(RESEARCH ARTICLE)

Social cohesion and optimal development outcomes: A study through graph theory, Markov Chain and Artificial Intelligence

Sugata Sen [1] [*], Soumya Sengupta [2] and Santosh Nandi [3]

[1] Department of Economics, Panskura Banamali College, Panskura, Purba Medinipur, West Bengal, India, 721152.
[2] Faculty of Computer Science, Panskura Banamali College, Panskura, Purba Medinipur, West Bengal, India, 721152
[3] Research Scholar, Panskura Banamali College Research Center in Science affiliated to Vidyasagar University, West Bengal, India, 721152.

Abstract

Lack of cohesion among human beings is a major problem in modern India. This absence of cohesion appeared due to many historical, geographical and economic factors. It has ultimately culminated in differentiated development outcomes in India. This work wants to evaluate the effect of social cohesion on the development achievements. To that respect Graph theory, Markov chain analysis and Artificial Intelligence have been used. Here society has been conceived as a network of n agents. To initiate the Markov process the idea of Genetic Algorithm is used. On the other hand this work tries to measure the Cohesion Density Index for the concerned network and correlation with the time and apace complexity of that very network. This work actually compares the stationary states through Artificial Intelligence to find the most acceptable or optimum. It is concluded that the level of cohesion among different social groups can largely influence the expected outcome of any inclusionary development programme. This lack of cohesiveness within the Indian society due to some strong religio-philosophical reasons is one of the greatest challenges to India in achieving desired inclusive growth. In this paper it is shown that increasing the connectedness of the society can improve the efficiency of the development programs.

Keyword: Social Cohesion; Genetic Algorithm; Graph Theory; Markov Chain; Artificial Intelligence

JEL Code: C63, O15

1. Introduction

Lack of cohesion among human beings is a major problem in modern India. This absence of cohesion appeared due to many historical, geographical and economic factors. It has ultimately culminated in differentiated development outcomes in India. This work wants to evaluate the effect of social cohesion on the development achievements. To that respect Graph theory, Markov chain analysis and Artificial Intelligence have been used. It is found that greater cohesion leads to more efficient development outcomes.

2. Literature Review

Mukherjee (Mukherjee & L.Roy) et. al. tried to find out women's perspective levels and patterns of social cohesion and its relationship with development in India. Social cohesion as common value and civic culture, affects social order and social control. Social cohesion as social networks and social capital place attachment and identity. Social cohesion is used to measure quality of societies (Schmitt, 2000). The idea of artificial intelligence can be used to study social

---

[*] Corresponding author: Sugata Sen

cohesion. Rainer (Schnell, 1991), Ramesh et. al. (Ramesh, Kambhampati, Monson, & Drew, 2004) , Malenic et.el. (Claudé & Combe), have tried to provide a better understanding of the role of human and Artificial Intelligence in the organization decision making process. Tin Miller (Miller, 2018) applied Artificial Intelligence in social sciences, explicitly explaining decisions or action to a human observer

2.1. Hypothesis

Greater cohesion within the society leads to efficient development outcomes.

---

3. Methodology

Society can be conceived as a network of n agents. Naturally there can be $2^{\frac{n(n-1)}{2}}$ = R social network structures on the basis of the existence of inter-nodal incidences. If we express each of the structures as a matrix then there can be same number of matrices as the number of available graphs. From each of these matrices we can create unique transitional matrices to carry forward Markov process. To initiate the Markov process the idea of Genetic Algorithm is used. Naturally each of the social network graphs is treated as a mating pool. Each of the Markov process converges these mating pools to unique stationary states. On the other hand Cohesion Density Index (CDI) measures the cohesiveness of the society. It is expressed as the ratio of the actual number of incidences within the network to the maximum possible incidences within that network. This work tries to measure the CDI for each of the network and correlative with the time and apace complexity of that very network. Time complexity has been measured through the number of iteration to the stationary state. The correlation between the number of iteration and CDI is measured through the correlation coefficient. This work actually compares the stationary states through Artificial Intelligence to find the most acceptable or optimum. In each of the cliques is considered as a mating pool then this mating pools constitutes the first layer of Artificial Neural Networks. The outputs obtained through Markov process can constitute the second layer. In the third layer the nest network is obtained through the correlation between CDI and number of iteration.

3.1. Model

In this study the use of genetic algorithm has been shown as a use of Markov chain. In a society a household or a chromosome can parallelly be members of different social groups on the basis of different inherent characteristics. This multiplicity of respondent characteristics will definitely reduce the society to different inserting subsets. If there are n characteristics of the society and if each characteristic has m phases then we would have m^n mutually exclusive intersections within the society. If we treat each of these intersections as a sub set of the whole society then it can be said that the society consists of m^n = M sub sets.

Let us assume that the length of the chromosomes in all the sub sets is L. From each of the sub sets we can select representative chromosomes through a selection operator. This selection operator will work through a fitness function F. Thus from the population we create a mating pool (P) of m^n chromosomes. Then iteration t delivers M solution or chromosome string. Such that

P^t = { $S_1^{(t)}$, $S_2^{(t)}$, $S_3^{(t)}$, … … … $S_M^{(t)}$}

Each string $S_i^{(t)}$ is evaluated on the basis of its inclusion status expressed through the fitness function F. Members of P(t) can undergo reproduction through crossover and mutation to create new population of solution for the next iteration.

Hamming Distance between the strings of the mating pool is expressed as

HA = (ha_{ij}| i = 1,…M, j = 1,…M, i ≠ j) and

ha$_{ij}$ = $\sum_{k=1}^{L} D_{ij}^k$

Now if $S_i^{(t)}$ = (γ_k | k = 1,… L) where γ_k is the individual gene or the capability variable in the string i. Then D_{ij}^k is the intensity of mismatch between string i and string j with respect to γ_k. Divergence between the strings in a particular mating pool exists when

|ha$_{ij}$| > 0.

Here the genetic algorithm is used to minimize $|ha_{ij}|$ or to maximize a function f(x), x ϵ D where D is a finite set. The problem here is to find x_{opt} such that

$f(x_{opt}) ≥ f(x)$; for all x ϵ D

Here D is a discrete domain since D is finite. If L is the number of parameters to be considered using GA and A_i represents the finite set of possible values of the i th parameter then $D= A_1 X A_2 \dots A_L$

We can consider any mating pool as a state of a Markov Chain thus the probability of creating a offspring through crossover between s_i and s_j can be expressed as $η_{ij}$ such that i=1…m, j=1…..m and i≠j. Thus a transition matrix of i X j is created where $\sum_{j=1}^{M} η_{ij} = 1$ and $0≤ η_{ij} ≤ 1$. Thus this transition matrix T is a m x m square matrix. Each element of T represents the transitional probability from S_i to the offspring create through the crossover of s_i and s_j.

