Disentangling networks: defining and analyzing cohesive subgroups*

Joachim De Weerdt[†] and Dirk Van de gaer[‡]

December 2005

Abstract

A cohesive subgroup is a subset of a network that displays some minimal level of cohesion. We axiomatize two cohesion-based orderings of subgroups. We show how cohesive subgroups can be identified in a population and propose a bootstrap procedure to analyze how these subgroups differ from randomly formed subgroups. We apply our framework to the informal insurance network of a small, rural village in the Kagera Region of Tanzania. We find that the flow of remittances within the subgroups we identify is 8-10 times higher than average and identify kinship, geographic distance, clan membership, religious affiliation and wealth to be the most important correlates of membership.

Keywords: networks, cohesive subgroups, cohesion, informal insurance, set evaluation

JEL Classification: D710, O17

*We wish to acknowledge financial support from the FWO-Vlaanderen and the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Federal office for scientific, technical and cultural affairs, contract no.P5/21. Comments on an earlier draft by Stefan Dercon, Geert Dhaene, Roland Iwan Luttens and Arne Schollaert are gratefully acknowledged.

[†]CES, Katholieke Universiteit Leuven, Naamsestraat 69, B-3000 Leuven and EDI, P.O. Box 393, Bukoba, Kagera region, Tanzania.

[‡]Corresponding author. SHERPPA, FEB, Ghent University, Hoveniersberg 24, B-9000 Gent and CORE, 34 Voie du Romain Pays, B-1348 Louvain la Neuve, Belgium.

1 Introduction

Agents use their network to overcome problems of information and free riding (e.g., Coleman (1984), North (1990) and Greif (1993)). McMillan and Woodruff (1999) find that the emergence of a credit relation between two firms in Vietnam depends on the duration of the trading relationship between them and whether managers learned about their customer before trading with him through their business associates or not. Murgai, Winters, Sadoulet and de Janvry (2002) show how family ties will influence the formation of insurance groups by easing the flow of information and the enforcement of implicit contracts. If link formation depends on the characteristics of the agents, one can expect closely-knit subgroups to form within the population that are more homogeneous with respect to the relevant characteristics than randomly formed subgroups. For example, if ethnic origin is important in business networks, we may expect closely-knit subgroups to form along ethnic lines.

In this paper we analyze real actual insurance networks in which the characteristics of agents and pre-existing links are crucially important. The purpose of the present paper is to develop a procedure that allows us to define, identify and analyze cohesive subgroups on the basis of actual network data. Our focus on this level, which is in between of the village and the individual, is motivated by recent empirical and theoretical findings. First, empirically, the identification of cohesive subgroups is important, for instance to test the extent of insurance in village economies. Townsend (1994), Dercon and Krishnan (2003) and many others perform such tests at the village level. Townsend (1994, p. 541) argued that "kinship groups or networks among family and friends might provide a good, if not better, basis for testing the risk sharing theory". Indeed, Fafchamps and Lund (2003) find that mutual insurance takes place through networks of friends and relatives: they provide gifts and loans when a household is subject to adverse shocks. Dercon and De Weerdt (2005) provided evidence of the importance of the change in average consumption of the households to which one is linked for the change in own non food consumption. We search for cohesive subgroups of households that say they can rely on each other in case of adverse shocks and verify that, in the follow-up surveys, they did indeed have higher than average flows of remittances among each other. More importantly, we develop a procedure that allows us to see in which dimensions such groups differ from randomly formed subgroups and we discover several new dimensions in which cohesive subgroups differ from randomly formed subgroups. Second, recently important contributions to the theory of endogenous network formation have been made -see, e.g., Jackson and Wolinsky (1996), Bala and Goyal (2000), Kranton and Minehart (2001), Watts (2002), Jackson and Watts (2002), Ligon, Thomas and Worrall (2002) and Genicot and Ray (2003). These theories have been applied to insurance networks by Bloch, Genicot and Ray (2005) and Bramoullé and Kranton (2005). Both papers obtain equilibrium insurance networks that are not connected, such that full insurance at the village level is excluded. Naturally, then, the relevant focus for theories of risk sharing becomes subgroups of the village.

We present a mathematically tight definition of cohesive subgroups and cohesion. A cohesive subgroup is defined as a subset of the population displaying some minimal level of cohesion. To find a consistent measure of the latter, we use an axiomatic approach to characterize orderings of social structures of subgroups that reflect two well known views -see, e.g., Alba (1973). The first view maintains that the cohesion within a subset of the population depends only on the inside links: the links between the elements of that subset. We call this internal cohesion. The second view says that it depends on the balance between the "centripetal-centrifugal" dimension of subgroups: a comparison of the inside links to the inside-outside links, the links of the subgroup with the rest of the population. This is called relative cohesion. The density of a subgraph and Bock and Hussain's (1950) measure incorporate the first and second view, respectively. The density is the fraction of possible inside links that are actually realized. Bock and Hussain's measure equals the density, divided by the fraction of possible inside-outside links that are present in the social structure. In the sociological literature many other proposals have been made to define cohesive subgroups - see, e.g. Frank (1995) and Wasserman and Faust (1997) for a discussion. These criteria are for instance based on the idea that each individual should be able to reach the other in the subgroup in a maximal number of steps, or that each actor has to share a minimal number of interactions with others in the subgroup. Frank (1995) proposes a statistical criterion to divide the population into a given number of non overlapping subgroups based on the concentration of interaction within subgroups relative to the extend of interaction between subgroups. This is a relative view of cohesion, that differs from ours, because our cohesive subgroups can be overlapping and the number of cohesive subgroups that we find is not specified a priori.

Having defined a cohesive subgroup, we simply compute the level of cohesion for all possible subgroups in the population and maintain only those that have at least a specified minimal level of cohesion. We describe an algorithm for the density of the subgraph that drastically diminishes the number of computations necessary to identify all cohesive subgroups in a population. Next, we use a bootstrap technique and find that internally cohesive subgroups are more likely to receive gifts and loans from other households in the village when hit by an important shock than randomly formed subgroups. The same procedure is used to determine in which dimensions cohesive subgroups differ from randomly formed subgroups. The basic idea behind the bootstrap is simple. We compare the cohesive subgroups with what these subgroups look like (with respect to some variable) if their composition was randomly determined within the population. We apply this procedure to the informal mutual insurance network of a community in rural Tanzania. The network structure is based on a survey question in which people were asked to identify the people on who they can personally rely for help or that can rely on them for help. We find that cohesive subgroups seem to be formed along kinship ties, geographical proximity, clan membership, religious affiliation and wealth. They are quite homogeneous concerning the age of the household head and wealth. Poor households have difficulties to become a member of a cohesive subgroup. In addition, if they are a member, this subgroup will typically be one whose members have lower than average wealth and thus less insurance potential. Despite the presence of the informal insurance mechanism this suggests that poor households remain vulnerable to adverse shocks.

The empirical literature on network formation relies on parametric regression analysis. McMillan and Woodruff (1999) run regressions with as dependent variable the amount of trust in business transactions, measured by the proportion of the amount due that firms receive from their customers after the delivery of the goods. They include network variables, such as the way firms first learned about their customers, as explanatory variables. Murgai et al. (2002) similarly add network variables to explain the probability that two households are linked in a probit specification. Contrary to these approaches, our approach is non parametric and allows us to identify cohesive subgroups in a clear, systematic way rather than through visual inspection of the data.

The advantage of our axiomatic approach over previous contributions is that judgments about cohesion are made transparent and theoretically consistent.[1] To obtain the axiomatizations, the problem is formulated as a problem of set evaluation, familiar from social choice theory -see, e.g., the survey by Barberà, Bossert and Pattanaik (2001). A recent contribution by Pattanaik and Xu (2002) axiomatizes measures of personal connections, taking indirect links into account. We do not take such indirect links into account for two reasons. First, due to the way our question was formulated, it is very likely that people also mention those links that have never been used before, but can be relied on nevertheless. Such links are probably 'indirect links' in the network structure obtained on the basis of really provided help. By focussing on the potential links, we take the relevant indirect links into account. Second, more generally, accounting for indirect links is definitely relevant for the individual. Our purpose, however, is to identify and analyze cohesive subgroups. For such an exercise, accounting for indirect links will not make a substantial difference.

In section 2 we develop the axiomatic approach to the measurement of cohesion and define cohesive subgroups. Section 3 describes the procedure to identify cohesive subgroups and the bootstrap procedure. We apply our ideas to the informal mutual insurance network of a community in rural Tanzania. Section 4 concludes.

[1] See, e.g., Thomson (2001) for a discussion of the axiomatic approach.

2 Defining cohesive subgroups

2.1 Notation and axioms

We use the following representation of the social structure in a population. Let the set of individuals be $N = \{1, \ldots, n\}$ and Σ be the set of all non empty subsets of N that contain at least two individuals. Take $S \in \Sigma$. $\Omega_S = \{\{x, y\} \mid x, y \in S, x \neq y\}$. Since we interpret $\{x, y\}$ to mean "household x and y are linked", Ω_S is the set of all possible links between the individuals in S. Γ_S denotes the set of all subsets of Ω_S. Pairs of households in S that are actually linked are in A_S; this set is the set of inside links.[2]

As stated in the introduction, arguments can be made to let the amount of cohesion in S also depend on the strength of the links between members of S and the rest of the population. Such links are called inside-outside links. Let $S^c = N \setminus S$. To capture the inside-outside links, we define $\Omega_{SS^c} = \{\{x, y\} \mid x \in S, y \in S^c\}$. The set of all subsets of Ω_{SS^c} is denoted by Γ_{SS^c}. The social structure between S and S^c will be given by $A_{SS^c} \in \Gamma_{SS^c}$. We can now define Ξ_N, the set of all possible social structures of subgroups in the population N. It consiss of S, a subset of N, A_S, the set of links they have among each other, and A_{SS^c}, the set of links they have with the rest of the population.

