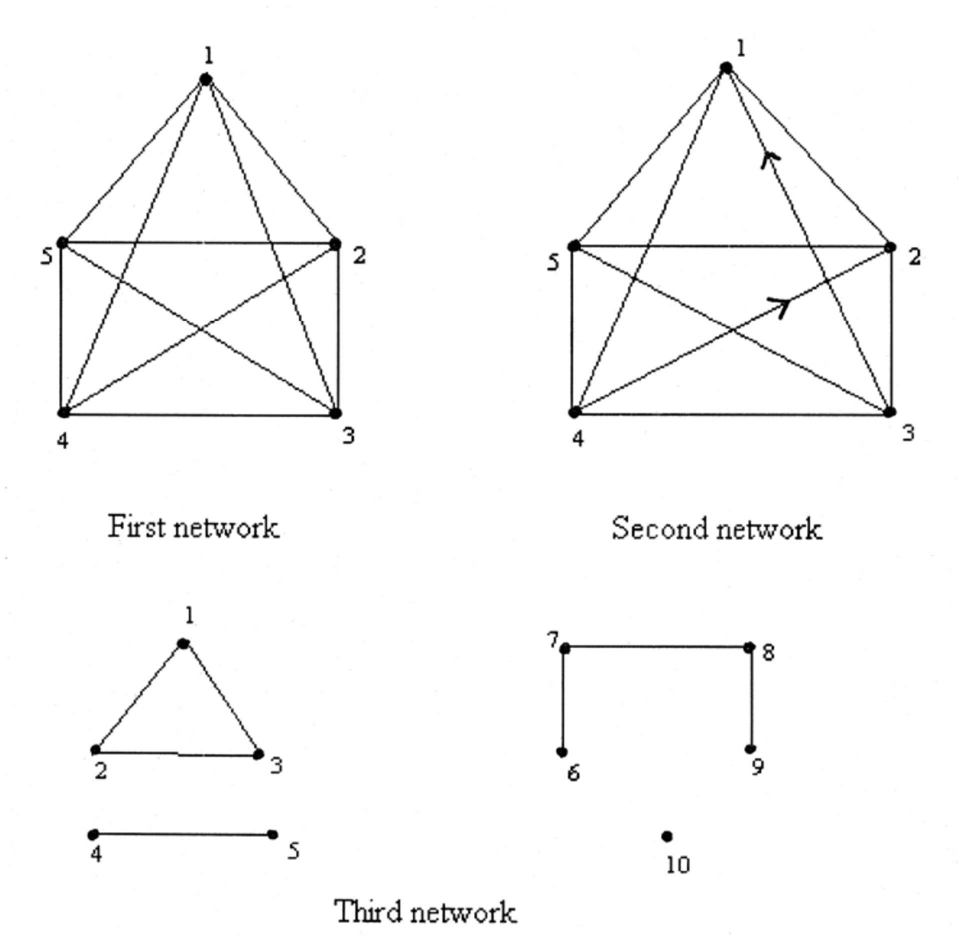
For analytical purposes, a social network is considered as a *directed graph* (in short, *digraph*) if direction of a tie is considered and as a *graph* if not. For example: 'x goes to visit y' gives a digraph, whereas 'x and y are kins' gives an undirected graph (simply, graph).

It is represented by a line with arrowhead indicating the direction of relationship. A tie with direction is called an *arc* and a tie without direction is called an *edge*.

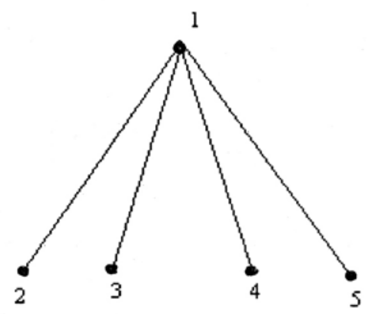
In addition to whether there is a tie or not between a pair of nodes, one could also observe the value or volume or frequency of flow of transaction among them and use it as the weight of the tie and thus obtain a network which would then be a *weighted digraph* or *multidigraph* if the direction of relation is considered and *weighted graph* or *multigraph* if the direction of relation is not considered.

The following hypothetical social networks (DD) are considered for understanding how the pattern of relationships can vary in different ways (here by a line between two nodes without arrow head we mean they are reciprocally tied up; and by an arc between two nodes with arrow head we mean that only one goes to the other to whom arrow head is marked.):

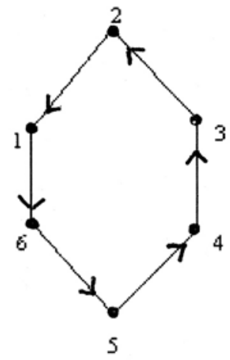
Figure 1



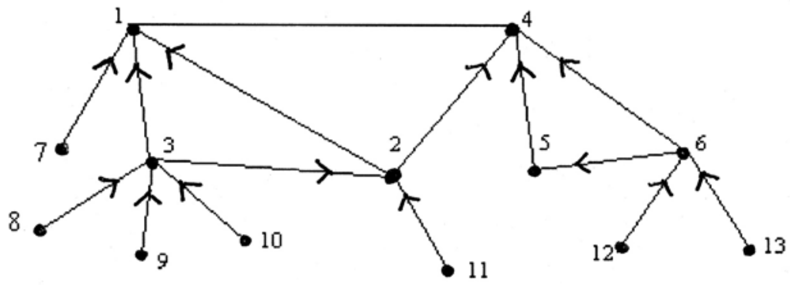
(i) The first network shows an ideal situation where everybody goes to everybody else directly; this is an ideal situation.
(ii) The second network is close to the ideal situation.
(iii) In the third, the ties are all reciprocated but the network is highly fragmented.



Fourth network



Fifth network



Sixth network

(iv) The fourth network is connected but highly centralized and shows concentration of power, being held together by a single node whose disappearance will cause the disintegration of the network.
(v) The fifth network is also connected in the sense that everybody can go to everybody else directly or indirectly, possibly through a large number of intermediaries.
(vi) The sixth one displays a strong hierarchy and though it is connected, there is a high tendency of flow only in one direction.

What does SNA differs from "standard" social science approaches ?

The following dimensions are considered in SNA, but not considered in standard approaches:

(i) concepts are relational
(ii) data are based relationship among the social units

(iii) relational properties (e.g., structures, reciprocity, etc.) are studied

Let me take an example for understanding it:
Consider a study seeking corporate behaviour in a large metropolitan area in terms of the level and types of monetary support given to local non-profit and charitable organizations (Galaskiewicz 1985). The standard approaches would:

- define a population of relevant units (here corporations)
- take a random sample of them if needed
- measure a variety of characteristics (such as size, industry, profitability, level of support for non-profit organizations, and so on)

The key assumption in the standard approaches is that the behaviour of a specific unit does not influence any other units. But this may not be true in reality. Because, the corporations tend to look at the behaviors of each other and even attempt to mimic each other. The network analysts take exception to this assumption. In order to get more complete picture of such behavior, it needs to study the relationships among the corporations, such as

- membership on each others' boards of directors
- acquaintanceships of corporate officers
- joint business dealings
- other relational variables

Some relevant terms and parameters in SN

Terms:

- Nodes (or, actors)/vertices
- Ties (or, links)/ arcs/edges
- Reciprocal (or, mutual/symmetric/two-way) relation
- Asymmetric (or, one-way) relation
- Null (or, no) relation

Parameters

- Out-degree
- In-degree
- Isolated vertex
- Reciprocity
- Distance
- Reachabilty
- Strong component
- Weak component
- Source
- Sink ; and so on.