A distribution matrix Q^t corresponding to P^t can be concept with the relative fit of each string in a certain phase. Thus Q^t is a 1 X M matrix. Thus this matrix is created following the bellow stated arguments:

Calculate the fitness value $F(S_i)$, for each chromosome S_i (i= 1,2,3….M).

Find the total fitness of the population $F = \sum_{i=1}^{M} F(S_i)$

Calculate the g_j^t, as the element of Q^t corresponding to S_j as $g_i = \frac{F(Si)}{F}$ where S_i (i = 1,2,….M)

The distribution matrix Q^t together with T constitute with Markov process which will systematically carry forward the process of genetic algorithm till x_{opt} is reached. The number of rows and columns in T is M. Every element of T is $η_{ij}$ which denotes the probability of transition from s_i to the offspring of crossover between S_i and S_j. Let $η_{ij}^t$ be the probability of the GA results in Q^t at the t th state given the initial stage as Q^0. Thus $Q^t = Q^{t-1} . T$ or in other words $Q^t = Q^0.T$.

The aim of this study is to derive efficiency of the Markov chain on the perspective of different social network structures. Depending upon different perspectives the society can take different shapes as well as density. The society can be a complete network or an incomplete graph. It can also be decentralized or highly centralized. If W is the set of all probable network structures of the society then

$W = \{W_1, W_2, \dots\dots W_n\}$

If the whole society is represented as a complete connected graph of M nodes, then it it can be written that $W_1= \{ S_i, e_{ij} \mid$ i=1,2,3,….M and i ≠ j\} where the total number of edge is $\frac{M(M-1)}{2}$.

On the other hand if we think the society about a incomplete decentralized network then the society can be represented through a Hamiltonian circuit. From the available social structure W_1 , H can be generated where

$H = \{ h_i \mid i = 1,2,3\dots..α\}$

Here h_i is Hamiltonian circuit where

$h_i = \{ S_i, e_{ij} \mid$ i=1,2,3,….M and i ≠ j\}

    From H , h_{opt} will be selected through selection operator using a fit function on the basis of minimum cumulated edge weight of a particular Hamiltonian circuit. Thus $W_2= \{ S_i, e_{ij} \mid$ i=1,2,3,….M and i ≠ j\} where the total number of edge is M and

$W_2 = h_{opt} = \min_{1-α} \sum_{i,j=1}^{M} e_{ij}^{β}$ where i ≠ j and β = 1,2,…. A

This Hamiltonian structure is nothing but a decentralized circular structure of the society.

On the other hand we can think about a centralized star network structure of the society through defining a node as the centre. This centrality can be defined through a selection operator on the basis of a certain fit function. In this structure

if s* is at the centre then the society W_3 = = { S_i, e_{ij} | i=1,2,3,....M and i ≠ j} where degree of incidence of s* is M-1 and the degree if incidence of other M-1 nodes is 1.

Thus the whole work reduces to the comparison of the Markov process $Q^t = Q^{t-1} . T$ or in other words $Q^t = Q^0.T$ in W. The stable state of the genetic operation of each of the network structure is achieved when $Q^* = Q^*.T$ at each structure.

The number of iteration performed by a search, given that the corresponding path length l, is l+1 counting the initial iteration. If the total number of elements are p, which is positive integer, so initial path length l(p). Then the average number of iteration for successful search

$$T(p) = 1 + \frac{l(p)}{p}$$

If we use binary search technique then the problem is reduced to calculating the internal path length with p items.

Our aim is to find the node from the second layer on the basis of the length of the iteration. Higher the length the iteration lower will be the efficiency of that network. So our identification function to create the third layer is Minimize (l(p)).

Let the CDI of the Rth network is $CDI_R = \frac{i_R}{n-1}$ where i is the actual number of incidence of the Rth network. Again the number of iterations to achieve the stationary states is measured through the path length $l(p) = \sum_{k=1}^{p} log_2 (k)$.

The correlation between CDI and l(p) is expressed through

$$\rho = \frac{1}{p-1} \left\{ \frac{\sum_{CDI} \sum_{l(p)} (CDI - CDI')(l(p) - l(p)')}{S_{CDI} \ S_{l(p)}} \right\}$$

The best network is located in the third layer through the consideration of the ρ. This can be achieved where $\min_{CDI \to 1}$ (l(p)). So the optimum will be achieved at CDI → 1 or when l(p) is minimum.

---

4. Results

In this work we consider 320 different households from different economic factors. Each household represents a node of a graph. So it is clear that we can create a graph with 320 different nodes. As a whole 320 nodes represent the overall society. With the help of Graph theory we tried to establish inter- nodal relationship among society. This is because we tried to create as many potential combinations as possible. We have seen that $2^{\frac{320(320-1)}{2}}$ or 2^{51040} different graphs are being created. Using the Markov logic on each graph, we were able to reach each of the graph's separate stationary state. It is interesting to observe that the graph which reaches the stationary state through the lesser iteration has the highest cohesion index. After identifying the most acceptable social graph using Artificial Intelligence, it is the main responsibility of the work to show the similarity between the graphs with the highest cohesion index. The Artificial Intelligence technique tells that inter-nodal relationships among the graphs are the best results. The graph with highest connectivity will show the greatest value of social cohesion.

The society as a whole can be denoted as the population. This society can be characterized as the combinations of different characteristics like location of residence, gender and caste. Such that a society consists of the rural and urban areas; males and females as well as scheduled tribe communities and other communities. It is also observed that the inclusion status of any household is significantly influenced by the intersection of these characteristics. As these characteristics are independent of each other we can get 2^3 combinations or intersections of these characteristics. These intersections are shown as follows:

Table 1 Node selection scheme

	Location of residence and gender			
	Urban (U)		Rural (R)	
Caste	Female (F)	Male (M)	Female (F)	Male (M)
Scheduled tribe (S)	USF (S_1)	USM (S_2)	RSF (S_5)	RSM (S_6)
Other castes (O)	UOF (S_3)	UOM (S_4)	ROF (S_7)	ROM (S_8)

Each of these intersecting sets consists of several households conceived as chromosomes or strings of genes. Each of these genes or capability variables has been used to determine the exclusion status of the households on the basis of their characteristics. Such that each of these chromosomes are of equal length which is here 17 i.e. each chromosome has 17 genes. Each of the intersecting sets are treated as a node represented by the best chromosome of the node concerned. The best chromosome from each of the nodes are selected through the inclusion status. Here the fitness is defined through the existence of highest number of zero in the chromosome. Our capability variables are binary in nature where 1 means exclusion on the concerned variable and zero is the inclusion on that gene. The inclusion exclusion status of each of the intersecting sets represented by their representing chromosomes are expressed through the following table.