Definition 1 *: Set of possible social structures of subgroups in N:*
$\Xi_N = \{(S, A_S, A_{SS^c}) \mid S \in \Sigma, A_S \in \Gamma_S, A_{SS^c} \in \Gamma_{SS^c}\}$.

The problem of the evaluation of cohesion in $S \in \Sigma$ is a problem of the evaluation of $(S, A_S, A_{SS^c}) \in \Xi_N$. Let $\succeq$ be a complete, reflexive and transitive binary relation defined over $\Xi_N \times \Xi_N$. For all (S, A_S, A_{SS^c}) and $(S', B_{S'}, B_{S'S'^c})$ in Ξ_N, we have that $(S, A_S, A_{SS^c}) \succeq (S', B_{S'}, B_{S'S'^c})$ means that "S has at least as much cohesion as S'"[3]. The binary relations $\succ$ and $\sim$ denote the asymmetric and symmetric factors of $\succeq$, respectively. $(S, A_S, A_{SS^c}) \succ (S', B_{S'}, B_{S'S'^c})$ denotes "S has more cohesion than S'", while $(S, A_S, A_{SS^c}) \sim (S', B_{S'}, B_{S'S'^c})$ denotes "S has as much cohesion as S'".

A first axiom that we want our ordering to satisfy requires that adding a new link to the inside links does not change the ranking of social structures that have all inside-outside links in common. This is part (1) in axiom I. Part (2) states that adding a new inside-outside link does not change the ranking of social structures that have all inside links in common.

Axiom 1 (I) *: Independence:*
$\forall (S, A_S, A_{SS^c}), (S, B_S, A_{SS^c}),$

[2] Our analysis deals with non-directional links. The entire framework can be easily adjusted to deal with directional links, however.

[3] It should be clear that subgroups can also be ordered on the basis of other concerns. The framework can be easily accommodated to derive orderings based on these concerns.

$(S, A_S \cup \{\{x,y\}\}, A_{SS^c}), (S, B_S \cup \{\{x,y\}\}, A_{SS^c}),$
$(S, A_S, A_{SS^c} \cup \{\{x,v\}\}), (S, A_S, B_{SS^c} \cup \{\{x,v\}\}) \in \Xi_N :$
(1) if $\{x,y\} \in [\Omega_S \setminus (A_S \cup B_S)] : (S, A_S, A_{SS^c}) \succeq (S, B_S, A_{SS^c}) \Longleftrightarrow$
$(S, A_S \cup \{\{x,y\}\}, A_{SS^c}) \succeq (S, B_S \cup \{\{x,y\}\}, A_{SS^c}),$
(2) if $\{x,v\} \in [\Omega_{SS^c} \setminus (A_{SS^c} \cup B_{SS^c})] : (S, A_S, A_{SS^c}) \succeq (S, A_S, B_{SS^c}) \Longleftrightarrow$
$(S, A_S, A_{SS^c} \cup \{\{x,y\}\}) \succeq (S, A_S, B_{SS^c} \cup \{\{x,y\}\})$.

Anonymity requirements constitute a second axiom. We impose very weak anonymity requirements: we only require anonymity with respect to a particular type of link if there is only one such link present.

Axiom 2 (A) *: Anonymity:*
(1) $\forall (S, \{\{x,y\}\}, A_{SS^c}), (S, \{\{v,w\}\}, A_{SS^c}) \in \Xi_N :$
$(S, \{\{x,y\}\}, A_{SS^c}) \sim (S, \{\{v,w\}\}, A_{SS^c}),$
(2) $\forall (S, A_S, \{\{x,y\}\}), (S, A_S, \{\{v,w\}\}) \in \Xi_N :$
$(S, A_S, \{\{x,y\}\}) \sim (S, A_S, \{\{v,w\}\})$.

While the anonymity axiom is very weak, it becomes quite strong when it is combined with independence. To see this, start from the statement in A(1). Now add $\{s,t\}$ to the internal links. From I(1), $(S, \{\{x,y\}, \{s,t\}\}, A_{SS^c}) \sim (S, \{\{v,w\}, \{s,t\}\}, A_{SS^c})$. Hence, any set of internal links with two elements that contains $\{s,t\}$ must be equally cohesive. Similarly, from $(S, \{\{s,t\}\}, A_{SS^c}) \sim (S, \{\{q,r\}\}, A_{SS^c})$, because of I(1), $(S, \{\{s,t\}, \{x,y\}\}, A_{SS^c}) \sim (S, \{\{q,r\}, \{x,y\}\}, A_{SS^c})$. Combining this with the previous indifference, we get $(S, \{\{q,r\}, \{x,y\}\}, A_{SS^c}) \sim (S, \{\{v,w\}, \{s,t\}\}, A_{SS^c})$, or any social structure within S with two inside links is equally cohesive. Continuing the argument, it is possible to show that A(1) combined with I(1) implies that, keeping A_{SS^c} fixed, all sets of internal links having the same number of elements are equally internally cohesive.

Social cohesion depends on the links between the members of S. We say that a social structure with one internal link is more cohesive than a structure without internal links.

Axiom 3 (C) *: Cohesion:*
$\forall (S, \{\{x,y\}\}, A_{SS^c}) \in \Xi_N : (S, \{\{x,y\}\}, A_{SS^c}) \succ (S, \varnothing, A_{SS^c})$.

This cohesion axiom is very weak. However, combined with independence, it becomes quite strong. C implies that $(S, \{\{v,w\}\}, A_{SS^c}) \succ (S, \varnothing, A_{SS^c})$. Add $\{x,y\}$ to $\{\{v,w\}\}$ and $\varnothing$, to obtain, because of I(1), $(S, \{\{v,w\}, \{x,y\}\}, A_{SS^c}) \succ (S, \{\{x,y\}\}, A_{SS^c})$, while because of C $(S, \{\{x,y\}\}, A_{SS^c}) \succ (S, \varnothing, A_{SS^c})$. Actually, combined with the independence axiom, C implies that, if we add new links to a given set of internal links, the amount of cohesion always increases.

We distinguish between two types of cohesion. The first type says that cohesion only depends on the number of links within the set A_S. Because of independence, it suffices to require only that any situation with one inside-outside link has as much cohesion as the situation with no inside-outside links.

Axiom 4 (IIOL) *: Irrelevance of Inside-Outside Links:*
$\forall (S, A_S, \{\{x,y\}\}) \in \Xi_N : (S, A_S, \{\{x,y\}\}) \sim (S, A_S, \varnothing)$.

For the second view of cohesion, we allow inside-outside links to be relevant. We impose the following consistency axiom.

Axiom 5 (CON) *: Consistency:*
$\forall (S, A_S, A_{SS^c}), (S, A_S \cup B_S, A_{SS^c} \cup B_{SS^c}), (S, B_S, B_{SS^c}) \in \Xi_N$:
if $A_S \cap B_S = \varnothing$ *and* $A_{SS^c} \cap B_{SS^c} = \varnothing$, *then* $(S, A_S, A_{SS^c}) \sim (S, B_S, B_{SS^c})$
$\Rightarrow (S, A_S, A_{SS^c}) \sim (S, A_S \cup B_S, A_{SS^c} \cup B_{SS^c}) \sim (S, B_S, B_{SS^c})$.

Axiom CON tells us that, if we are indifferent between two social structures defined over S that have no inside or inside-outside links in common, we should be indifferent between those two social structures and the union of both structures.[4] This axiom ensures consistency of the way inside and inside-outside links are traded off against each other

We can now prove the following results[5].

Theorem 1 *: Orderings of subgroups based on the number of inside links:*
An ordering of social structures satisfies A,I, C and IIOL if and only if it is such that $\forall (S, A_S, A_{SS^c}), (S, B_S, B_{SS^c}) \in \Xi_N$:
$(S, A_S, A_{SS^c}) \succeq (S, B_S, B_{SS^c}) \Leftrightarrow |A_S| \geq |B_S|$.

Theorem 2 *: Ordering of subgroups based on the relative number of inside links:*
An ordering of social structures satisfies A, I, C and CON if and only if it is such that $\forall (S, A_S, A_{SS^c}), (S, B_S, B_{SS^c}) \in \Xi_N$:
$(S, A_S, A_{SS^c}) \succeq (S, B_S, B_{SS^c}) \Leftrightarrow \frac{|A_S|}{|A_{SS^c}|} \geq \frac{|B_S|}{|B_{SS^c}|}$.

Once the axioms A, I and C have been imposed, we can either impose that inside-outside links do not matter (IIOL), or that inside and inside-outside links have to be traded off against one another consistently (CON). If we impose the former, we get the ordering of theorem 1, if we impose the latter, we get the ordering of theorem 2. In both theorems, only the cardinalities of the sets of links matter. Note that, when CON is imposed, a higher number of inside-outside links makes a given set of internal links less cohesive. The resulting ordering ranks sets of links on the basis of the balance between inside and inside-outside links; this is the notion of relative cohesion.

[4] A more demanding version of this kind of axiom in a very different context has been used by, e.g., Van Hees (1998) -see his axiom C.2, p.180.

[5] Appendix 1 proofs these results and establishes the independence of the axioms.

2.2 Cohesion and cohesive subgroups

A natural way to cardinalize the ordering in theorem 1 is through an index proportional to the cardinality of the set of internal links, such that, if A_S has twice the number of links that B_S has, social cohesion in A_S is twice the social cohesion in B_S. A further requirement that can be imposed is that all complete sets of internal links ($A_S = \Omega_S$) are equally cohesive, irrespective of the (size of) the subsets of the population whose internal links are compared. This leads to $C_P\left(|A_S|, |\Omega_S|\right)$, defined in definition 2. To cardinalize the ordering in theorem 2, we follow a similar reasoning. Our cardinalization is such that it satisfies three properties. First, if $A_{SS^c} = B_{SS^c}$ and A_S has twice the number of links that B_S has, social cohesion in (S, A_S, A_{SS^c}) is twice the social cohesion in (S, B_S, B_{SS^c}). Second, if $A_S = B_S$ and A_{SS^C} has twice the number of links that B_{SS^C} has, social cohesion in (S, A_S, A_{SS^c}) is half the social cohesion in (S, B_S, B_{SS^c}). Third, all social structures with complete sets of inside and inside-outside links ($A_S = \Omega_S$ and $A_{SS^c} = \Omega_{SS^c}$) are equally cohesive, irrespective of the (size of) the subsets of the population whose social cohesion we want to evaluate. This results in $C_R\left(\frac{|A_S|}{|\Omega_S|}, \frac{|A_{SS^c}|}{|\Omega_{SS^c}|}\right)$, also defined in definition 2.