Matrix representation of SN and its use

A simple social network of N actors can be represented by a N x N square 0-1 matrix X = ((X_{ij})) with 'structurally zero' diagonals, where

X_{ij} = 1 if i has relation to j

= 0 if not.

Transpose of X: the degree of similarity between X and trans(X) indicates the degree of symmetry, i.e., the correlation between X and trans(X) is a measure of the degree of reciprocity of ties.

Suppose, X for money transaction (MT) and Y for goods transaction (GT). Then

Addition of X and Y: Z = X + Y

Z_{ij} = 0 → no relationship between i & j

= 1 → relationship either in MT or in GT

= 2 → relationship in both MT and GT

Subtraction of X & Y: Z = X - Y

Z_{ij} = -1 → only GT relationship between i & j

= 0 → either no relationship both in MT and GT <u>or</u> relationship both in MT and GT

= +1 → only MT relationship between i & j

Multiplication of X by itself:

X.Y → no interpretation

But X.X = X^2 → each cell value indicates the no. of walks of length 2;

…….

Reachability matrix:

$R^T = X + X^2 + X^3 + \ldots + X^p$

Where

R^T = ((r_{ij})) is called reachability matrix and r_{ij} counts the total no. of walks of length p or less.

Distance matrix:

$D^p = X + X^{2^\wedge} + X^{3^\wedge} + \ldots + X^{p^\wedge}$

$D^p = ((d_{ij}))$ is called distance matrix and dij gives the length of the shortest path linking i & j.

$X^{p\wedge} = X^p$ in which all non-zero elements are set equal to p except for those elements that have been replaced by zero because they were non-zero in matrix X raised to some power less than p.

- The longest path distance in a network will equal to p when X^{p+1} has no zero elements in lower power of the matrix.

In case of simple SN (DD),

$$d_i = \sum_{j=1}^{n} X_{ij} = X_{i+} = i\text{-th row sum of the matix};$$

$$e_j = \sum_{i=1}^{n} X_{ij} = X_{+j} = j\text{-th column sum of the matrix}$$

$$m = \sum_{i=1}^{n} \sum_{j=1}^{n} X_{ij} = X_{++} = \text{grand total; and}$$

$$s = \sum_{i<j} \sum X_{ij} X_{ji} = \text{no. of reciprocal pairs}.$$

In case of weighted SN (DV),

$$d_i = \sum_{j=1}^{n} X_{ij} = X_{io} = i\text{-th row sum of the matix};$$

$$e_j = \sum_{i=1}^{n} X_{ij} = X_{oj} = j\text{-th column sum of the matrix}$$

$$m = \sum_{i=1}^{n} \sum_{j=1}^{n} X_{ij} = X_{oo} = \text{grand total; and}$$

$$s = \sum_{i<j} \sum \min(X_{ij}, X_{ji}) = \text{total reciprocity count}.$$

Different measures:

Suppose,
n = no. of hhs in the village
m = no. of ties within the village
p = no. of strong components (*strong component* is a set of actors in a digraph, who are tied each other directly or indirectly)
q = no. of weak components (weak component is a set of actors in a undirected graph, who are tied each other directly or indirectly)

→ Measures of reciprocity (to be discussed later);
→ Indices for
- Density (within village ties) = 100m/n(n-1);
- Strong connectedness = 100(n-p)/(n-1);
- Weak connectedness = 100(n-q)/(n-1);
- Strong fragmentation = 100(p-1)/(n-1);
- Weak fragmentation = 100(q-1)/(n-1);
- Hierarchy = 100(p-q)/(n-1).

Measures of reciprocity for DD SN data:
Suppose,
n = no. of hhs,
m = total ties,
d_i = out-degree of the i-th hh,
e_i = in-degree of the i-th hh,
s = variable defining no. of reciprocal pairs in a social network,
s_0 = observed no. of reciprocal pairs in a given social network,
s_{min} = minimum value of s under a model on certain given conditions,
s_{max} = maximum value of s under a model on certain given conditions,
μ_s = expected value of s under a model on certain given conditions,
v_s = variance of s under a model on certain given conditions,
σ_s = standard deviation of s = positive square-root of v_s.

Then there are two types of standardized measures to be calculated: deterministic (or graph-theoretic) and probabilistic, as given below respectively:
$s_d = [(s_0 - s_{min})/(s_{max} - s_{min})] \times 100$

$s_p = (s_0 - \mu_s)/\sigma_s$

Table-1

Deterministic approach for measures of reciprocity

Model	Specifications	Measures of reciprocity
I	n given	**s_d (I) = (200 s_0)/ n(n-1)**
II	n, m given	**s_d (II) = (200 s_0)/(m – ε)**, when $m < \binom{n}{2}$ and where ε is 0 or 1 according as m is even or odd.
III	n, d_1, …, d_n given and hence m given	**s_d (III) = (200 s_0)/(m – ε)** [= **s_d (II)**], when $d_{max} \le (n-1)/2$ and $d_{max} \le 2\sqrt{z}$ -2, where d_{max} = maximum out-degree, z = no. of hhs having non-zero out-degrees and ε is 0 or 1 according as m is even or odd.
IV	n, d_1, …, d_n and e_1, …, e_n given.	**s_d (IV) ≈ (200 s_0)/∑min(d_i, e_i),** assuming some conditions on the differences $e_i - d_i$ (Rao, 1984) which generally hold in real life data.

Table-2

Probabilistic approach for measures of reciprocity

Model	Specifications	Measures of reciprocity
I	n given, all possible networks are equally likely, P(i chooses j) = ½, distinct pairs independent	$\mu_s(I) = n(n-1)/8$, $v_s(I) = 3n(n-1)/32$ $\mathbf{s_p(I) \approx (3.27\, s_o/n) - (n/2.45)}$, when n is large (> 15, say)
II	N, m given, all possible networks are equally likely, m arcs chosen at random from n(n-1) possible pairs	$\mu_s(II) = m(m-1)/2\{n(n-1)-1\}$ $v_s(II) = \mu_s(II)\,[\,1 - \mu_s(II) + (m-2)(m-3)/2\{n(n-1)-3\}]$ $\mathbf{s_p(II) \approx (s_o n\sqrt{2}/m) - m/(n\sqrt{2})}$, when n is large (>10, say) and m/n(n-1) is small (<0.1, say).
III	n, d_1, …, d_n given, all possible networks are equally likely, P(i chooses j) = $d_i/(n-1)$, different vertices chose independently.	Define $t_k = \sum_{i=1}^{n} d_i^k$, $\mu_s(III) = \dfrac{t_1^2 - t_2}{2(n-1)^2}$, $v_s(III) = \mu_s(III)$ $+ \dfrac{1}{(n-1)^3(n-2)}(t_1^2 t_2 - t_2^2 - 2t_1t_3 + 2t_4 - t_1^3 + 3t_1t_2 - 2t_3)$ $- \dfrac{1}{2(n-1)^4}(2t_1^2 t_2 - t_2^2 - 4t_1t_3 + 3t_4)$. $\mathbf{s_p(III)} \approx \dfrac{s_o - \{n\bar{d}^2/2(n-1)\}}{\{\bar{d}\sqrt{n}/\sqrt{2(n-1)}\}\{1 - \bar{d}/(n-1)\}}$, when all d_is do not differ much and $\bar{d} = m/n$.
IV	n, d_1, …, d_n and e_1, …, e_n given, all possible networks are equally likely.	First, estimate the distribution of s by simulation applying MCMC method (Rao, Jana, Bandyopadhyay, 1996); find $\mu_s(IV)$ and $v_s(IV)$ from the estimated distribution of s; and calculate $\mathbf{s_p(IV)}$.