Table 2 Node-wise inclusion exclusion status

Nodes	Level of exclusion	Level of inclusion
S_1	09	08
S_2	07	10
S_3	06	11
S_4	04	13
S_5	11	06
S_6	09	08
S_7	05	12
S_8	05	12

The $e_{i,j}$ of the edge weight matrix can be determined through the normalized Euclidian distance in the exclusion space. The determination of the edge weight has been discussed in full details in Appendix - II. The calculated inter-nodal edge weights are as follows.

Table 4 Edge weight matrix under complete network structure

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	-	0.4	0.25	0.5747	0.7139	0.5763	0.6376	0.5787
S_2		-	0.4527	0.6887	1.0469	0.8918	0.8691	0.8186
S_3			-	0.3288	0.6988	0.529	0.4301	0.3747
S_4				-	0.8234	0.6537	0.2589	0.2581
S_5					-	0.1796	0.6665	0.6043
S_6						-	0.525	0.4537
S_7							-	0.0824
S_8								-

Using these edge weights, the society or the mating pool P^0 can be described with the following complete graph.
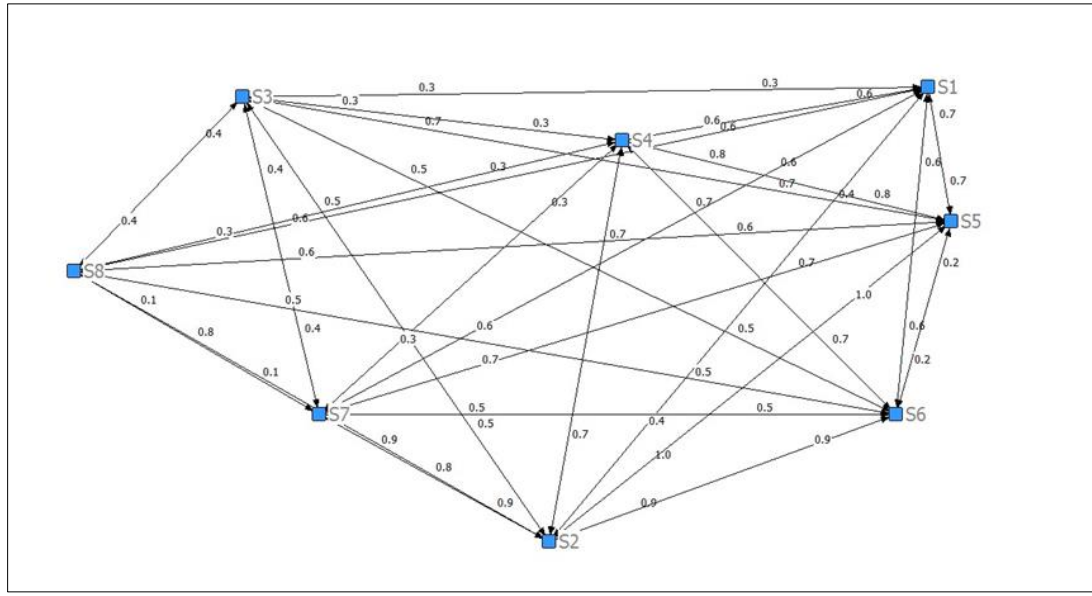
Figure 1 Initial Mating Pool

The above depicted complete network is the initial mating pool under W_1. From this mating pool we can generate the initial distribution matrix and stochastic transition matrix. Here the initial distribution matrix Q^0 corresponding to P^0 is a [1 X 8] matrix. To create this matrix fitness value $F(s_i^0)$ of each of the chromosome s_i^0 is calculated. The fitness is determined through the inclusion status or the number of zero in each string s_i^0. After calculating the $F(s_i^0)$ the total fitness F of the mating pool is calculated. Thus the following table is generated.

Table 5 Fitness Table of Q0 under W1

Chromosome	Fitness	Relative Fitness $g_i^0 = \dfrac{F(s_i^0)}{F^0}$
s_1^0	8	0.0987654320987654
s_2^0	10	0.123456790123456
s_3^0	11	0.135802469135802
s_4^0	13	0.160493827160493
s_5^0	6	0.0740740740740741
s_6^0	8	0.0987654320987654
s_7^0	12	0.148148148148148
s_8^0	13	0.160493827160493
Population	81	1

Each element of the initial distribution matrix Q^0 is nothing but the relative fitness of each of the string available in the initial mating pool. Thus

$g_i^0 = \frac{F(s_i^0)}{F^0}$ where i= 1,2,....8

This Markov process reaches the stable state after the 6 iterations. This means that

$Q^6 = Q^7$

or $Q^6 = Q^6$. T

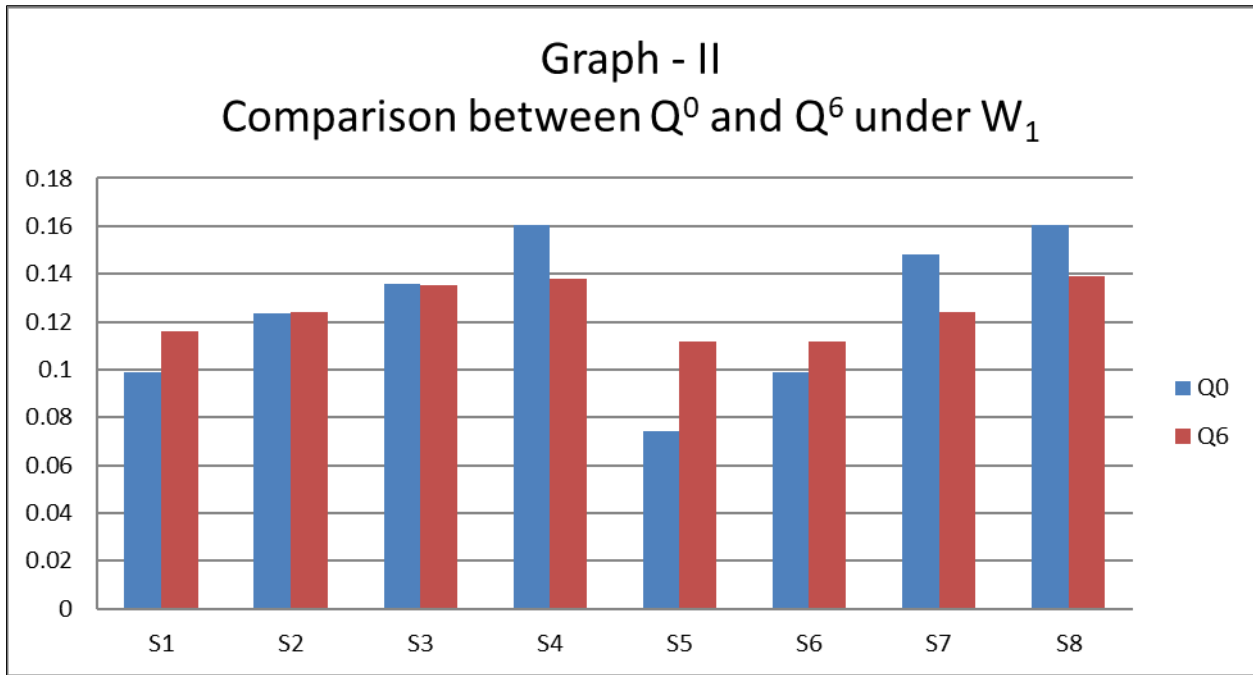Now comparing the Q^6 with Q^0 we get the following diagram.