Definition 2 *: Measures of cohesion :*

(a) $C_P\left(|A_S|, |\Omega_S|\right) = \frac{|A_S|}{|\Omega_S|} = \frac{|A_S|}{\frac{1}{2}|S|[|S|-1]}$.

(b) $C_R\left(\frac{|A_S|}{|\Omega_S|}, \frac{|A_{SS^c}|}{|\Omega_{SS^c}|}\right) = \frac{\frac{|A_S|}{|S|[|S|-1]}}{\frac{|A_{SS^c}|}{|S||S^c|}}$.

Measure $C_P\left(|A_S|, |\Omega_S|\right)$ gives us the probability that a randomly chosen individual x in S will have a link with another randomly chosen individual in S. It is well known in the network literature as the density (see, e.g., Wasserman and Faust (1997, p.101-103)). $C_R\left(\frac{|A_S|}{|\Omega_S|}, \frac{|A_{SS^c}|}{|\Omega_{SS^c}|}\right)$ is the proportion of the possible links in S that are actually present in the social structure, divided by the proportion of possible links between S and S^c that are present in the social structure. Measure $C_R\left(\frac{|A_S|}{|\Omega_S|}, \frac{|A_{SS^c}|}{|\Omega_{SS^c}|}\right)$ was first proposed in the literature by Bock and Husain (1950). Alba (1973) points out that its numerator $\left(C_P\left(|A_S|, |\Omega_S|\right)\right)$ is a measure of the cohesiveness of a subgroup, while the denominator is a measure of the frequency of ties to individuals outside the subgroup.

The value $C_P\left(|A_S|, |\Omega_S|\right)$ lies between zero and one. It is equal to zero only when there are no inside links and is equal to one when all elements of S are linked to each other. This measure of cohesion can be additively decomposed[6]: the cohesion in a union of subgroups can be written as a weighted sum of the cohesion in the subgroups. Keeping the social structure between the subgroups and the subgroups and the rest of the population constant, we have that, if the

[6]Similar decomposition properties are frequently used in the measurement of inequality -see, e.g., Bourguignon (1979), Shorrocks (1984) or Cowell (2000).

cohesion in one of the subgroups of the union increases, then the cohesion of the union increases. Note that the sum of the weights attached to the subgroups is smaller than unity. The reason for this is that the number of possible inside links in the union of sets is larger than the sum of the potential number of inside links in the sets. As a consequence, the amount of cohesion in the union of sets will always be smaller than the amount of cohesion in the subset with the highest amount of cohesion if there are no links between the subsets. Measure $C_R\left(\frac{|A_S|}{|\Omega_S|}, \frac{|A_{SS^c}|}{|\Omega_{SS^c}|}\right)$ always has a positive value. It is equal to zero when there are no inside-outside links, is equal to one when all possible inside and inside-outside links are present and becomes infinite when a set has no inside-outside links. The latter is perhaps a drawback of this measure.

We can now provide a formal definition of cohesive subgroups. We define P_l^τ and R_l^k, the sets of cohesive subgroups, as the sets of subsets of N with $|S| \geq l$ and minimal levels of within subset cohesion τ and relative cohesion κ, respectively:

Definition 3 *: cohesive subgroups:*
$\forall S \in \Sigma$ *with* $|S| \geq l$:
(a) $S \in P_l^\tau \Leftrightarrow C_P\left(|A_S|, |\Omega_S|\right) \geq \tau$ *with* $\tau \in [0,1]$.
(b) $S \in R_l^\kappa \Leftrightarrow C_R\left(\frac{|A_S|}{|\Omega_S|}, \frac{|A_{SS^c}|}{|\Omega_{SS^c}|}\right) \geq \kappa$ *with* $\kappa \in [0,\infty]$.

The definition tells us, for different values of τ and κ, which subsets of N are cohesive. The cut-off values for τ and κ and the value of l have to be determined by the researcher. Practical considerations can play a role here. For instance, when we determine all the P_l^τ cohesive subgroups in our village, we exclude subsets with only two or three members since the data contain too many triplets with complete social structure.

3 Identifying and analyzing cohesive subgroups

3.1 Methodology

As a first step in our analysis, we illustrate the meaning of our measures of cohesion by dividing the population into mutually exclusive subgroups on the basis of several factors that will later be used to analyze cohesive subgroups. We calculate the level of cohesion in these subgroups. This also provides a way to introduce the data and the social structure in the village studied.

Next, we develop a procedure that identifies cohesive subgroups in a population. The bulldozer approach to finding these cohesive subgroups is to actually form all possible subsets S, calculate the measures of cohesion for each of them and retain only those subsets that have values above a certain threshold (τ or κ). For populations of some size, this requires going through many combinations of households and calculating C_P or C_R for each of them, a task which is impossible to complete within a reasonable time span on an ordinary PC. Fortunately, to identify P_l^τ, we can use an algorithm that drastically reduces the number of subsets to consider. It exploits the fact, shown in appendix 2, that a subgroup of size $|S|$ with $C_P \geq \tau$ will always be a superset of a subgroup of size $|S|-1$ with $C_P \geq \tau$. It is straightforward to identify all pairs with $C_P \geq \tau$: for $\tau > 0$ this will simply be all linked households. To find the combinations that could potentially be triplets with $C_P \geq \tau$ we need only consider those subgroups which add a third element to these pairs. The procedure is repeated to find subgroups with four, five and more members. We apply the algorithm to identify all elements of $P_4^{0.75}$ in the mutual insurance network of Nyakatoke. Unfortunately we have not found a similar algorithm to determine subgroups based on C_R. We therefore opt for a partial analysis of C_R-based subgroups: we restrict the analysis to search through the combinations having three members.

To see whether cohesive subgroups are better insured against bad shocks, we used follow-up data on actual transfers that flowed between cohesive subgroup members. We make a one-to-one mapping of the original households of the network onto new, random households (e.g., in our application household 5 becomes household 27 in all the subgroups, etc...) and calculate the flow of transfers within the randomly composed subgroups. We repeat this randomization R times (R will always be 1000 in this paper). Finally, we compare the results of these R randomizations with the results obtained for the actual cohesive subgroups. Note that after the mapping exercise the data retain the original structure. For example, the same kind of overlap between the subgroups will exist before and after the mapping. The structure will only be generated by different households. This bootstrap procedure enables us to test for the significance of the different amounts of remittances between cohesive subgroups and randomly formed subgroups.

We next turn to the characteristics of cohesive subgroups. We would like to know which households form cohesive subgroups. We follow two approaches. First we compare the characteristics of those that are a member of a cohesive subgroup with the characteristics of those that are not. Second, we use the bootstrap method explained above. We first calculate what the cohesive subgroups look like with respect to some variable. Next, we make a one-to-one mapping of the original households onto new, random households and calculate what the subgroups look like with respect to the variable under consideration. We repeat this randomization R (=1000) times and compare the results of these R randomizations with the results obtained from the actual cohesive subgroups. Compared to the probit model used by, e.g., Murgai et al (2002), the procedure has the advantages that no parametric assumptions have to be made. The disadvantage is that the effects are not controlled for the influence of other characteristics.

3.2 Data

We use the Nyakatoke Household Survey to illustrate how our measures of cohesion and our definition of cohesive subgroups can be used. The survey was administered in Nyakatoke, a Haya community in the Bukoba Rural District of the Kagera Region of Tanzania. In February 2000, all the 119 households in the community were interviewed. One of the questions asked to all adult individuals in the community was:

"Can you give a list of people from inside or outside of Nyakatoke, who you can personally rely on for help and/or that can rely on you for help (in cash, kind or labor)".

The answers to this question allow us to get a picture of the structure of the mutual insurance network inside the village. The 224 respondents listed a total of 1,126 network partners. About two thirds of them (738 in total) live in Nyakatoke. In what follows we only use the 738 intra-village links. To simplify the computation and analysis of the cohesive subgroups, we aggregate to household level, which leaves us with 490 links between 119 households.[7] The relevance of the network to smooth consumption and expenditure is likeli to be significant because formal savings and insurance institutions are absent in this area. When probed for the strategies they depended on in response to the two most important adverse shocks of the past 10 years, mutual insurance was most frequently mentioned (De Weerdt, 2002). Understanding mutual insurance networks is key to gaining insights into why households in developing countries are vulnerable to shocks.

In a follow-up survey the flow of actual remittances within the village was tracked for a year. We found that the bulk of these remittances (72% in terms of total value) was between members of our subgroup, while they constitute only 14% of the area of the network graph.

The formation of links in a mutual insurance context is endogenous to numerous factors related to smooth information flows, norms, trust, the ability to punish, discount rates and the potential gains of cooperation (Platteau (1991), Fafchamps (1992), Coate and Ravallion (1993)). For each household, we have information about several factors that are relevant for the formation of voluntary insurance networks. We know the age of the household head, and the value of the household's livestock and land holdings. The latter are good measures of wealth in Haya society (Reining (1967)). We have information on several important categorical variables such as the clan to which the household belongs, its religious

[7] Some links were bilaterally mentioned, but others (350 in total) were mentioned by only one of the two households. A unilaterally mentioned link should not be taken to mean that there are false expectations or the relation is not reciprocal. Remember that the question was framed as "who can you rely on and/or who can rely on you", so no directional meaning can be attached to it. We have repeated all computations and statistics for an alternative set-up, in which a bilaterally mentioned link is counted twice, while a unilaterally mentioned is counted only once. This makes no difference to the results.

affiliation and whether or not the household head completed primary education. To limit the size of our report of the bootstrap results, each of these categorical variables is given a numerical representation.