Basic idea of MCMC method for generating random digraphs with given out-degrees and in-degrees

Let H denote the set of all networks of n vertices with given out-degrees $d_1, \ldots, d_n$ and in-degrees $e_1, \ldots, e_n$.

Consider a digraph D with n vertices which can also be represented by its adjacency matrix $X = ((X_{ij}))$, where $X_{ij} = 1$ if there is an arc from i to j and 0 otherwise.

Alternating rectangle in D: If there exist four *distinct* vertices t, u, v, w such that t goes to u and v to w but v does not go to u and t not to w, i.e., $X_{tu} = 1$ and $X_{vw} = 1$ but $X_{vu} = 0$ and $X_{tw} = 0$, then the cells (t,u), (t,w), (v,w) and (v,u) of X form an '*alternating rectangle*' as shown in Figure 2.1A where the entries are alternately 1's and 0's as one goes around the rectangle.

Switching along the alternating rectangle: By it, we mean interchanging 1's and 0's in this rectangle, i.e., dropping the arcs tu, and vw, and introducing the new arcs tw and vu. Note that the out-degree and the in-degree of each vertex in D remain unchanged but we obtain a digraph D' different from D.

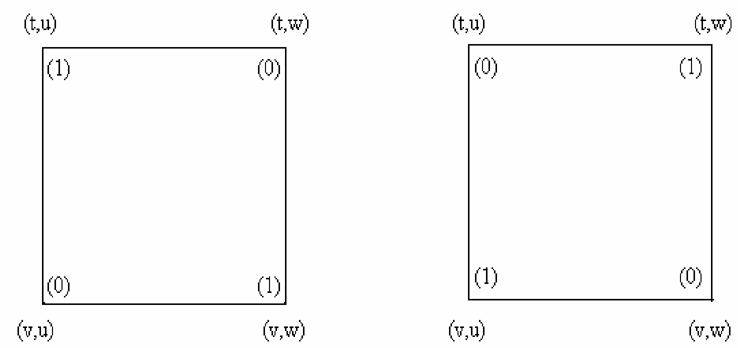


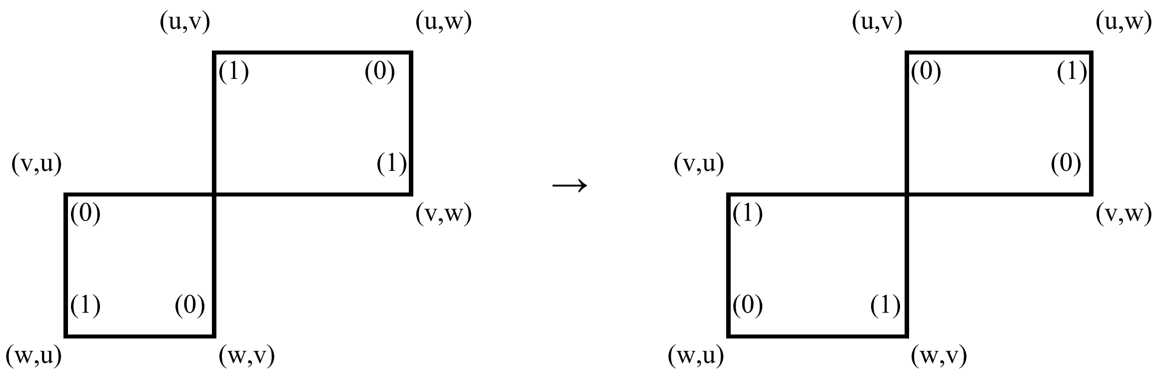
Figure2.1A: Before interchange

Figure2.1B: After interchange

Compact alternating hexagon in D: If there exist three *distinct* vertices u, v and w such that u goes to v, v to w, and w to u but v does not go to u, w not to v, and u not to w, i.e., $X_{uv} = 1$, $X_{vw} = 1$, and $X_{wu} = 1$ but $X_{vu} = 0$, $X_{wv} = 0$, and $X_{uw} = 0$, then the six cells (u,v), (w,v), (w,u), (v,u), (v,w), and (u,w) of X form an '*Compact alternating hexagon*', as shown in the following example and figures, where the entries are alternately 1's and 0's as one goes around the hexagon.

Switching along the compact alternating hexagon: By it, we mean interchanging 1's and 0's in this hexagon, i.e., dropping the arcs uv, vw, and wu and introducing the new arcs vu, wv, and uw. Note that the out-degree and the in-degree of each vertex in D remain unchanged but we obtain a digraph D' different from D.

$$\begin{array}{c} \\ u \\ v \\ w \end{array}\begin{array}{c} \begin{array}{ccc} u & v & w \end{array} \\ \left[\begin{array}{ccc} 0 & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & 0 & \mathbf{1} \\ \mathbf{1} & \mathbf{0} & 0 \end{array}\right] \end{array} \rightarrow \begin{array}{c} \\ u \\ v \\ w \end{array}\begin{array}{c} \begin{array}{ccc} u & v & w \end{array} \\ \left[\begin{array}{ccc} 0 & \mathbf{0} & \mathbf{1} \\ \mathbf{1} & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & 0 \end{array}\right] \end{array}$$



Now we use the term ***alternating cycles*** to mean alternating rectangles and compact alternating hexagons as defined above.

Step-1: Start with an initial matrix X

Step-2: Find t = min (no. of 1's and no. of non-structural 0's)

Step-3: List the alternating cycles of the current matrix and count their number (n, say)

Step-4: Choose one of the n alternating cycles at random with a modified way (discussed in Rao, Jana and Bandyopadhyay, 1996) and switch along it to get a new matrix.

Step-5: Consider the new matrix as initial matrix and repeat the step-3 to step-4 upto 3t times and the last matrix will be a random matrix.

Step-6: Repeat step-1 to step-5 as many times as one intends to generate the random matrices (>10000, say).

Some empirical findings obtained by analyzing the SN data of Kabilpur, collected in 1970 and 1998

Table3
Caste/Comm. & occupation-wise % distribution of households in Kabilpur(1970)

Caste/Comm.	Occupation				
	Rural poor*	Farmers (M & B)	Business	Others	**Total**
Brahmins	0.42	0.84	0.00	0.00	1.26
Mandals	9.62	8.37	0.00	0.00	17.99
Pals	3.77	2.51	0.00	0.00	6.28
Bagdis	**40.17 (n=96)**	0.00	0.42	4.60	45.19
Santhals	**22.18 (n=53)**	0.00	0.00	0.42	22.59
Others	4.18	0.00	0.00	2.51	6.69
Total	80.33	11.72	0.42	7.53	100.00 (n=239)

***Rural Poor: Factory labourers, Agricultural labourers, unspecified/other labourers, sharecroppers, marginal and small farmers.**

Observations from the Table-3:

- Both Mandals and Pals are Sadgopes. Initially, Pals were descendants of Kumhar (Potter-maker) caste and were later accepted as Sadgopes through an affair of marriage with Sadgopes about two generation ago.
- Mandals were economically and politically dominant; one of them being a big farmer with more than 100 acres of own cultivable land and an ex-MLA.