Figure 2 Comparison between Q0 and Q6 under W1

It appears from the above discussion that nodes S_1, S_3, S_5, and S_6 have registered improvement in their respective relative fitness while S_2, S_4, S_7 and S_8 have registered deterioration in the same.

The following table can be derived from the above.

Table 7 Character-wise improvement and deterioration after Markov process under W1

	Number of nodes where present	Number of nodes where improved	Number of nodes where deteriorated	Improved nodes : deteriorated nodes
Scheduled Tribe	4	4	0	Very large
Other Castes	4	0	4	0
Females	4	2	2	1
Males	4	2	2	1
Rural	4	2	2	1
Urban	4	2	2	1

It appears that under W_1 that is under the complete network structure the inclusion status of the scheduled tribe community and the females will improve through applying the rules of genetic algorithm. In other words, it can be said that the existing development programmes can deliver better results through establishing greater interaction among the different ethnic groups and the males and females.

We can reduce this networked complete graph to a set of Hamiltonian circuits. The Hamiltonian circuit having the minimum cumulated edge weight is used as the initial structure to apply the genetic cross-over. The algorithm to determine the minimum Hamiltonian circuit is given in Appendix - IV. Thus in this hypothetical situation W_2 the mating pool is presented through the following matrix.

Table 8 Edge weight matrix under Minimum Hamiltonian circuit or W2

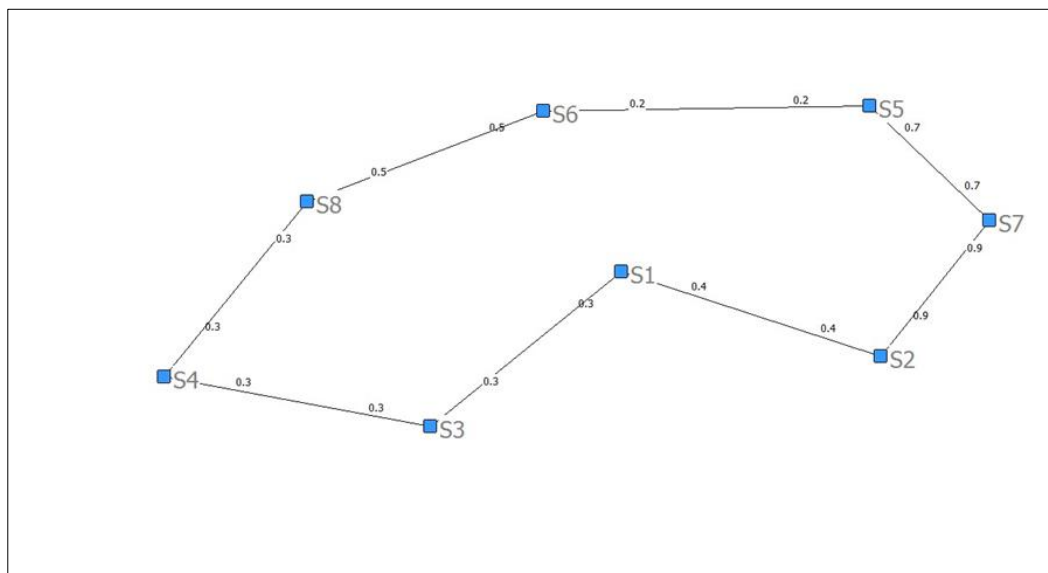|	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	-	0.45	0.25	-	-	-	-	-
S_2	0.45	-	-	-	-	-	0.8691	-
S_3	0.25	-	-	0.3288	-	-	-	-
S_4	-	-	0.3288	-	-	-	-	0.2581
S_5	-	-	-	--	-	0.1796	0.6665	-
S_6	-	-	-	-	0.1796	-	-	0.4537
S_7	-	0.8691	-	-	0.6665	-	-	-
S_8	-	-	-	0.2581	-	0.4537	-	-

Figure 3 Hamiltonian Circuit

The above depicted minimum Hamiltonian circuit is the initial mating pool under W_2. From this mating pool we can generate the initial distribution matrix and stochastic transition matrix. Here the initial distribution matrix Q^0 corresponding to P^0 is a [1 X 8] matrix as under W_2. To create this matrix same procedure as under W_1 is followed.

This Markov process reaches the stable state after the 57 iterations. This means that

$Q^{57} = Q^{58}$

or $Q^{57} = Q^{57}. T$

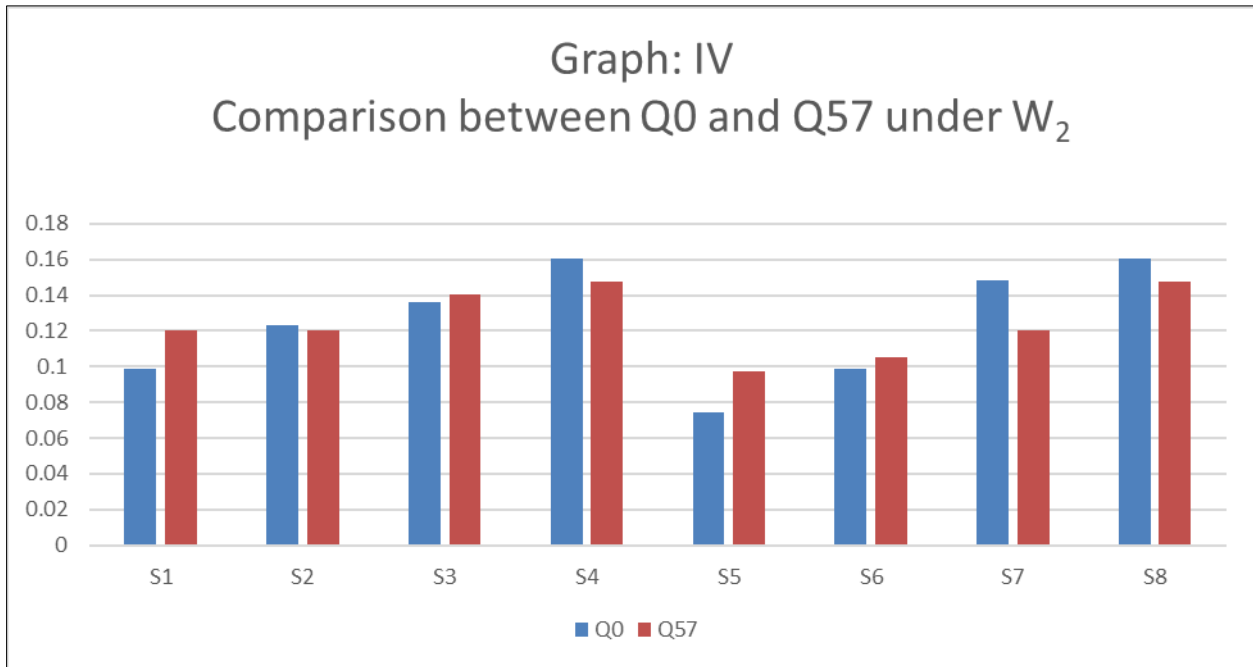Now comparing the Q^{57} with Q^0 we get the following diagram.