To investigate whether cohesive subgroups are more homogeneous than randomly formed subgroups, we use a measure which Taylor, Lewis and Hudson (1972) originally proposed to quantify ethno-linguistic fractionalization (ELF). Adapted to our set-up, it represents the probability that two households, drawn randomly without replacement from the subgroup, will belong to a different category. If there are G categories for a particular categorical variable, we get

$ELF = 1 - \sum_{i=1}^{G} \left(\frac{|s_i|}{|S|}\right)\left(\frac{|s_i|-1}{|S|}\right)$,

where $|s_i|$ is the number of households in category i and $|S|$ is the number of households in the subgroup. The lower this index, the more homogeneous the subgroup is with respect to the categorical variable under consideration.

Geographical proximity might stimulate the formation of insurance links. Haya homesteads are typically surrounded by dense fields of banana trees, intercropped with other crops. These form a natural barrier with the household's closest neighbor, who can only be reached by a short walk through the field or along a path. Houses are never built adjacent to each other. All the homesteads were plotted on an electronic map and the distance between each pair was calculated. These data allow us to calculate the average distance between the members of each subgroup.

Data are also available on all the kinship ties between households in the village. These can be used to calculate a kinship index of a subgroup as the mean kinship index across all possible pairs of households in the subgroup. We use Wright's coefficient of relationship, which is the expected proportion of one individual's genes that are identical to the genes of a second individual (Wright, 1922).[8]

3.3 Cohesion in exogenously given subgroups

We split the village into mutually exclusive subgroups based on clan membership, religious affiliation, wealth, education, or age of the household head. The first columns of Table 1 present values for $|S|, |A_S|, |A_{SS^c}|$, denoting the number of members in each subgroup, the number of links they have between each other and the number of links they have with the rest of the population. This contains all the information we need to calculate C_P and C_R and to rank the subgroups in terms of their cohesiveness. Note that there is no exogenous subgroup which is unconnected to the outside, i.e. $|A_{SS^c}|$ is never zero.

[8] This coefficient is 0.5 for parents, children and siblings. It is 0.25 for uncles, aunts, grandparents and grandchildren. At 0.125 and below we cannot specify further, given the nature of the data we have. Therefore we give all other relations a value of 0.125. However, as a robustness check, we have also repeated the analysis by giving the latter category a value of .0625 and .03125. This makes no difference for the results.

Nyakatoke has 25 different clans and every household belongs to one of them. Table 1 gives C_P and C_R for the ten largest clans. The Muyego have no inside links and thus have $C_P = 0$ and $C_R = 0$. The Mumiro is the most cohesive clan with $C_P = .67$ and $C_R = 7.53$. The former means that within the Mumiro clan a randomly chosen household will be linked to another randomly chosen household from the same clan with a probability of .67. The latter means that the proportion of possible inside links present within the clan to the proportion of possible inside-outside links is 7.53.

There are three different religions in Nyakatoke: Lutheran, Catholic and Muslim. The Table shows that the group of Muslims is more cohesive than Lutherans or Catholics. Also, when we consider Lutherans and Catholics as a single subgroup (Christians), they are not as tightly-knit as the Muslims. There are only minor shifts in ranking when one uses C_R instead of C_P to rank the clan or religious subgroups. The fraction of possible outside links that actually occur, C_P/C_R does not vary that much from one subgroup to another. C_P varies much more among different clans.

For the other variables, the population was split into two groups, implying that both groups have the same number of actual and potential inside-outside links. The subgroup of households whose head completed primary education does not seem to be more cohesive than its counterpart. For continuous variables like the age of the household head and wealth, we artificially split the village into two groups of approximately equal size. The first group contains all values that are equal to or smaller than the median and the second contains all values that are strictly greater than the median. We see that households with older heads form a more cohesive subgroup than households with younger heads. The differences in cohesiveness between subgroups based on wealth are striking. Table 1 shows that the top half of the wealth distribution is more cohesive than the bottom half. The values of C_P tell us that two randomly chosen rich households are twice as likely to have a link among each other than two randomly chosen poor households.

3.4 Identifying cohesive subgroups

Due to the multitude of factors that influence the formation of cohesive subgroups, cohesive subgroups do not coincide with a collection of households delineated on the basis of just one factor like religion, clan or wealth. We apply the algorithm described in section 3.1. to identify all cohesive subgroups in the mutual insurance network of Nyakatoke. We say that a subgroup is cohesive if and only if $|S| \geq 4$ and $C_P \geq .75$.[9] Table 2 reports the results of this exercise. Cohesive subgroups have between four and eight members. The largest complete subgroups ($A_S = \Omega_S$) have five members. For $C_P = .83$ and $C_P = .80$ the number of cohesive

[9] τ has been set at .75, because below this value the number of cohesive subgroups explodes. By requiring $|S| \geq 4$, we exclude the 490 pairs and 315 triplets that have a complete social structure.

subgroups is huge. Note that by comparing the values for $|S|$ and $|A_S|$ one can see that all cohesive subgroups are connected: every member of a cohesive subgroup has at least one link with another member of the cohesive subgroup.

For computational reasons, we opt for a partial analysis of C_R-based subgroups: we restrict ourselves to search through the $2.74 * 10^5$ combinations of 3 members. Setting $\kappa = 20$, we find 48 cohesive subgroups. Out of these, 31 have a complete social structure (i.e. have three internal links), 15 have two internal links and two have only one internal link. Note that the latter are unconnected graphs.

When moving on to the statistical analysis of these cohesive subgroups, we check whether our results are robust for different values of τ. For this purpose we order the C_P based subgroups on the basis of their value for C_P, irrespective of their size. For $C_P = .86$ and $C_P = .93$ too few subgroups exist for any statistical analysis and they are therefore grouped together with the nearest value of C_P for which sufficient cohesive subgroups can be identified. This gives us a total of 8 C-groups. C-Group 1 contains the subgroups with highest C_P and C-Group 8 the subgroups with lowest C_P. The ninth C-group contains all 1889 cohesive subgroups together. Table 3 summarizes the nine C-groups and provides some data on their characteristics. Note that most C-groups, except C-Group 6, are quite homogenous with respect to the number of members they contain. The number of households that belong to at least one cohesive subgroup ranges from 18% in C-Group 5 to 88% in C-Group 4 and 9. Next, we define C-group 10 as all 48 subgroups of three members with $C_R \geq 20$ and C-group 11 as all 46 subgroups of three members with $C_R \geq 20$ and having at least two links (i.e. they are connected graphs). Finally, we define C-group 12 as all 315 subgroups of three members with $C_P \geq .75$ (these are all triplets with complete social structure). Table 3 shows that 92% of the population belongs to a subgroup from C-group 12. C-groups 10 and 11 each contain 55% of the households.

3.5 Cohesive subgroups and risk sharing

If cohesive subgroups matter in risk-sharing then we may expect a reasonable amount of remittances to be exchanged within them. Follow-up surveys in Nyakatoke mapped all remittances that occurred within the village in the course of a year. Figure 1 plots the different C-groups from Table 3 on the X-axis. The values of the actual (not randomized) average yearly flow of remittances within the subgroups are connected by straight lines. The vertical bars are confidence intervals for the randomization experiments: 90% of the 1000 randomization experiments yielded an average yearly flow of remittances on this line, 5% lay above the line and 5% below it. The square lying on the vertical line indicates the mean value of this variable across the 1000 randomizations. Figure 1 shows that network traffic is 8 to 10 times higher within the subgroups than outside of them. Subgroups also matter at aggregate level. Around 72% of total network traffic in terms of gifts occurred within the subgroups we identified, while only 14% of all dyads are member of at least one subgroup.

3.6 Analyzing cohesive subgroups

Table 4 compares the characteristics of households that are included in a cohesive subgroup to those that are excluded. The differences in wealth (livestock and land) are striking. It seems that cohesive subgroups determined on the basis of C_P are mainly made up of rich households. Interestingly, exactly the opposite is true for subgroups based on C_R: in C-groups 10 and 11 the households that belong to cohesive subgroups are, on average, poorer than the others. This suggests that the poor have difficulties forming links, tend to be excluded from internally cohesive subgroups and belong to the relatively more isolated ones (i.e. those with few outside links). Strongly cohesive C_P-groups (groups 1-3) contain more households that completed primary education, while C_R-groups contain more households with lowly educated household heads. C_P-groups are composed of households that have older heads. The bootstrap results in Figure 2 confirm the results in Table 4. The different C-group numbers are plotted on the X-axis. The values of the actual (not randomized) average intra-subgroup variable for each C-group are connected by straight lines. The vertical bars are again confidence intervals for the randomization experiments, just like in the previous Figure. We can conclude that differences in wealth (livestock and land) and head's age are statistically significant. The head's education does not appear to be significant.

Figure 3 analyzes other characteristics of cohesive subgroups further. For kinship and geographical distance we calculate the intra-subgroup mean, i.e. the mean over all possible intra-subgroup pairs. From Figure 3.1, we see that the kinship index is much higher in the actual cohesive subgroups than in the randomized subgroups. Figure 3.2 shows that geographical distance too is very important, suggesting a strong neighborhood effect in subgroup formation. To see whether cohesive subgroups are more homogeneous than randomized subgroups with respect to the age of the household head, the value of the livestock and the size of the land holdings, we split the population in two groups of equal size. Next, we compare the ELF-value of the cohesive subgroups to the ELF-value of the randomized subgroups. Figures 3.3 to 3.5 contain the results. Cohesive groups are more homogeneous with respect to all three variables, although the wealth variables are significant at 10% only. Homogeneity with respect to wealth variables is quite worrying for poor households. If they are part of a cohesive subgroups, they tend to be part of relatively poor cohesive subgroups, subgroups with less insurance potential. Figure 3.8 finds no evidence that cohesive subgroups are homogeneous with respect to the head's education. To analyze the effect of the categorical variables (religion, clan and education) we compute the ELF-value for the actual and bootstrapped cohesive subgroups. Figures 3.6 and 3.7 show that subgroups are relatively homogeneous with respect to clan and religion.