Since, in 1970, m= 871 and sd(III) = 89% (Rao and Bandyopadhyay,1987) which may be considered as high, it needs to go through the structure of the diagram of SN of Kabilpur(1970) which was reconstructed.

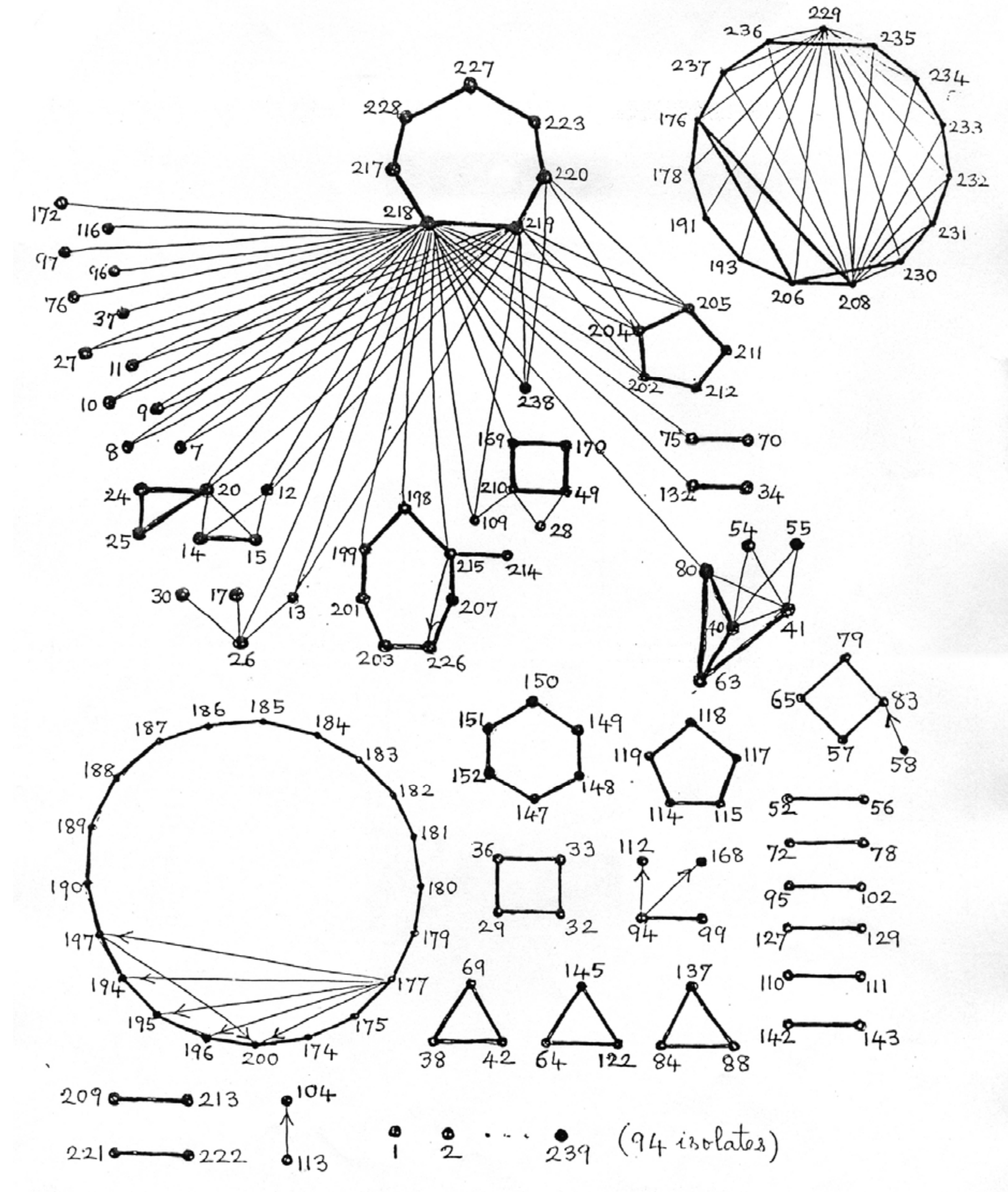


Diagram-A: Social network of Kabilpur, 1970

Observations from the diagram of SN, 1970 (Diagram-A):

- The network (1970) is divided into one pyramid, two large cliques, many smaller cliques and 94 isolates, even though high reciprocity. That is, the village is fragmented in spite of showing high reciprocity. This fact says that, reciprocity, by itself, does not tell us whether the village is close-knit and hence it leads to study various other aspects like connectedness and distances in the network.
- The households of the two large cliques belong to upper caste (i.e., Sadgope).
- The households of the low castes (Bagdi and Santhal) form a few small cliques or a majority being isolates.

Changes in social relations of 'help' in 1970 & 1998:

Table4
Caste/Comm. & occupation-wise % distribution of households in Kabilpur(1998)

Caste/Comm.	Occupation				
	Rural poor*	Farmers (M & B)	Business	Others	**Total**
Brahmins	0.00	0.23	0.00	1.39	1.62
Mandals	10.88	1.85	0.69	0.46	13.89
Pals	6.71	1.16	0.23	0.23	8.33
Bagdis	**46.99 (n=203)**	0.00	0.46	3.24	50.69
Santhals	**18.75 (n=81)**	0.00	0.23	0.93	19.91
Others	3.01	0.00	0.00	2.55	5.56
Total	86.34	3.24	1.62	8.80	100.00 (n=432)

Table-5
Occupation-wise % distribution of 'Rural poor' Bagdi and Santhal households in Kabilpur(1970 & 1998)

Year	Occupation				
	Small & marginal farmer or sharecropper	Agricultural labourer	Factory labourer	Unspecified / other labourer	**Total**
Panel A: Bagdi					
1970	4.17	26.04	0.00	69.79	100.00 (n=96)
1998	29.56	43.84	18.72	7.88	100.00 (n=203)
Panel B: Santhal					
1970	5.66	32.08	0.00	62.26	100.00 (n=53)
1998	44.44	43.21	11.11	1.24	100.00 (n=81)

Observations from the Tables 4&5:

- Many households of the caste Sadgope who were big and medium farmers have become small farmers because of land reforms and partitioning of families, and have lost their preeminent position and political power as well.
- The lot of Bagdis and Santhals has improved considerably with employment in factories, etc., several becoming share-croppers and marginal and small farmers, though the most are poor still.