Figure 4 Comparison between Q0 and Q57 under W2

It appears from the above discussion that nodes S_1, S_3, S_5 and S_6 have registered improvement in their respective relative fitness while S_2, S_4, S_7 and S_8 have registered deterioration in the same.

The following table can be derived from the above.

Table 10 Character-wise improvement and deterioration after Markov process under W2

	Number of nodes where present	Number of nodes where improved	Number of nodes where deteriorated	Improved nodes : deteriorated nodes
Scheduled Tribe	4	3	1	3
Other Castes	4	1	3	0.33
Females	4	3	1	3
Males	4	1	3	0.33
Rural	4	2	2	1
Urban	4	2	2	1

It appears that under W that is under the decentralized circular structure the inclusion status of the scheduled tribe community and the females will improve through applying the rules of genetic algorithm. In other words, it can be said that the existing development programmes can deliver better results through establishing greater interaction among the different ethnic groups and the males and females.

Alternative to W_1 and W2 we can think about a hypothetical network structure where the society takes the form of a centralized star structure – described as W_3. Thus in the hypothetical situation W_3 the mating pool is presented through the following matrix and the centralized star structure.

Table 11 Edge weight matrix under centralised star structure under W3

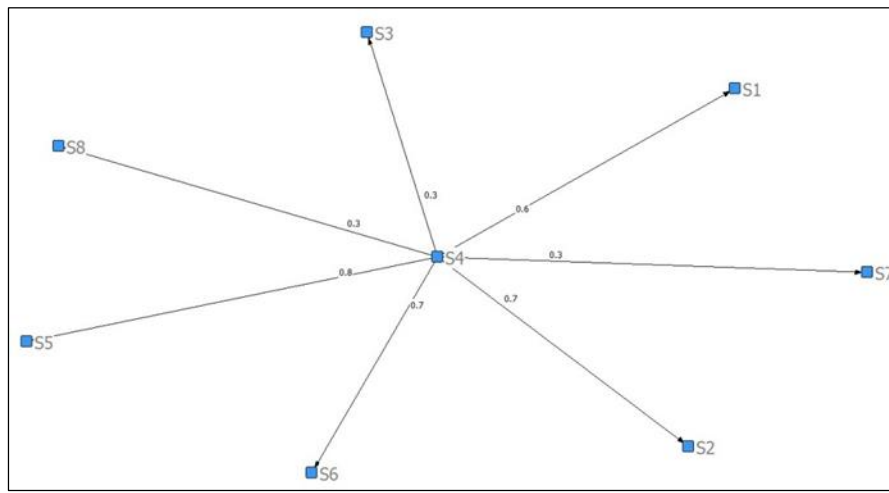|	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
S_1	-	-	-	0.5747	-	-	-	-
S_2	-	-	-	0.6887	-	-	-	-
S_3	-	-	-	0.3288	-	-	-	-
S_4	-	-	-	-	-	-	-	-
S_5	-	-	-	0.8234	-	-	-	-
S_6	-	-	-	0.6537	-	-	-	-
S_7	-	-	-	0.2589	-	-	-	-
S_8	-	-	-	0.2581	-	-	-	-

Figure 5 Star Structure

Thus the above depicted star structure is the initial mating pool P^0 under W_3. From this mating pool we can generate the stochastic transition matrix. But the initial distribution matrix under W_3 will be the same as under W_1.

This Markov process reaches the stable state after the 146 iterations. This means that

$Q^{146} = Q^{147}$

or $Q^{146} = Q^{146} \cdot T$

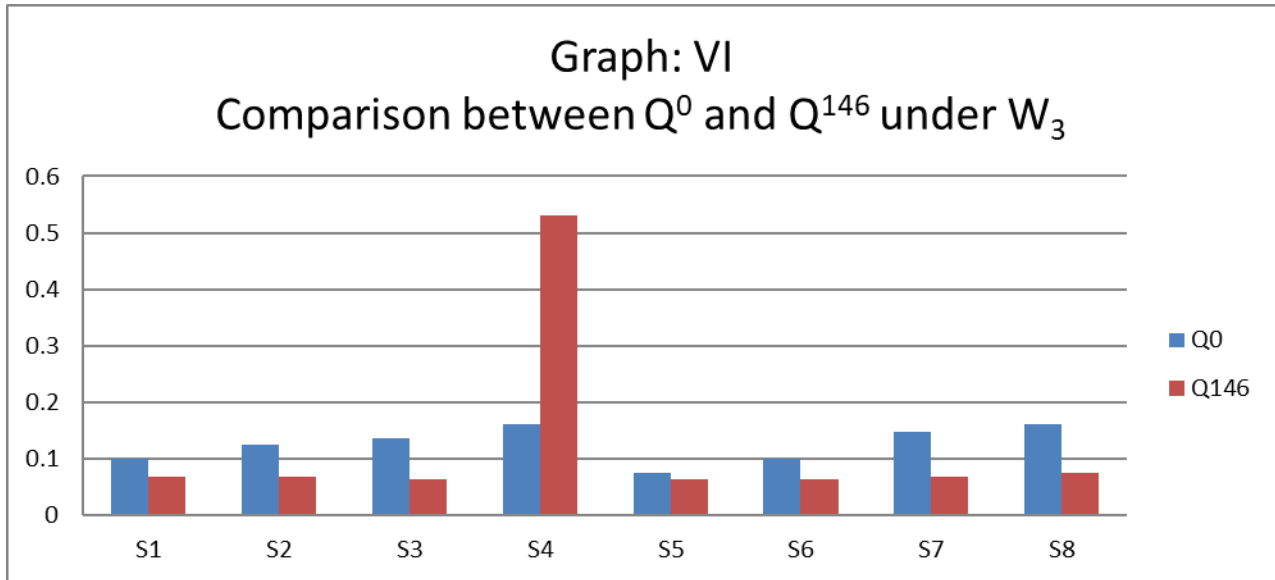Now comparing the Q^{146} with Q^0 we get the following diagram.