4 Conclusions

In this paper we focus on subgroups of networks. As such, the perspective is in-between an individual and network perspective. Different criteria can be used to identify particular types of subgroups, such as internal cohesion and relative cohesion. To axiomatize internal and relative cohesion, we used a framework of set evaluation. Cohesive subgroups are subsets of the network that have a certain minimal level of cohesion. This leads to a coherent definition of cohesive subgroups and opens the possibility to identify cohesive subgroups systematically. Such an identification is in itself desirable for a variety of purposes. For instance, in line with a suggestion by Townsend (1994), our cohesive subgroups can be used as a basis for a test of the risk sharing hypothesis at the level of the cohesive subgroup.

Our measures of cohesion and our definition of cohesive subgroups were used to analyze the informal mutual insurance system in Nyakatoke, Tanzania. We found that close to three quarters of the remittances that occur within the village occur within these subgroups, while they constitute only 14% of the network graph. The formation of cohesive subgroups is not a trivial matter, but members of cohesive subgroups are likely to belong to the same clan, religion or family or live in each other's neighborhood. This confirms the hypothesis that these factors ease the flow of information and enhance the enforcement of informal contracts. It may also point at serious problems to use data from real networks to test predictions about network formation based on the assumption that households are identical such that the network structure is anonymous[10]. Fortunately, advances have been made to deal with network formation when players are heterogeneous - see, e.g. Galeotti and Goyal (2004). Most importantly, we found evidence that poor families have difficulties to form ties with other households, tend to be excluded from internally cohesive subgroups, belong to relatively more isolated subgroups and subgroups with lower wealth and thus less insurance potential. As a consequence, poor households will remain vulnerable to adverse shocks, despite the presence of the informal insurance system.

[10]Tests of the standard theories on the basis of experimental data is more straightforward, see, e.g., Falk and Kosfeld (2003) or Callander and Plott (2005).

Appendix 1: proof of the theorems

1.1. Proof of theorem 1

Lemma 1 *: Positive cardinality based ordering of inside links:*
An ordering of social structures satisfies A(1), I(1) and C,
$\Longleftrightarrow \forall \left(S, A_S, A_{SS^c}\right), \left(S, B_S, A_{SS^c}\right) \in \Xi_N :$
$\left(S, A_S, A_{SS^c}\right) \succeq \left(S, B_S, A_{SS^c}\right) \Longleftrightarrow |A_S| \geq |B_S|$

Proof. It is easy to verify that the positive cardinality based ordering satisfies A(1), I(1) and C. Hence we only prove the sufficiency part. Since S and A_{SS^c} are kept constant in all evaluations, the notation used in this proof suppresses them.

First consider $A_S, B_S \in \Gamma_S \setminus \varnothing$. Pattanaik and Xu (1990) -see also Barberà, Bossert and Pattanaik (2001, p.23-24)- use three axioms to characterize the positive cardinality based ordering over this domain. These three axioms are A(1) (labelled INC in their framework: indifference between no choice situations, I(1) (in their framework IND) and SM. The latter is defined as follows:

SM: Strong Monotonicity:

$\forall \{x, y\} \neq \{v, w\} \in \Omega_S : \{\{x, y\}, \{v, w\}\} \succ \{\{x, y\}\}$

SM is implied by C ($\{v, w\} \succ \varnothing$) and I(1) (adding $\{x, y\}$). Hence, over this domain the lemma follows directly from Pattanaik and Xu.

Second, consider $B_S = \varnothing, A_S \in \Gamma_S \backslash \varnothing$. Let $A_S = \left\{\{a_{11}, a_{12}\}, \ldots, \left\{a_{|A_S|1}, a_{|A_S|2}\right\}\right\}$. SM holds on this domain. This implies that

$\forall \{a_{i1}, a_{i2}\} \neq \{a_{11}, a_{12}\} : \{\{a_{11}, a_{12}\}, \{a_{i1}, a_{i2}\}\} \succ \{\{a_{11}, a_{12}\}\}$.

Start from this relationship with $i = 2$. Add $\{a_{31}, a_{32}\}$, to get from I(1)

$\{\{a_{11}, a_{12}\}, \{a_{21}, a_{22}\}, \{a_{31}, a_{32}\}\} \succeq \{\{a_{11}, a_{12}\}, \{a_{31}, a_{32}\}\}$

By SM, $\{\{a_{11}, a_{12}\}, \{a_{31}, a_{32}\}\} \succ \{\{a_{11}, a_{12}\}\}$. Transitivity of $\succeq$ then implies

$\{\{a_{11}, a_{12}\}, \{a_{21}, a_{22}\}, \{a_{31}, a_{32}\}\} \succ \{\{a_{11}, a_{12}\}\}$.

Add $\{a_{41}, a_{42}\}$ to this last relationship and application of I(1) and SM together with transitivity of $\succeq$ yields

$\{\{a_{11}, a_{12}\}, \{a_{21}, a_{22}\}, \{a_{31}, a_{32}\}, \{a_{41}, a_{42}\}\} \succ \{\{a_{11}, a_{12}\}\}$. Continue to add the elements of A_S until we get $A_S \succ \{\{a_{11}, a_{12}\}\}$, so that by virtue of C, $A_S \succ \varnothing = B_S$.

Third, if $B_S = A_S = \varnothing$, then the result $A_S \sim B_S$ follows immediately from reflexivity of $\sim$. ■

Lemma 2 *: An ordering of social structures satisfies IIOL and I(2) iff*
$\forall \left(S, A_S, A_{SS^c}\right) \in \Xi_N : \left(S, A_S, A_{SS^c}\right) \sim \left(S, A_S, \varnothing\right)$

Proof. Let $A_{SS^c} = \left\{\{a_{11}, a_{12}\}, \ldots, \left\{a_{|A_S|1}, a_{|A_S|2}\right\}\right\}$. By IIOL,
$(S, A_S, A_{S^c}, \{\{a_{11}, a_{12}\}\}) \sim (S, A_S, A_{S^c}, \varnothing)$.

Adding $\{a_{21}, a_{22}\}$ yields, because of I(2) and IIOL

$(S, A_S, A_{S^c}, \{\{a_{11}, a_{12}\}, \{a_{21}, a_{22}\}\}) \sim (S, A_S, A_{S^c}, \{\{a_{21}, a_{22}\}\}) \sim (S, A_S, A_{S^c}, \varnothing)$

Continue to expand the set at the left and right-hand side (I(2)) and using IIOL results in the lemma. ■

Proof of theorem 1. Take the social structure (S, B_S, B_{SS^c}). Apply lemma 2 twice to get

$(S, B_S, B_{SS^c}) \sim (S, B_S, \varnothing) \sim (S, B_S, A_{SS^c})$,

such that, by lemma 1

$(S, A_S, A_{SS^c}) \succeq (S, B_S, B_{SS^c}) \Leftrightarrow |A_S| \geq |B_S|$. ■

Proof of the independence of the axioms in theorem 1. (1) Consider ordering sets of internal links in S on the basis of a weighted average of the number of links within S, $\sum_{x \in S} \alpha(x) l(x)$, where $l(x)$ is the number of links that x has with other members of S, and $\alpha(x) > 0$ is the weight, which is different for at least two elements of S. This criterion satisfies all axioms, except A.

(2) Define $k(x) = 0$ if $l(x) = 0$ and $k(x) = \sum_{i=1}^{l(x)} \beta^i$ if $l(x) > 0$. Let $0 < \beta < 1$. Consider ranking sets of internal links on the basis of $\sum_{x \in S} k(x)$. This criterion satisfies all axioms, except I.

(3) Consider ordering sets of internal links in S on the basis of $-|A_S|$. This criterion satisfies all axioms, except C.

(4) Consider ordering sets of internal links in S on the basis of the criterion in theorem 2. This criterion satisfies all axioms, except IIOL. ■

1.2. Proof of theorem 2

Lemma 3 *: If an ordering of social structures satisfies A and I then*

$\forall (S, A_S, A_{SS^c}), (S, B_S, B_{SS^c}) \in \Xi_N$:

$|A_S| = |B_S|$ *and* $|A_{SS^c}| = |B_{SS^c}| \Rightarrow (S, A_S, A_{SS^c}) \sim (S, B_S, B_{SS^c})$.

Proof. There are four possibilities to consider.

(i) $A_S = B_S$ and $A_{SS^c} = B_{SS^c}$, then the result follows immediately from the reflexivity of $\sim$.

(ii) $A_S \neq B_S$ and $A_{SS^c} = B_{SS^c}$, then the result follows from A(1) and I(1).

The proof that $(S, B_S, A_{SS^c}) \sim (S, A_S, A_{SS^c})$ is similar to the proof of steps 1 and 2 in the proof of Pattanaik and Xu (1990). Only A(1) and I(1) are used in these parts of their proof.

(iii) $A_S = B_S$ and $A_{SS^c} \neq B_{SS^c}$, then the result follows from A(2) and I(2). See (ii) above.