<u>Measures of various aspects of SNs (1970 and 1998)</u>:

Table-6
Measures of various aspects of social networks in 1970 & 1998

Measures of various aspects	**Values of the measures for**	
	1970	1998
No. of hhs.	239	432
No. of ties within village	871	734
No. of ties within castes	800	366
No. of ties between castes	71	368
No. of ties going outside village	**0**	**274**
Total no. of ties	871	1008
Average out-degree	3.6	1.7
No. of reciprocal pairs	387	46
Reciprocity (Deterministic measure): s_d**(III)**	**89%**	**13%**
Reciprocity (Probabilistic measure): s_p**(III)**	**149.5**	**26.23**
No. of isolates (within village ties)	94	40
No. of isolates (considering both inside and outside vill. ties)	**94**	**3**
Percentage of reach. pairs (dir.)	**2.5%**	**15.1%**
No. of strong components	149	340
Size of largest strong comp.	20	70
Average finite distance (dir.)	1.2	5.8
Maximum finite distance (dir.)	4	16
Percentage of reach. pairs (undir.)	**7.7%**	**78.2%**
No. of weak components	114	45
Size of largest weak comp.	59	382
Average finite distance (undir.)	2.3	4.4
Maximum finite distance (undir.)	6	10

Table-7
Percentage distribution of ties by caste groups vs. caste groups in 1970

	Brahmins	Mandals	Pals	Bagdis	Santhals	Others	Total
Brahmins	**0.00**	0.00	0.00	0.00	0.00	0.00	0.00
Mandals	0.00	**54.31**	0.00	0.00	0.00	0.57	54.88
Pals	0.00	0.00	**21.58**	0.00	0.00	0.00	21.58
Bagdis	0.00	2.53	0.00	**5.40**	0.80	0.11	8.84
Santhals	0.00	0.80	0.00	0.80	**8.27**	0.57	10.45
Others	0.00	1.38	0.00	0.00	0.57	**2.30**	4.25
Total	0.00	59.01	21.58	6.20	9.64	3.56	100.00

- Ties within caste: 91.85%
- Ties between caste groups: 8.15%

Table-8
Percentage distribution of ties by caste groups vs. caste groups in 1998

	Brahmins	Mandals	Pals	Bagdis	Santhals	Others	Total
Brahmin	**0.82**	0.68	0.27	0.00	0.14	0.27	2.18
Mandals	0.27	**12.40**	3.00	0.41	0.27	0.68	17.03
Pals	0.14	3.00	**8.45**	0.00	0.00	0.14	11.72
Bagdis	2.72	12.81	8.99	**19.75**	0.54	2.45	47.28
Santhals	0.95	3.27	2.45	0.14	**7.08**	0.54	14.44
Others	0.95	3.00	1.50	0.27	0.00	**1.63**	7.36
Total	5.86	35.15	24.66	20.57	8.04	5.72	100.00

- Ties within caste: 50.14%
- Ties between caste groups: 49.86%

Table-9
Percentage distribution of reciprocal ties by caste groups vs. caste groups in 1970

	Brahmins	Mandals	Pals	Bagdis	Santhals	Others	Total
Brahmin	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mandals		58.66	0.00	0.00	0.00	1.29	59.95
Pals			21.45	0.00	0.00	0.00	21.45
Bagdis				5.17	1.81	0.00	6.98
Santhals					8.01	1.29	9.30
Others						2.32	2.32
Total	-	-	-	-	-	-	100.00

- Reciprocal ties within caste groups: 95.61%
- Reciprocal ties between caste groups: 4.39%

Table-10
Percentage distribution of reciprocal ties by caste groups vs. caste groups in 1998

	Brahmins	Mandals	Pals	Bagdis	Santhals	Others	Total
Brahmin	6.53	0.00	0.00	0.00	0.00	4.35	10.88
Mandals		30.44	2.17	4.35	2.17	2.17	41.30
Pals			21.74	0.00	0.00	2.17	23.91
Bagdis				10.87	0.00	2.17	13.04
Santhals					6.52	0.00	6.52
Others						4.35	4.35
Total	-	-	-	-	-	-	100.00

- Reciprocal ties within caste groups: 80.45%
- Reciprocal ties between caste groups: 19.55%

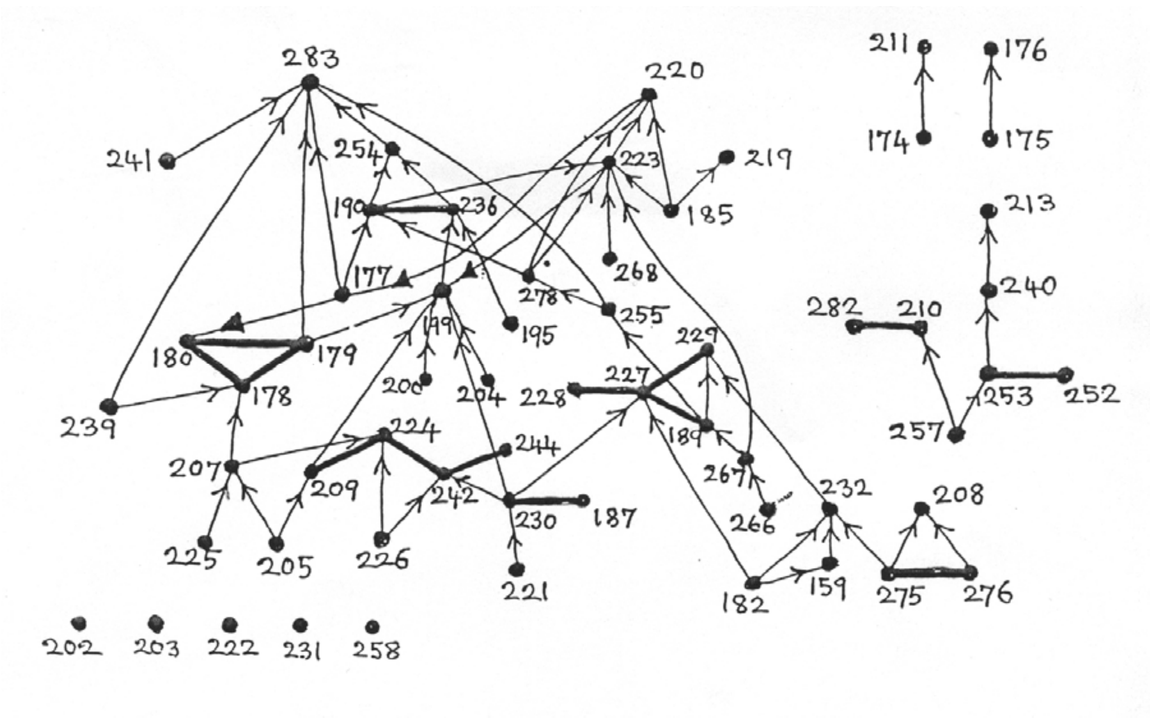


Diagram-B: Social network within Sodgope community
(excluding 'Pal' community) of Kabilpur, 1998