Figure 6 Comparison between Q0 and Q146 under W3

It appears from the above discussion that only the central node S_4 has registered spectacular improvement with respect to fitness while nodes S_1, S_2, S_3, S_5, S_6, S_7 and S_8 have registered deterioration in the same.

The following table can be derived from the above.

Table 13 Character-wise improvement and deterioration after Markov process under W3

	Number of nodes where present	Number of nodes where improved	Number of nodes where deteriorated	Improved nodes : deteriorated nodes
Scheduled Tribe	4	0	4	0
Other Castes	4	1	3	0.33
Females	4	0	4	0
Males	4	1	3	0.33
Rural	4	0	4	0
Urban	4	1	3	0.33

It appears that under W_3 that is under the centralized star structure the inclusion status of the other caste communities, the males and the urban dwellers will improve by applying the rules of genetic algorithm. The exclusion status of the existing excluded groups like the scheduled tribes, females and the rural peoples will deteriorate further. In other words, it can be said that the existing development programmes will extend the exclusionary processes further instead of allowing the excluded to participate.

Table 14 Comparison of relative fitness of chromosomes under initial state and stable state of W1, W2 and W3

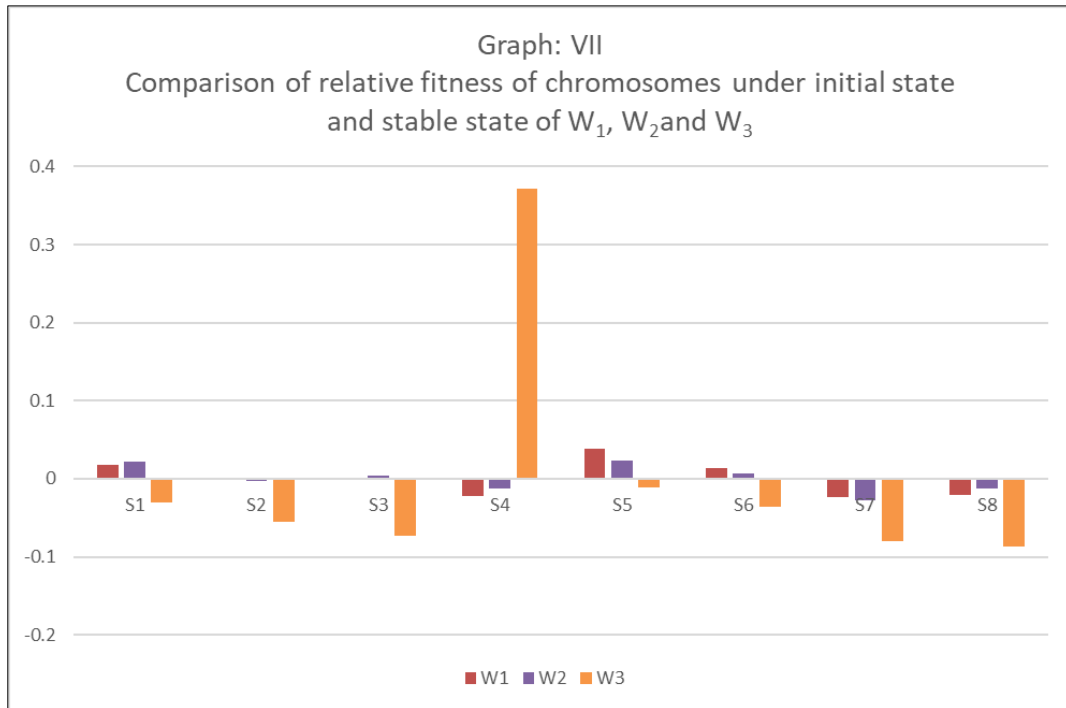|	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
W_1	0.017196	0.000691	-0.00074	-0.0227	0.037795	0.013104	-0.024	-0.02134
W_2	0.021857	-0.00283	0.004275	-0.01263	0.023202	0.006293	-0.02753	-0.01263
W_3	-0.03034	-0.05504	-0.07264	0.371085	-0.01092	-0.03561	-0.07973	-0.08681

Figure 7 Comparison of relative fitness of chromosomes under initial state and stable state of W1, W2and W3

Table 15 Comparison between W1, W2 and W3

Character	Improvement Ratio Under W_1	Improvement Ratio Under W_2	Improvement Ratio Under W_3
Scheduled Tribe	Very large	3	0
Other Castes	0	0.33	0.33
Females	1	3	0
Males	1	0.33	0.33
Rural	1	1	0
Urban	1	1	0.33
1 means no change in exclusion status. > 1 means improvement in exclusion status with highest value as infinitely large. <1 means deterioration in exclusion status with lowest value as 0.			

Table 16 Cohesion density and time to achieve stationarity under different network structures

	W_1	W_2	W_3
Cohesion density index	1	0.28	0.25
Shape	Complete network	Incomplete decentralised Hamiltonian circuit	Incomplete centralised star structure
Time for stationary	6 iterations	57 iterations	146 iterations

The correlation between the CDI and the path length l(p) is given below.

Table 17 Correlation Coefficient

	Complete Network	Incomplete decentralized Hamiltonian circuit	Incomplete centralized star network
CDI	1	0.28	0.25
l(p)	6	57	146
Coefficient Correlation	-0.79987		

It appears from the above discussion that improvement of cohesion density or inclusionary status varies largely with the change in social network conditions. Under the highly centralized network, none of the social characteristics demonstrated improvement with respect to the inclusionary status. On the other hand, under decentralized Hamiltonian structure the inclusionary status of two characteristics improved, that of two characteristics deteriorated and that of two characteristics remained the same. But the inclusionary status of the traditionally excluded like STs improved. The inclusionary status of the STs improved spectacularly under the complete network structure. Here the status of one characteristic deteriorated while that of four characteristics remained the same. Thus, with the increase in the cohesion density within the society the inclusionary status of the traditionally excluded castes improved. Again, the time to achieve stationery state improves with the improvement of cohesion density. Thus cohesion density is a major factor to achieve the desired outcomes under inclusionary development programmes. So, at the time of formation of inclusionary development programmes network cohesion should be given due importance and network specific inclusionary programmes should be developed. To get the desired outcome the plan formulators should primarily take steps to estimate the social network structure. After estimating the structure steps should be taken to improve cohesion within the society. State should play a proactive role to transform the society to the desired structure. It is true that the lack of cohesiveness within the Indian society has a long history and has the support of *Brahmanical* philosophical thoughts. To that respect the social outlook of the Hindu society should be changed. Peoples should embrace each other irrespective of caste, creed and economic status to achieve higher level of development with a given resource.