(iv) $A_S \neq B_S$ and $A_{SS^c} \neq B_{SS^c}$, then we have, after applying first (ii) and then (iii) that,

$(S, B_S, B_{SS^c}) \sim (S, A_S, B_{SS^c}) \sim (S, A_S, A_{SS^c})$. ■

Lemma 4 *: If an ordering of social structures satisfies A, I and C, then*
$\forall \left(S, A_S, A_{SS^c}\right), \left(S, B_S, B_{SS^c}\right) \in \Xi_N :$
$|A_S| > |B_S|$ *and* $|A_{SS^c}| = |B_{SS^c}| \Rightarrow \left(S, A_S, A_{SS^c}\right) \succ \left(S, B_S, B_{SS^c}\right).$

Proof. Again, we distinguish two cases:
(i) $A_{SS^c} = B_{SS^c}$. From lemma 1 the result follows immediately;
(ii) $A_{SS^c} \neq B_{SS^c}$. First apply lemma 3, then lemma 1 to get
$(S, A_S, A_{SS^c}) \sim (S, A_S, B_{SS^c}) \succ (S, B_S, B_{SS^c})$. ■

Lemma 5 *: If an ordering of social structures satisfies A, I, and CON, then*
$\forall \left(S, A_S, A_{SS^c}\right), \left(S, B_S, B_{SS^c}\right) \in \Xi_N :$
$\frac{|A_S|}{|A_{SS^c}|} = \frac{|B_S|}{|B_{SS^c}|} \Rightarrow \left(S, A_S, A_{SS^c}\right) \sim \left(S, B_S, B_{SS^c}\right).$

Proof. Under the condition stated in the lemma, we can write
$|A_S| = (n+q)\,|B_S|$ and $|A_{SS^c}| = (n+q)\,|B_{SS^c}|$
with $n \in \mathcal{N}^+, q \in Q^+$ and $0 \leq q < 1$.
Case 1: $q = 0$.
Define n disjunct subsets of A_S and A_{SS^c} such that for all A_S^i and $A_{SS^c}^i, i = 1, \ldots n$ we have that $|A_S^i| = |B_S|\,, |A_{SS^c}^i| = |B_{SS^c}|$ and $A_S = \cup_{i=1}^n A_S^i$ and $A_{SS^c} = \cup_{i=1}^n A_{SS^c}^i$. Due to CON we have
$(S, A_S, A_{SS^c}) = (S, \cup_{i=1}^n A_S^i, \cup_{i=1}^n A_{SS^c}^i) \sim (S, A_S^i, A_{SS^c}^i)$.
Lemma 3 implies $(S, A_S^i, A_{SS^c}^i) \sim (S, B_S, B_{SS^c})$.
Case 2: $q > 0$.
Since $|A_S|$ and $|A_{SS^c}|$ are natural numbers, we must have that $q\,|B_S|$ and $q\,|B_{SS^c}|$ are also natural numbers. Let $k_1, k_2 \in \mathcal{N}^+$ be the two smallest numbers such that $q = \frac{k_1}{k_2}$. Thus we have that $\frac{k_1|B_S|}{k_2}$ and $\frac{k_1|B_{SS^c}|}{k_2}$ must be natural numbers. Due to the way k_1 and k_2 are constructed, this can only be the case if both $\frac{|B_S|}{k_2}$ and $\frac{|B_{SS^c}|}{k_2}$ are natural numbers.
First, define k_1 disjunct subsets of B_S and B_{SS^c} such that for all C_S^i and $C_{SS^c}^i$, $i = 1, ...k_1$ we have that $|C_S^i| = \frac{|B_S|}{k_2}, |C_{SS^c}^i| = \frac{|B_{SS^c}|}{k_2}$ and $B_S = \cup_{i=1}^{k_1} C_S^i$ and $B_{SS^c} = \cup_{i=1}^{k_1} C_{SS^c}^i$. Due to CON we have
$(S, B_S, B_{SS^c}) = \left(S, \cup_{i=1}^{k_1} C_S^i, \cup_{i=1}^{k_1} C_{SS^c}^i\right) \sim \left(S, C_S^i, C_{SS^c}^i\right).$
Second, note that
$|A_S| = (n+q)\,|B_S| = (nk_2 + k_1)\,\frac{|B_S|}{k_2}$
$|A_{SS^c}| = (n+q)\,|B_{SS^c}| = (nk_2 + k_1)\,\frac{|B_{SS^c}|}{k_2}$.
Define $nk_2 + k_1$ disjunct subsets of A_S and A_{SS^c} such that for all D_S^i and $D_{SS^c}^i, i = 1, \ldots nk_2 + k_1$ we have that $|D_S^i| = \frac{|B_S|}{k_2}, |D_{SS^c}^i| = \frac{|B_{SS^c}|}{k_2}$ and $A_S = \cup_{i=1}^{nk_2+k_1} D_S^i$ and $A_{SS^c} = \cup_{i=1}^{nk_2+k_1} D_{SS^c}^i$. Due to CON we have
$(S, A_S, A_{SS^c}) = \left(S, \cup_{i=1}^{nk_2+k_1} D_S^i, \cup_{i=1}^{nk_2+k_1} D_{SS^c}^i\right) \sim \left(S, D_S^i, D_{SS^c}^i\right).$
Finally, since $|D_S^i| = |C_S^i|$ and $|D_{SS^c}^i| = |C_{SS^c}^i|$ lemma 3 yields
$(S, C_S^i, C_{SS^c}^i) \sim (S, D_S^i, D_{SS^c}^i)$, such that $(S, A_S, A_{SS^c}) \sim (S, B_S, B_{SS^c})$. ■

Lemma 6 *If an ordering of social structures satisfies A, I, C and CON, then*
$\forall \left(S, A_S, A_{SS^c}\right), \left(S, B_S, B_{SS^c}\right) \in \Xi_N$:
$\frac{|A_S|}{|A_{SS^c}|} > \frac{|B_S|}{|B_{SS^c}|} \Rightarrow \left(S, A_S, A_{SS^c}\right) \succ \left(S, B_S, B_{SS^c}\right).$

Proof. Under the condition stated in the lemma, we can write
$|A_S| = (n_1 + q_1) |B_S|$
$|A_{SS^c}| = (n_2 + q_2) |B_{SS^c}|$
with $(n_1 + q_1) > (n_2 + q_2)$, $n_1, n_2 \in \mathcal{N}^+, q_1, q_2 \in Q^+$ and $0 \leq q_1, q_2 < 1$.
Suppose the ordering satisfies C. Define a social structure $\left(S, \widetilde{A}_S, A_{SS^c}\right)$ where $\left|\widetilde{A}_S\right| = (n_2 + q_2) |B_S|$. By lemma 4 we have
$\left(S, A_S, A_{SS^c}\right) \succ \left(S, \widetilde{A}_S, A_{SS^c}\right).$
From lemma 5 we have
$\left(S, \widetilde{A}_S, A_{SS^c}\right) \sim \left(S, B_S, B_{SS^c}\right).$ ■

Lemma 7 *If an ordering of social structures satisfies A, I, C and CON, then*
$\forall \left(S, A_S, A_{SS^c}\right), \left(S, B_S, B_{SS^c}\right) \in \Xi_N$:
$\frac{|A_S|}{|A_{SS^c}|} < \frac{|B_S|}{|B_{SS^c}|} \Rightarrow \left(S, B_S, B_{SS^c}\right) \succ \left(S, A_S, A_{SS^c}\right)$

Proof. Under the condition stated in the lemma, we can write
$|A_S| = (n_1 + q_1) |B_S|$
$|A_{SS^c}| = (n_2 + q_2) |B_{SS^c}|$
with $(n_1 + q_1) < (n_2 + q_2)$, $n_1, n_2 \in \mathcal{N}^+, q_1, q_2 \in Q^+$ and $0 \leq q_1, q_2 < 1$.
Suppose the evaluation of social structures satisfies C. Define a social structure $\left(S, \widetilde{A}_S, A_{SS^c}\right)$ where $\left|\widetilde{A}_S\right| = (n_2 + q_2) |B_S|$. By lemma 4 we have
$\left(S, \widetilde{A}_S, A_{SS^c}\right) \succ \left(S, A_S, A_{SS^c}\right).$
From lemma 5 we have
$\left(S, \widetilde{A}_S, A_{SS^c}\right) \sim \left(S, B_S, B_{SS^c}\right).$ ■

Proof of theorem 2. Sufficiency follows immediately from the previous lemmas 5, 6 and 7. Necessity can be shown by contradiction. Suppose, e.g. $\left(S, A_S, A_{SS^c}\right) \succ \left(S, B_S, B_{SS^c}\right)$, but $\frac{|A_S|}{|A_{SS^c}|} < \frac{|B_S|}{|B_{SS^c}|}$ or $\frac{|A_S|}{|A_{SS^c}|} = \frac{|B_S|}{|B_{SS^c}|}$. In the first case, $\left(S, A_S, A_{SS^c}\right) \prec \left(S, B_S, B_{SS^c}\right)$ (lemma 7). In the second case $\left(S, A_S, A_{SS^c}\right) \sim \left(S, B_S, B_{SS^c}\right)$ (lemma 5). Both contradict $\left(S, A_S, A_{SS^c}\right) \succ \left(S, B_S, B_{SS^c}\right)$. ■

Proof of independence of the axioms in theorem 2. (1) Consider ordering sets of internal links in S on the basis of a weighted average of the number of links within S, $\frac{\sum_{x \in S} \alpha(x) l(x))}{|A_{SS^c}|}$, where $l(x)$ is the number of links that x has with other members of S, and $\alpha(x) > 0$ is the weight, which is different for at least two elements of S. This criterion satisfies all axioms, except A.

(2) Define $k(x) = 0$ if $l(x) = 0$ and $k(x) = \sum_{i=1}^{l(x)} \beta^i$ if $l(x) > 0$. Let $0 < \beta < 1$. Consider ranking sets of internal links on the basis of $\frac{\sum_{x \in S} k(x)}{|A_{SS^c}|}$. This criterion satisfies all axioms, except I.

(3) Consider ordering sets of internal links in S on the basis of $-\frac{|A_S|}{|A_{SS^c}|}$. This criterion satisfies all axioms, except C.

(4) Consider ordering sets of internal links in S on the basis of the criterion in theorem 1. This criterion satisfies all axioms, except CON. ■

Appendix 2: finding cohesive subgroups

First we define $P^{k,\tau}$ the cohesive subgroups with k members:
$\forall k \in N, k \geq 2 : S' \in P^{k,\tau} \Leftrightarrow C_P\left(|A_{S'}|, |\Omega_{S'}|\right) \geq \tau$ and $|S'| = k$.
Evidently, $P_4^\tau = \cup_{k=4}^{|N|} P^{k,\tau}$. We now have the following proposition.