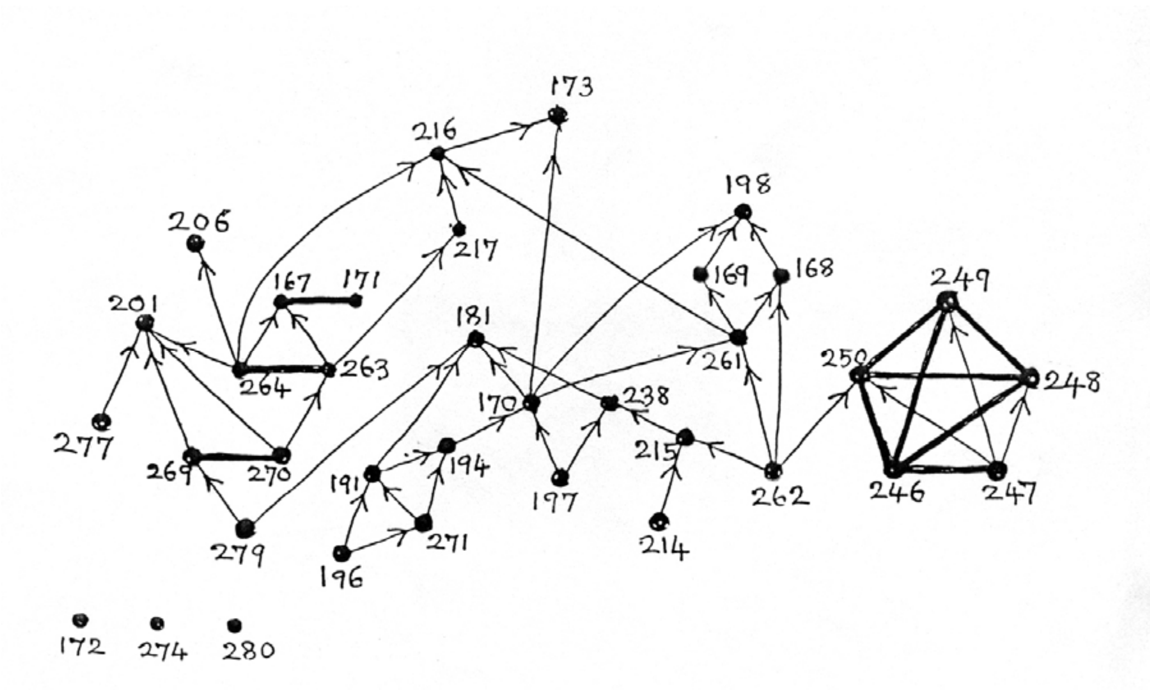


Diagram-B′: Social network within 'Pal' community of Kabilpur, 1998

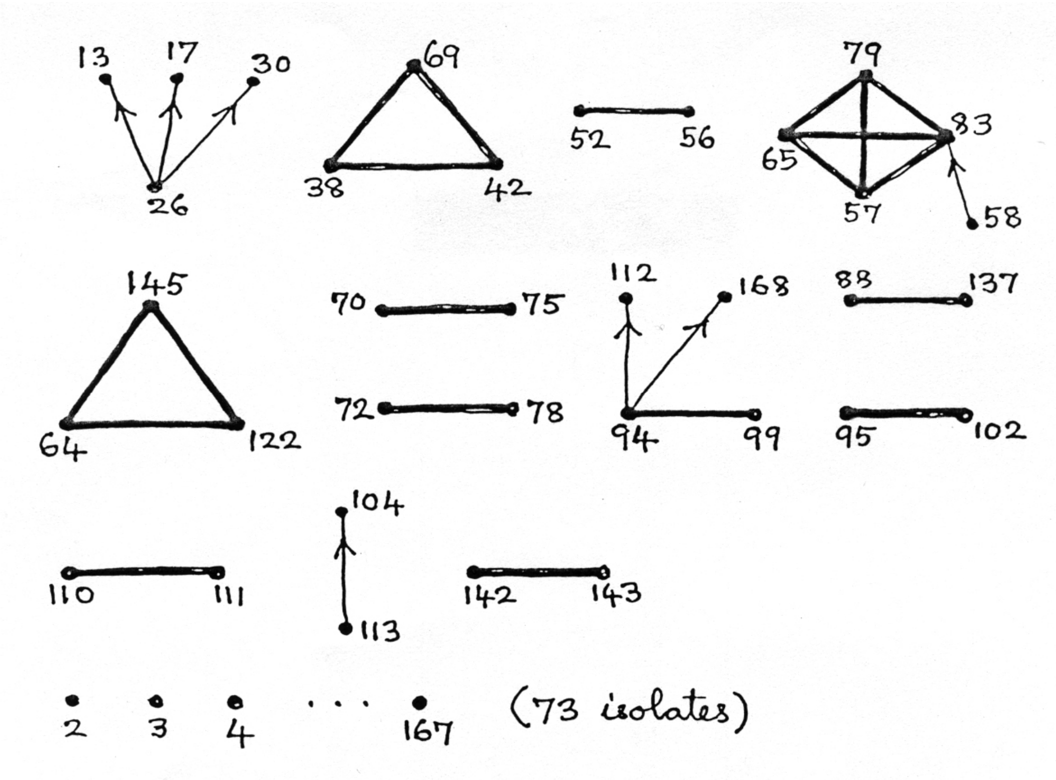


Diagram-C: Social network within Bagdi community of Kabilpur, 1970

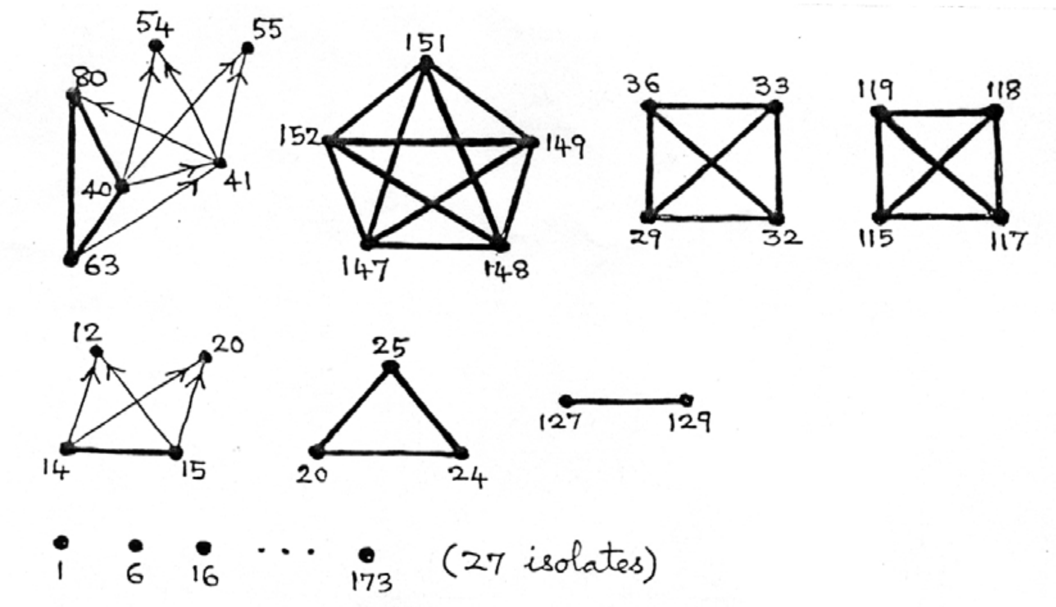


Diagram-D: Social network within Santhal community of Kabilpur, 1970

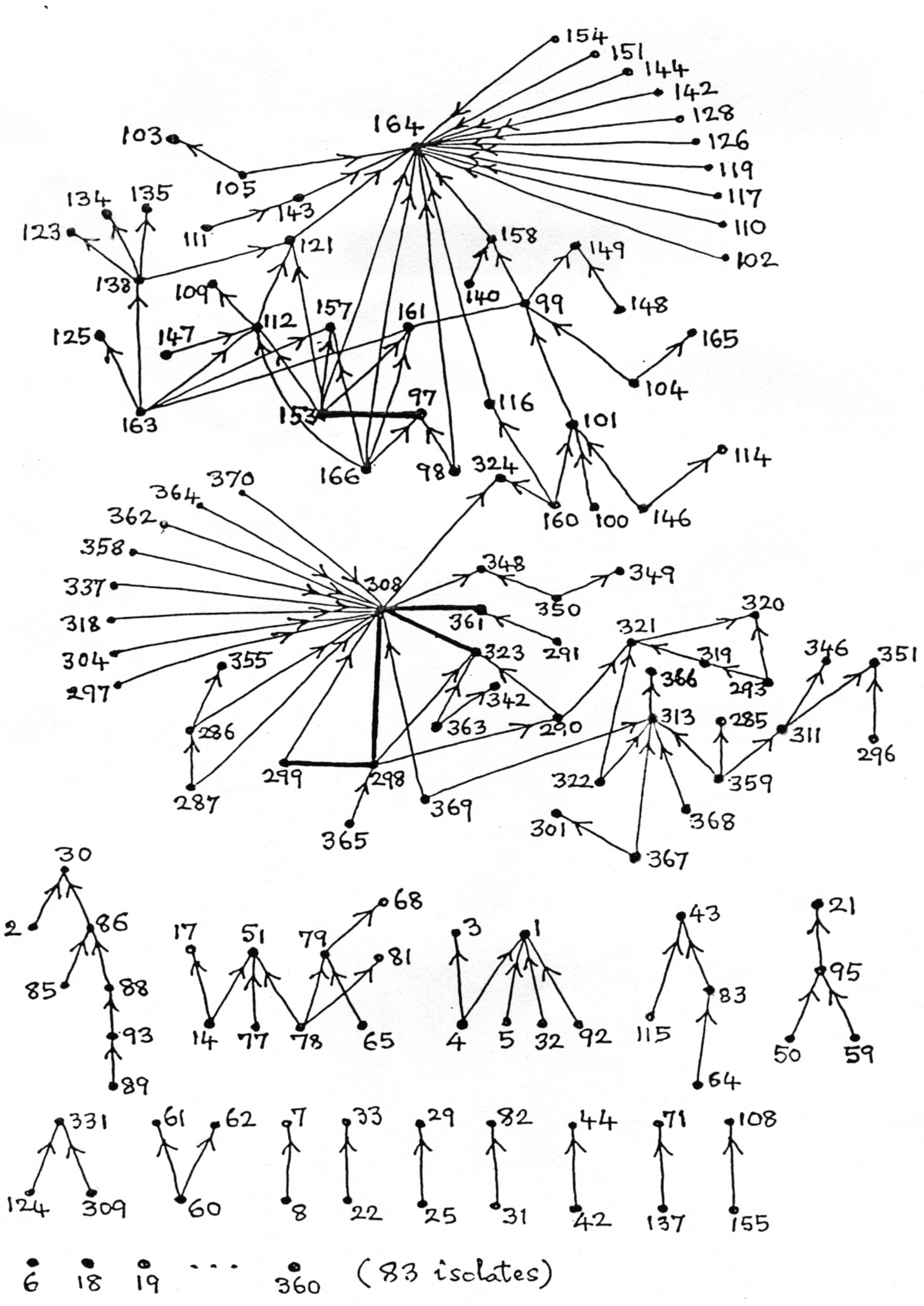


Diagram-E: Social network within Bagdi community of Kabilpur, 1998

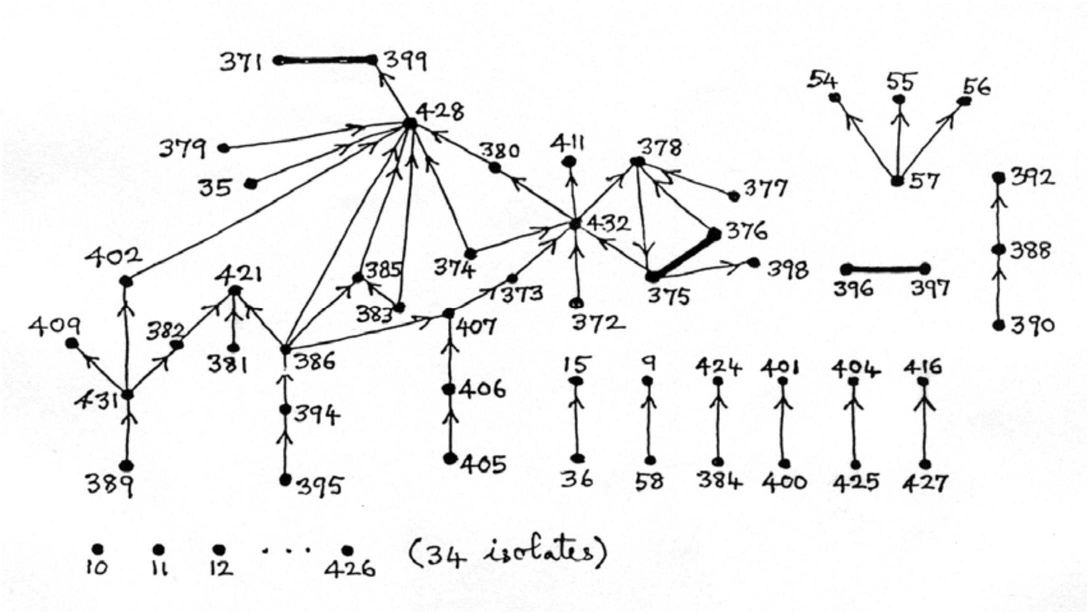


Diagram-F: Social network within Santhal community of Kabilpur, 1998

Major Findings:

- All the major castes have developed ties outside the village in 1998.
- Reciprocity has decreased drastically within each caste except in the case of Brahmins and this fall has been more steep within Bagdis and within Santhals and less steep within Mandals and within Pals. This reduction is mainly because of the disintegration of the cliques.
- The number of ties within the caste has decreased drastically in the case of Mandals and Pals and marginally in the case of Santhals and has increased in the case of Bagdis. Incidentally, the Brahmins were totally isolated in 1970, having felt alienated as a result of the passing of the lease of the village lands to the Sadgopes earlier, but by 1998, there were 16 ties originating from the Brahmins and 43 ties termination with them, 5 of these being reciprocal.
- The number of ties going out of each caste has increased considerably for all castes. Even, earlier Pals had no ties with others but, in 1998, they have several; there are 22 ties from Pals to Mandals and 22 ties from Mandals to Pals.