5. Conclusion

On the basis of the above discussion it can be concluded that the level of cohesion among different social groups can largely influence the expected outcome of any inclusionary development programme. This lack of cohesiveness within the Indian society due to some strong religio-philosophical reasons is one of the greatest challenges to India in achieving desired inclusive growth. In this paper it is shown that increasing the connectedness of the society can improve the efficiency of the development programs.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Abrahams, D. (2003). Building Hybrid Systems with Boost.Python.

[2] Bamshad, M., & al., e. (2001, May 8). Genetic Evidence on the Origins of Indian Caste Populations. Retrieved May 7, 2017, from Genome Research: www.genome.org

[3] Beauvais, C., & Jeason, J. (2002, May). Social Cohesion : Updating the State of the Research. CPRN Discussion paper .

[4] Bret Larget, D. L. (1999). Markov Chain Monte Carlo Algorithms for the Bayesian Analysis of Phylogenetic Trees. Molecular biology and evolution , 750-759.

[5] Bureau of Applied Economics and Statistics, Govt. of West Bengal. (2009). Statistical Abstract West Bengal 2008. Kolkata: Bureau of Applied Economics and Statistics, Govt. of West Bengal.

[6] Church, k. W., & Mercer, R. L. (1993). Introduction to the Special Issue on Computational Linguistics Using Large Coprora. Computational Linguistics , 19 (1), 1-24.

[7] Deotale, N. (2014). Automated Train using Genetic Algorithm for Railway Scheduling with GPS. International Journal of Current Engineering and Technology , 4 (3), 1553-1555.

[8] Ganguly, V. (2005). Displacement, Rehabilitation and Resettlement: The Case of Maldhari Families of Gir Forest . Sage Journal , 3-17.

[9] Horowitz, E. (2006). Fundamentals of Computer Algorithms. Galgotia Publications Pvt. Ltd.

[10] John A. Lewis, S. M. (1991). An empirical study of the object-oriented paradigm and software reuse. 91 Conference proceedings on Object-oriented programming systems, languages, and applications (pp. 184-196). New York: ACM SIGPLAN.

[11] K. Krishna, M. N. (June, 1999). Genetic K-Means Algorithm. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS , 433-439.

[12] Lee, D. G. (1989). Inheritance in Actor Based Concurrent Object-Oriented Languages. The Computer Journal , 297-304.

[13] Lieberherr, K. I. (1988). Object-oriented programming: An objective sense of style. ACM SIGPLAN Notices 23 , 232-234.

[14] Marmelstein, R. (1997). Application of Genetic Algorithms to Data Mining. Proceedings of 8th Midwest Artificial Intelligence and Cognitive Science Conference (pp. 58-65). AAAI.

[15] Martin Hitz, B. M. (1995). Measuring Coupling and Cohesion In Object-Oriented Systems. 25-27.

[16] Mat Marcus, J. J. (2007). Runtime Polymorphic Generic Programming—Mixing Objects and Concepts in ConceptC++. GPCE '07 Proceedings of the 6th international conference on Generative programming and component engineering , (pp. 73-82). New York.

[17] Max J. Egenhofer, A. U. (1992). Object-Oriented Modeling for GIS. URISA Journal 4 , 3-19.

[18] Mukherjee, P., & L.Roy. (n.d.). Lavels and Patterns of Social Cohesion and Its Relationship With Development in India. Centre for the Study of Regional Development .

[19] Oh, B. e. (2010). Genetic Algorithm-based Dynamic Vehicle Route Search using Car-to-Car Communication. Advances in Electrical and Computer Engineering , 81-86.

[20] Rama, N., & Lakshmanan, M. (2009). A New Computational Schema for Euphonic Conjunctions in Sanskrit Processing. International Journal of Computer Science Issues , 5, 43-51.

[21] Runjhun Tandon, N. T. (2016). Effect of seeding on sucrose polymorphism. Journal of Chemical and Pharmaceutical Research , 531-546.

[22] Saxena, S., & Agrawal, R. (2013). Sanskrit as a Programming Language and Natural Language Processing. Global Journal of Management and Business Studies , 3 (10), 1135-1142.

[23] Sengupta, S. (2015). "A New Generation of Optimization process using Genetic Algorithm. International Journal of Innovative Technology & Adaptive Managemen , 2 (7).

[24] Siek, J. G. (1995). Gradual Typing for Functional Languages. Journal of functional programming , 111-130.

[25] Smith, J. M. (2002). An Elemental Design Pattern Catalog.

[26] Su, R. e. (2014). Optimization of High-Speed Train Control Strategy for Traction Energy Saving Using an Improved Genetic Algorithm. Journal of Applied Mathematics , 2014, 7.

[27] Tang, M. (2015). Efficient Processing of Hamming-Distance-Based Similarity-Search Queries Over MapReduce. 18th International Conference on Extending Database Technology .

[28] The Task Force on Higher Education in Developing Countries. (2000). Higher Education in Developing Countries: Peril and Promise. Washington DC: The World Bank.

[29] Yee Leung, Y. G. (1997). Degree of Population Diversity—A Perspective on Premature Convergence in Genetic Algorithms and its Markov Chain Analysis. IEEE TRANSACTIONS ON NEURAL NETWORKS , 1165-1176.

Appendix – II

Determination of edge weight or inter nodal distance

To calculate the inter-nodal distance in the exclusion space or the edge weight we have tried to measure the level of social exclusion of each node or the representative chromosome. To develop an appropriate measure of social exclusion from development programmes we have mainly followed Chakraborty & D'Ambrosio(Chakraborty & D'Ambrosio, 2002), Camara *et.al*(Camara, Monteiro, Ramos, Sposati, & Koga, 2002),Poggi (Poggi, 2003), Bosset, D'Ambrosio & Peragine(Bossert, D'Ambrosio, & Peragine, 2004) and Acharya(Acharya, 2010). Chakraborty and D'Ambrosio have looked into social exclusion as functioning deprivation and tried to locate relevant functioning those have greater effect on the life. They first looked at functioning failures, the number of functioning from which the person is excluded. In this respect they marked certain indicators over a limited number of domains or functioning. They referred to it as the deprivation score or the number of exclusions of the person concerned. A person's exclusion in a given domain has been obtained by adding up his exclusions over the concerned variables. At the same time Camara *et.al* used the term dimension for domains or functioning. Like Chakraborty and D'Ambrosio, in the study of Camera *et. al.* also each dimension is captured by a set of variables. They have put 0 for attaining the desired level, 1 for over attainment and -1 for under attainment with respect to each variable. In the same line Poggi has also tried to define the functioning failures. His paper identified the socially excluded individual using Sen's capability approach. Like Camara *et.al,* Poggi also identified certain items under each functioning and for each selected item he assigned to each individual a score of zero or one. A score of one means that the individual can afford the item, has the item or does not have the 'problem'. Instead a score equal zero means that the individual is deprived in that item. He sum up the score of each item representing the same functioning and gave equal weights to the items. Then divided each functioning score by the number of items belonging to such functioning in order to be able to compare the distribution of different functionings. Thus for each functioning, an individual receives a score between zero and one. A score equal one means that the functioning has been fully achieved. Finally, he has used vector analysis for measuring the appropriate level of social exclusion. Bosset, D'Ambrosio and Peragine have said that Social exclusion manifests itself in the lack of an individual's access to functionings. GraphThey have also calculated social exclusion through the number of functioning failures. Their notion of social exclusion is obtained as an aggregate of the levels of deprivation experienced by an individual in each of the functioning. In a final step, these individual indicators of exclusion are aggregated across individuals to arrive at a class of measures of exclusion for society as a whole. In all cases, they use the arithmetic mean as the requisite aggregator function. Indicators of discrimination as discussed by Acharya tried to capture discrimination in different spheres, forms and personnel who may practice discrimination. Prevalence of discrimination was measured by simple percentage. Fixed scores were awarded to different levels of discrimination. The average score for each respondent was computed for sphere, form and provider separately – which may be called sectoral indexes. These average scores were used to compute a composite index of discrimination.