Proposition 3 $\forall k \geq 3 : S' \in P^{k,\tau} \Rightarrow \exists S \in P^{k-1,\tau} : S \subset S'$

Proof. The theorem is true if
$\forall S' \in P^{k,\tau} : \left[\exists S = S' \setminus x : C_P\left(|A_S|, |\Omega_S|\right) \geq C_P\left(|A_{S'}|, |\Omega_{S'}|\right)\right]$
Note that $|A_{S'}| = |A_S| + |A_{Sx}|$ where $S' = S \cup \{x\}$ and A_{Sx} contains the links between elements of S and x. Consequently,
$C_P\left(|A_{S'}|, |\Omega_{S'}|\right) = \frac{|A_S|}{|\Omega_S|}\frac{|\Omega_S|}{|\Omega_{S'}|} + \frac{|A_{Sx}|}{|\Omega_{S'}|}$
$= C_P\left(|A_S|, |\Omega_S|\right)\frac{|\Omega_S|}{|\Omega_{S'}|} + \frac{|A_{Sx}|}{|\Omega_{S'}|}$
And thus
$|\Omega_{S'}|\left[C_P\left(|A_{S'}|, |\Omega_{S'}|\right) - C_P\left(|A_S|, |\Omega_S|\right)\right] = C_P\left(|A_S|, |\Omega_S|\right)\left[|\Omega_S| - |\Omega_{S'}|\right] + |A_{Sx}|$
$= -C_P\left(|A_S|, |\Omega_S|\right)|S| + |A_{Sx}|$
This expression will be negative (as required above) if and only if
$\frac{|A_{Sx}|}{|S|} \leq C_P\left(|A_S|, |\Omega_S|\right).$
It is always possible to partition S' in S and x such that this condition holds true. We just have to identify an $x \in S'$ that has fewer than average links with the other elements in S. ■

The following procedure, based on the proposition is used to obtain P_4^τ:
(1) compute all $S \in P^{2,\tau}$;
(2) $\forall k > 2$, add an element to the sets in $P^{k-1,\tau}$ and verify whether the resulting set belongs to $P^{k,\tau}$. Then use the fact that $P_4^\tau = \cup_{k=4}^{|N|} P^{k,\tau}$.

References

Alba, R.D. (1973), A graph-theoretic definition of a sociometric clique, *Journal of Mathematical Sociology* 3, 113-126.

Bala, V. and S. Goyal (2000), A noncooperative model of network formation, *Econometrica* 68, 1181-1229.

Barberà, S., Bossert, W. and P.K. Pattanaik (2001), *Ranking sets of objects*, cahier 02-2001, Centre de recherche et développement économique, Université de Montréal, forthcoming in S. Barberà, P. Hammond and C. Seidl (eds.), Handbook of Utility Theory, Vol.2, Kluwer, Dordrecht.

Bloch, F., Genicot, G. and D. Ray (2005), Informal insurance in social networks, working paper.

Bock, R.D. and S.Z. Husain (1950), An adaptation of Holzinger's B-coefficients for the analysis of sociometric data, *Sociometry* 15, 206-219.

Bourguignon, F. (1979), Decomposable income inequality measures, *Econometrica* 47, 901-920.

Bramoullé, Y. and R. Kranton (2005), Risk-sharing networks, working paper.

Callander, S. and C.R. Plott (2005), Principles of network development and evolution: an experimental study, *Journal of Public Economics* 89, 1469-1495.

Coate, S. and M. Ravallion (1993), Reciprocity without commitment: characterisation and performance of informal risk-sharing arrangements, *Journal of Development Economics* 40, 1-24.

Coleman, J. (1984), Introducing social structure into economic analysis, *American Economic Review:* papers and proceedings 74, 84-88.

Cowell, F.A. (2000), Measurement of inequality, in: Handbook of Income distribution, A.B. Atkinson and F. Bourguignon (eds.), North Holland, 87-166.

Dercon, S. and P. Krishnan (2003), Risk sharing and public transfers, *The Economic Journal* 113, C96-C94.

Dercon, S. and J. De Weerdt (2005), Risk sharing networks and insurance against illness, *Journal of Development Economics*, forthcomming.

De Weerdt, J. (2002), Social Networks, Transfers and Insurance in Developing Countries, Ph.D., KULeuven.

Fafchamps, M. (1992), Solidarity networks in preindustrial societies: rational peasants with a moral economy, *Economic Development and Cultural Change* 41, 147-174.

Fafchamps, M. and S. Lund (2003), Risk-sharing networks in rural Philippines, *Journal of Development Economics* 71, 261-287.

Falk, A. and M. Kosfeld (2003), It's all about connections: evidence on network formation, mimeo University of Zurich.

Frank, K.A. (1995), Identifying cohesive subgroups, *Social Networks* 17, 27-56.

Galeotti, A. and S. Goyal (2004), Network formation with heterogeneous players, working paper.

Genicot, G. and D. Ray (2003), Group formation in risk-sharing arrangements, *Review of Economic Studies* 70, 87-113.

Greif, A. (1993), Contract enforceability and economic institutions in early trade: the Maghribi traders' coalition, *American Economic Review* 83, 525-548.

Jackson, M.O. and A. Wolinsky (1996), A strategic model of social and economic networks, *Journal of Economic Theory* 71, 44-74.

Jackson, M.O. and A. Watts (2002), The evolution of social and economic networks, *Journal of Economic Theory* 106, 265-295.

Kranton, R.E. and D.F. Minehart (2001), A theory of buyer-seller networks, *American Economic Review* 91, 485-508.

Ligon, E., Thomas, J.P. and T. Worrall (2002), Informal insurance arrangements with limited commitment: theory and evidence from village economies, *Review of Economic Studies* 69, 209-244.

McMillan J. and C. Woodruff (1999), Interfirm relationships and informal credit in Vietnam, *Quarterly Journal of Economics* 114, 1285-1320.

Murgai, R, Winters, P, Sadoulet, E. and A. de Janvry (2002), Localized and incomplete mutual insurance, *Journal of Development Economics* 67, 245-274.

North, D.C. (1990), *Institutions, institutional change and economic performance*, Cambridge University Press, 152p.

Pattanaik, P.K. and Y. Xu (1990), On ranking opportunity sets in terms of freedom of choice, *Recherches Economiques de Louvain* 56, 383-390.

Pattanaik, P.K. and Y. Xu (2002), On measuring personal connections and the extent of social networks, Department of Economics, University of California Riverside, working paper 02-04.

Platteau, J.P. (1991), Traditional systems of social security and hunger insurance: past achievements and modern challenges, in Ahmad, E., Drèze, J., Hills, J. and Sen, A. (eds.), Social Security in Developing Countries, Oxford Clarendon Press, 112-170.

Reining, P. (1967), The Haya: the Agrarian System of a Sedentary people, Ph.D. dissertation, University of Chicago.

Shorrocks, A. (1984), Inequality decomposition by population subgroups, *Econometrica* 52, 1369-1385.

Taylor, C., C. Lewis and M.C. Hudson (1972), World Handbook of Political and Social Indicators, Second Edition, (Reprinted in 1975), Yale University.

Thomson, W. (2001), On the axiomatic method and its recent applications to game theory and resource allocation, *Social Choice and Welfare* 18, 327-386.

Townsend, R.M. (1994), Risk and insurance in village India, *Econometrica* 62, 539-591.

Van Hees, M. (1998), On the analysis of negative freedom, *Theory and Decision* 45, 175-197.

Watts, A. (2002), Non-myopic Formation of Circle Networks, *Economics Letters* 74, 277-282.

Wasserman, S. and K. Faust (1997), *Social Network Analysis*, Cambridge University Press, 825 p.

Wright S. (1922), Coefficients of inbreeding and relationship, *American Naturalist* 56, 330-338.

Tables and Figures

Table 1: cohesion in subgroups of the population.

variable	category	$\|S\|$	$\|A_S\|$	$\|A_{SS^c}\|$	C_P	C_R
Clan	Muyego	3	0	19	.00	.00
(only clans with	Muyango	23	20	171	.08	.51
more than three	Musimba	20	29	125	.15	1.21
members are listed)	Muhimba	12	11	100	.17	1.07
	Muyozi	8	7	37	.25	3.00
	Muhunga	10	12	87	.27	1.67
	Musita	10	12	46	.27	3.16
	Mugaya	3	1	24	.33	2.42
	Musingo	4	3	16	.50	7.19
	Mumiru	6	10	30	.67	7.53
Religious affiliation	Lutheran	46	94	211	.09	.72
	Catholic	49	105	185	.09	.83
	Muslim	24	41	104	.15	1.63
	(Christian)*	(95)	(345)	(104)	(.08)	(.85)
Household head	no	47	73	234	.07	.49
completed primary?	yes	72	183	234	.07	.51
Age of household	bottom half	62	134	208	.07	.60
head category**	top half	57	148	208	.09	.79
Wealth category	bottom half	61	90	242	.05	.36
(livestock value)**	top half	58	158	242	.10	.70
Wealth category	bottom half	60	84	232	.05	.36
(land holdings)**	top half	59	174	232	.10	.78

Source: Nyakatoke Household Survey.

* Christians = Lutherans + Catholics.

** Group size may differ because all observations equal to the median value are placed in the bottom category.

Table 2: number of cohesive subgroups for different values of $|S|$ and C_P.

	$\|A_S\|$	C_P	Number of cohesive subgroups identified
$\|S\| = 4$	5	.83	749
	6	1.00	68
$\|S\| = 5$	8	.80	612
	9	.90	63
	10	1.00	7
$\|S\| = 6$	12	.80	189
	13	.87	21
	14	.93	4
	15	1.00	0
$\|S\| = 7$	16	.76	135
	17	.81	19
	18	.86	2
	19-21	.90-1.00	0
$\|S\| = 8$	21	.75	20
	22-28	.79-1.00	0

Source: Nyakatoke Household Survey.