- The maximum out-degree and the maximum in-degree have reduced among the Mandals and Pals because of the disintegration of the large cliques. But the maximum in-degree has increased considerably among the Bagdis and the Santhals, indicating that some of them like those with hh. serial nos. 164, 308 and 428 have become centres of power and influence. From our field visits, we also came to know that, politically and economically, the position of these two Bagdi households and one Santhal household has gone up considerably in 1998. It was also found that, generally, the Bagdis and Santhals now have wider choice of households from whom they can get help; and the help is more easily forthcoming.
- The number of isolates has reduced drastically among the Bagdis and marginally among the Santhals. If the ties outside the village are taken into account, there are only two Bagdi households and one Santhal household who are isolates.
- Though reciprocity has reduced, ties have spred out and connectedness and reachability in the undirected sense have increased. Thus the maximum number of households who can reach each other through paths when directions are ignored, has increased in each caste. The increase is dramatic in the case of Bagdis and Santhals. Of course, this reachability is tempered with having to go through long chains as the increase in the maximum (finite) distance shows.
- In case of occupation, in 1970, 60 of the 387 reciprocal pairs were within the big & medium farmers and big & medium businessmen. In 1998, there is no reciprocal pair within them. Between the rural poor and the rest, there is not much change in the situation between 1970 and 1998. Within the rural poor, there were 170 reciprocal pairs in 1970 (i.e., 44% of all reciprocal pairs) whereas in 1998, there are 34 reciprocal pairs (i.e., 76% of all the reciprocal pairs). Thus, a much larger proportion of the reciprocal pairs are within the rural poor in 1998 than in 1970.