In reference to the above studies we have also constructed our own measure of social exclusion from development programmes in West Bengal. Our survey of literature within this study has found that the most of the studies tried to functionalise social exclusion through the notion of functioning or capability deprivation, such that the main domains of social exclusion according to those studies are health, education and income. The domains or aspects accepted by this work to measure social exclusion are also health, education and income. Under each of these domains certain development programmes have been chosen to functionalise the idea. Some questions or variables under each domain are put forwarded to capture the views of the respondents.

We have used dummy or binary variable to incorporate the views of the respondents into the model. 0 is assigned to the answer 'yes' for each question and 1 otherwise. The answer 'yes' or assigning 0 to any question means the respondent is not excluded with respect to the concerned variable. On the other hand answering 'no' or assigning 1 to any question means that the respondent is excluded with respect to the concerned variable. The score from each variable of each respondent for exclusion from income domain is added and divided by 5 to get the average (as income domain has five variables). This average value may be regarded as the measure of exclusion in the field of income delivery mechanism. Mathematically this may be presented as follows:

$$ME_j = \frac{1}{5}\sum_{i=1}^{5} e_{ij}^M$$

where ME_j is the measure of exclusion in the field of income delivery programme of jth individual. e_{ij}^M , i = 1,2 ....5 is the score on each variable under the indicator of income delivery programme of jth individual.

The measure for exclusion in the ground of health delivery programmes has two components with equal weights. These are measure of exclusion in the field of curative health and measure of exclusion in the field of preventive health.

The score from each variable of each respondent for exclusion from different variables under curative health care is added and divided by 7 to get the average (as the domain of curative health has 7 variables). This average value may be regarded as the measure of exclusion in the field of curative health delivery mechanism. Mathematically this may be presented as follows:

$$CHE_j = \frac{1}{7}\sum_{i=1}^{7} e_{ij}^{CH}$$

where CHE_j is the measure of exclusion in the field of government sponsored curative health delivery mechanism of jth individual. e_{ij}^{CH}, i = 1,2 ....7 is the score on each variable under the indicator of curative health delivery mechanism of jth individual.

Likewise, the score of each respondent for exclusion from three variables under preventive health care is added and divided by 3 to get the average. This average value may be regarded as the measure of exclusion in the field of government sponsored preventive health delivery mechanism. Mathematically this may be presented as follows:

$$PHE_j = \frac{1}{3}\sum_{i=1}^{3} e_{ij}^{PH}$$

where PHE_j is the measure of exclusion in the field of government sponsored preventive health delivery mechanism of jth individual. e_{ij}^{PH}, i = 1 .... 3 is the score on each variable under the indicator of preventive health delivery mechanism of jth individual.

Thus the composite measure of exclusion from Govt. sponsored health delivery programme is the average of curative health exclusion measure and preventive health exclusion measure having equal weight to each component. Mathematically,

$$HE_j = \frac{1}{2}[CHE_j + PHE_j]$$

$$\Rightarrow HE_j = \frac{1}{2}[\frac{1}{7}\sum_{i=1}^{7} e_{ij}^{CH} + \frac{1}{3}\sum_{i=1}^{3} e_{ij}^{PH}]$$

In the domain of education we have two variables and like the others the answer s are expressed in binary form.

$$EE_j = \frac{1}{2}\sum_{i=1}^{2} e_{ij}^{E}$$

where EE_j is the measure of exclusion in the field of government sponsored education delivery mechanism of jth individual. e_{ij}^{E}, i = 1 .... 3 is the score on each variable under the indicator of education delivery mechanism of jth individual.

Here it is to be kept in mind that voluntary exclusion from any programme has been treated as inclusion under the said programme.

The above discussion ensures that each Sectoral Index (ME_j, HE_j and EE_j) takes the values from 0 to 1 i.e., $0 \leq$ Sectoral Index ≤ 1. The higher the value of the sectoral index the higher will be the level of exclusion on that particular sector. If 3 dimensions of exclusion from government programmes are considered, then a composite measure will be represented by a point Dj = (ME_j, HE_j and EE_j) on the 3 dimensional Cartesian space. In the 3 dimensional space, the point O = (0,0,0) represents the point indicating the best situation, representing no exclusion while the point I = (1,1,1) represents the highest level of exclusion. Then the measure of exclusion for jth individual is SE_j, ismeasured by the

normalized Euclidean distance of the point Di from the idealpoint 0= (0,0,0). The exact formula to calculate normalized Euclidean distance in an n dimension Cartesian space (Simmons, 1963)(Malik & Arora, 2010) is

$$\frac{1}{n}\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

In our three dimension space of ME_j, HE_j and EE_j the same can be written as

$$SE_j = \frac{1}{\sqrt{3}}\left[\sqrt{\left(ME_j - 0\right)^2 + \left(HE_j - 0\right)^2 + \left(EE_j - 0\right)^2}\right]$$

Household level data collected on the basis of primary level survey (Appendix – I) are used to find the social exclusion score of the representative household of each node.

To find the inter nodal distance in the exclusion space the same idea of normalized Euclidean Distance in the three dimensional Cartesian space has been used. Thus the inter nodal distance between two nodes in the exclusion space is

$$\frac{1}{\sqrt{3}}\left[\sqrt{\left(ME_i - ME_j\right)^2 + \left(HE_i - HE_j\right)^2 + \left(EE_i - EE_j\right)^2}\right]$$

Where i = 1,2,…,8 and j = 1,2,…,8.