Table 3: characteristics of the 12 C-groups.

C-group No.	contain all subgroups with	No. of subgroups	average No. of members	total No. of HHs that are member of a cohesive subgroup
1	$C_P = 1.00$	75	4.09	64 (54%)
2	$.90 \leq C_P \leq .93$	67	5.06	47 (39%)
3	$.86 \leq C_P \leq .87$	23	6.09	27 (23%)
4	$C_P = .83$	749	4.00	105 (88%)
5	$C_P = .81$	19	7.00	22 (18%)
6	$C_P = .80$	801	5.24	93 (78%)
7	$C_P = .76$	135	7.00	50 (42%)
8	$C_P = .75$	20	8.00	25 (21%)
9	$C_P \geq .75$	1889	4.88	105 (88%)
10	$\|S\| = 3, C_R \geq 20$	48	3.00	66 (55%)
11	$\|S\| = 3, C_R \geq 20, \|A_S\| \geq 2$	46	3.00	65 (55%)
12	$\|S\| = 3, C_P \geq .75$	315	3.00	109 (92%)

Source: Nyakatoke Household Survey.

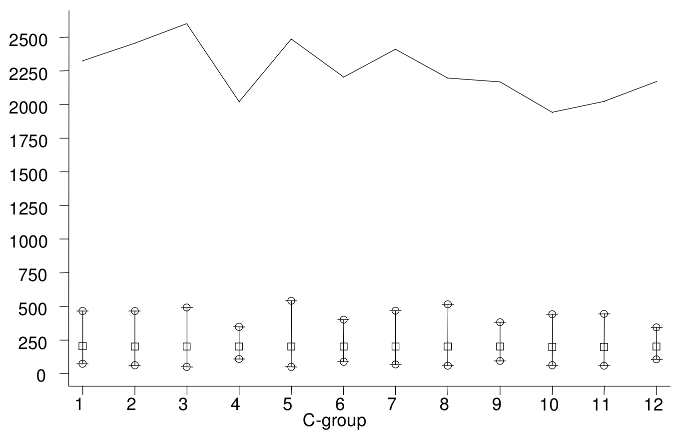
Table 4: Inclusion and exclusion from cohesive subgroups.

C-group No.	% heads who completed primary		average age of heads		average value of livestock (in TSh*)		average size of land (ha.)	
	mem	non-mem	mem	non-mem	mem	non-mem	mem	non-mem
1	64	56	44	46	83172	18658	1.57	1.07
2	66	57	45	45	87089	31333	1.58	1.18
3	63	60	50	44	143089	27020	1.76	1.21
4	59	71	45	44	57510	22186	1.37	1.09
5	59	61	52	44	175555	25639	1.88	1.21
6	60	62	45	45	63383	17485	1.48	.84
7	56	64	46	45	96360	22191	1.45	1.26
8	60	61	50	44	160944	24740	1.81	1.21
9	61		45		53355		1.34	
10	.55	.68	45	46	24182	89683	1.04	1.71
11	.54	.69	45	45	24492	88096	1.05	1.69
12	.59	.80	45	41	57301	10340	1.38	.82

Source: Nyakatoke Household Survey.

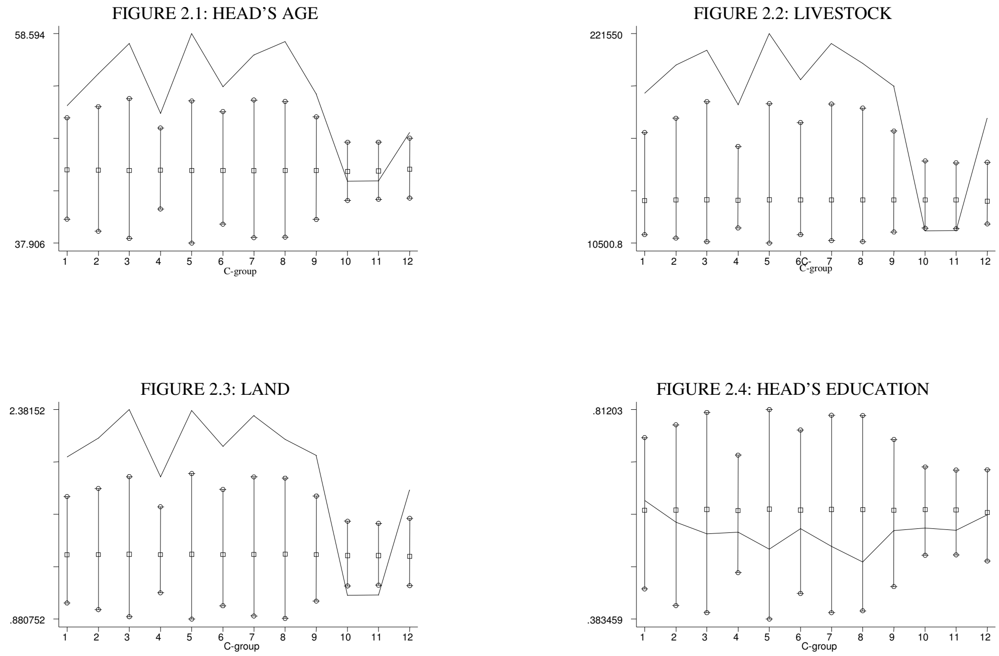
* TSh 800.00 = $1.00.

Figure 1: Flow of Gifts within Subgroups.



The values of the actual means within the subgroups are connected by straight lines. The vertical bars are confidence intervals for the randomization experiments: 90% of the 1000 randomization experiments yielded average yearly flows of remittances on this line, 5% lay above the line and 5% below it. The square lying on the vertical line indicates the mean value of this variable across the 1000 randomizations.

Figure 2: Characteristics of Subgroups.



The values of the actual means within the subgroups are connected by straight lines. The vertical bars are confidence intervals for the randomization experiments: 90% of the 1000 randomization experiments yielded average yearly flows of remittances on this line, 5% lay above the line and 5% below it. The square lying on the vertical line indicates the mean value of this variable across the 1000 randomizations.

Figure 3: Characteristics of Subgroups.

FIGURE 3.1: KINSHIP

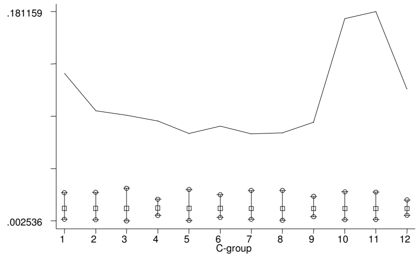


FIGURE 3.2: GEOGRAPHICAL DISTANCE

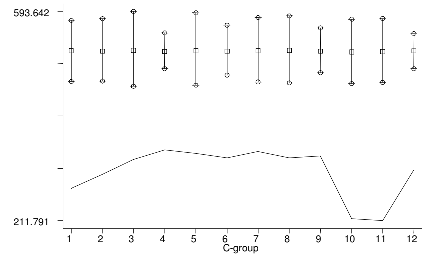


FIGURE 3.3: HEAD'S AGE

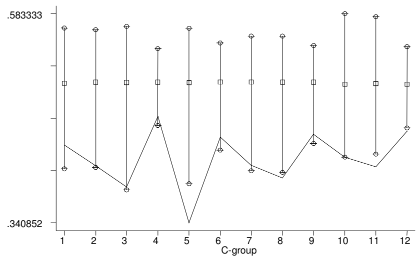


FIGURE 3.4: LIVESTOCK

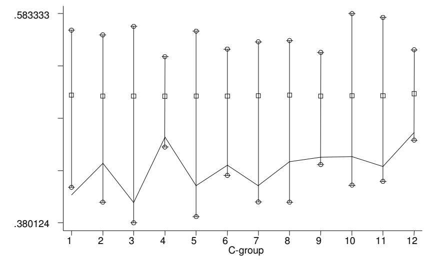


FIGURE 3.5: LAND

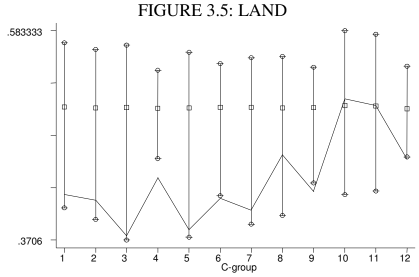


FIGURE 3.6: CLAN

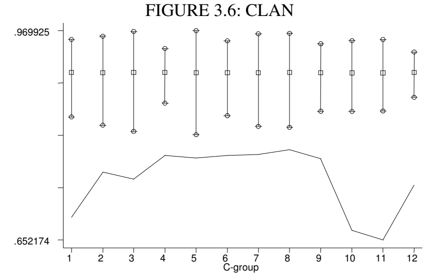


FIGURE 3.7: RELIGION

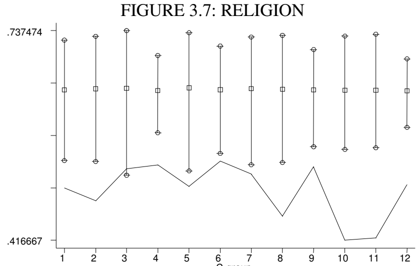
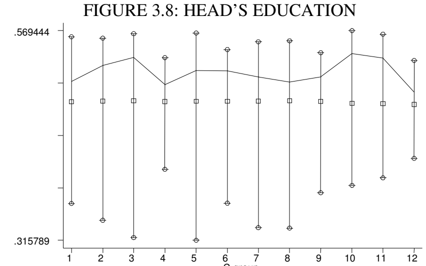


FIGURE 3.8: HEAD'S EDUCATION



The values of the actual means within the subgroups are connected by straight lines. The vertical bars are confidence intervals for the randomization experiments: 90% of the 1000 randomization experiments yielded average yearly flows of remittances on this line, 5% lay above the line and 5% below it. The square lying on the vertical line indicates the mean value of this variable across the 1000 randomizations.