Conclusion:

- Economic and political changes had a positive impact, in terms of social relations, particularly on those at the lower rungs of rural society in West Bengal.
- The lower rungs of the rural society in Bengal have opened up and have developed ties with the outside world. In fact, it was found that even women do approach and take help from their own employers as distinct from employers of their husbands.
- Reciprocal relations have not developed but mini-pyramidal hierarchies have developed among the lower sections.
- The networks of the entire village in 1970 and 1998 have been analysed, not on the basis of sample networks. Hence, one may examine the possibility of using sampling to study the various aspects of a full network with less cost.

References

Bandyopadhyay, S., Rao, A. R., and Sinha, Bikas K.(2010). *Models for Social Networks with Statistical Applications*. Sage Publications.

Jana, R., Bagchi, S.B., and Rao, A.R. (2003). A probabilistic model for weighted social network. Presented at the *Fifth International Triennial Calcutta Symposium on Probability and Statistics*. 28-31 December 2003, jointly organized by Department of Statistics, Calcutta University and Calcutta Statistical Association.

Knoke, D., and Kuklinski, J.H. (1988). *Network Analysis*. Sage Publications.
Moreno, J.L. (ed.) (1956). *Sociometry and the science of man*. Beacon House, New York.

Rao, A.R., and Bandyopadhyay, S. (1987). Measures of reciprocity in a social network. *Sankhy*. Series A, **49**, 141-188.

Rao, A.R., and Bondyopadhyay, S. (1998). *A Study of Changing Social Relations—Social Network Approach*. A Project report, funded by Survey Research and Data Analysis Centre, Indian Statistical Institute.

Rao, A.R., and Rao, S.B. (1992). Measuring reciprocity in weighted social networks. *Sankhy*. Special volume, **54**, 349-355.

Rao, A.R., Jana, R., and Bandyopadhyay, S. (1996). A Markov chain Monte Carlo method for generating random (0,1)-matrices with given marginals. *Sankhy*. Series A, **58**, Pt. 2, 225-242.

Scott, J. (1991). *Social Network Analysis: A Handbook.* Sage Publications.

Snijders, Tom A.B. (1991). Enumeration and simulation methods for 0,1 matrices with given marginals. *Psychometrika*. **56**, 397-417.

Wasserman, S., and Faust, K. (1999). *Social Network Analysis: Methods and Applications*. Cambridge University Press.

A bit about late Professor Suraj Bandyopadhyay

(*by Rabindranath Jana, Sociological Research Unit,
Indian Statistical Institute, Kolkata*)

Birth: 1 September, 1932 **Death: 29 September, 2010**

I would mainly say about his academic achievements briefly with a short background of his pioneering work on SNA at ISI.

Brief academic achievements

Professor Suraj Bandyopadhyay (PhD, Sociology, McGill University, Montreal, Canada) was, before his retirement on 31 August, 1992, Professor of Sociology and Head of the Sociological Research Unit, Indian Statistical Institute, Kolkata. After retirement, for sometime, he was also affiliated with the Sociological Research Unit and Theoretical Statistics-Mathematical Unit, Indian Statistical Institute, Kolkata, as an Honorary Visiting Scientist. For his academic performance at McGill University, he was awarded Bobbs-Merrill Award in Sociology in 1968. He has received a number of academic invitations and Fellowships from different international institutions, such as, the Canadian International Development Agency (CIDA, Ottawa), Centre for Developing-Area Studies (McGill University), International Data Library and Reference Service of the Survey Research Centre at the University of California (Berkeley, USA), and Overseas Development Group at the University of East Anglia (Norwich, UK). He has served as expert member in academic committees of different institutions and published more than thirty research papers. Before few months of his death, his book (entitled "Models for Social Networks with Statistical Applications" jointly with A.R. Rao and Bikas K. Sinha) has been published by Sage, USA.

Background of his pioneering work on SNA at ISI

His PhD work was a part of a larger study on "The Conditions of Rural Progress in India", jointly undertaken at Dept. of Sociology (McGill University, Canada) by Suraj Bandyopadhyay and Donald von-Eschen of McGill University and funded by Canadian International Development Agency, for understanding the causes of failure of the Indian Community Development Programme and the impact of that programme on village economy and social structure in 1970. The study covered almost 2700 households from the sampled 21 villages of Md. Bazar block in Birbhum district of West Bengal. Relational dimension was incorporated in the study along with other various socio-economic dimensions. During his field visits, he felt that relational structure may be useful in the study. When he analyzed the data, he observed from sociological point of view that, along with the other socio-economic factors, the study on the patterns of relational ties is very much needed to get a more complete picture of the causes of failure of the programme. After completion of his PhD work and other academic assignments in abroad, he came back to ISI. Then he exercised his collected SN data in various ways, specially on reciprocity obtained by dividing the number of reciprocal pairs by n(n-1)/2, n being the no. of hhs. During this period, he became surprised by seeing that a village network with low no. of reciprocal pairs shows more reciprocity value than the other village network with large no. of reciprocal pairs. To solve this problem, he was introduced to Prof. A.R. Rao (a graph-theorist of Stat.-Math. Division, ISI, Kolkata) by Prof. B.P. Adhikari, the then Director of ISI. Thereafter, they needed its standardization for comparison purpose. After a lot of their joint efforts, they developed deterministic and probabilistic measures under different models for standardizing the no. of reciprocal pairs which was published in Sankhya (1987), Series A. In 1998, SN data of Kabilpur (one of the earlier 21 villages in 1970) were again collected by the ISI team (including myself and others) leaded by Prof. A.R. Rao and Prof. Suraj Bandyopadhyay. Thus, till their death, both of them became more interested and continued jointly SN study at ISI with active association of myself, Dr. Anil K. Choudhuri and Dr. Arun K. Chatterjee of SRU, and Prof. Bikas K Sinha, Prof. S.B. Rao (former Director of ISI) and Prof. Debapriya Sengupta of Stat.-Math Division, Kolkata.

Though, at the end of his age, he spent more time for SNA, he had also enough capability in other areas of sociological research. He had strong coworker-feelings. He always encouraged the other colleagues of SRU to enrich the area of their research interest, specially the junior colleagues of SRU.

He passed away on 29 September, 2010 while, on an academic assignment, he was visiting the C.R. Rao Advanced Institute of Mathematics, Statistics and Computer Science (AIMSCS) at Hyderabad. Before that, both Prof. A.R. Rao and Dr. Arun K. Chatterjee had died untimely in 2006 and 2007 respectively.

Today I would again pay my homage to their memory.

Lastly, not the least, I would express my sincere gratitude to the organizers of the Workshop for inviting me to deliver the lecture as well as to participate in the Workshop and to say a bit about Prof. Suraj Bandyoapadhyay.