BIROn - Birkbeck Institutional Research Online

Krishnan, P. and Sciubba, Emanuela (2006) Links and architecture in village networks. Working Paper. Birkbeck, University of London, London, UK.

Downloaded from: https://eprints.bbk.ac.uk/id/eprint/26929/

Usage Guidelines:
Please refer to usage guidelines at https://eprints.bbk.ac.uk/policies.html or alternatively contact lib-eprints@bbk.ac.uk.

ISSN 1745-8587

School of Economics, Mathematics and Statistics

Birkbeck Working Papers in Economics & Finance

BWPEF 0614

Links and Architecture in Village Networks

Pramila Krishnan
University of Cambridge

Emanuela Sciubba
Birkbeck, University of London

May 2006

Links and Architecture in Village Networks*

Pramila Krishnan
University of Cambridge

Emanuela Sciubba
Birkbeck College

May 17, 2006

Abstract

This paper provides a theoretical framework of endogenous network formation that yields testable predictions for the network architectures generated by a particular informal institution common in village economies. We test the implications of the model on data from rural Ethiopia. In contrast to the current literature, we demonstrate the critical role of both number of links and architecture in determining the impact of social networks on outcomes. Social capital matters, but its impact differs by the architecture of the network to which one belongs.

JEL Classification: D85, Z13, O12, O17.
Keywords: Endogenous network formation, social networks, rural institutions.

*This is a competely revised version of an earlier paper "Endogenous Network Formation and Informal Institutions in Village Economies". We would like to thank Tessa Bold, Stefan Dercon, Jeremy Edwards, Robert Evans, Natalie Quinn, Timothy Rothery, Hamid Sabourian, Tassew Woldehanna and participants at seminars at Birkbeck College London, Birmingham, Cambridge, Leicester, London Business School, Oxford, University College London, at the EUDN meeting in Paris, ESWC meeting in London and EEA meeting in Amsterdam for their comments on the earlier version of the paper. This project was funded by ESRC grant RES-000-22-0497, Nuffield grant SGF/00914/A and EC FP6 Integrated Project 001907 (DELIS).

1 Introduction

There is a vast empirical literature aiming to establish a link between social networks and economic performance (for a recent review, see [3]). In parallel, there is a burgeoning theoretical literature which aims to model the endogenous formation of social networks (for a recent review, see [27]). While these two literatures are slowly becoming aware of each other, there is still little empirical work that uses the equilibrium predictions of a theoretical model of network formation to assess the impact of networks on outcomes. In particular, while the main focus of the theoretical literature has been to characterise the structure or architecture of equilibrium social networks, the empirical work has typically used only the number of links interconnecting agents to represent networks[1]. For example, this is the approach taken in Putnam [37] and by Glaeser et al. [22]

In this paper, we provide a theoretical framework that offers testable predictions for the network architectures generated by a particular informal institution common in village economies. In contrast to the current literature, we demonstrate the critical role of both number of links and architecture in determining the impact of social networks on outcomes.

Since two seminal articles by Jackson and Wolinsky [28] and Bala and Goyal [1], a lively strand of literature has developed, offering a theory of strategic network formation. According to this literature individuals rationally choose whether to form or delete links with other agents. The focus has been to identify the network structures that emerge in equilibrium under different assumptions about the process of network formation and on the nature of the network externalities in place, whether positive or negative. In this paper, we build on this theoretical literature to capture the essential features of an informal labour-sharing network and allow for heterogeneity amongst agents. This provides us with a framework that can be readily adapted for use in empirical analysis. We carry the predictions of our model to data from a household survey conducted in 1994 in rural Ethiopia. We have detailed information on 1477 households in 15 villages in different regions of the country. The data offer information on various aspects of household structure, incomes, consumption, assets, investment and participation in local informal networks. We use the data to test the consistency of the empirical evidence with the theoretical equilibrium characterisations, and illustrate the implications of the theoretical framework for inference on the impact of networks on economic performance.

Our focus on network formation is driven by the observation that important market and non-market transactions take place within social networks in these villages. These include informal insurance groups (largely fu-

[1] It should be acknowledged that the sociological literature on networks has been far richer in this respect and network characteristics beyond the number of links have been taken into account ([23], [31]).

neral societies, the *iddir*), rotating savings and credit associations (*equbs*), as well as various kinds of oxen-sharing, crop sharing and labour sharing arrangements. Among this plethora of informal institutions, labour sharing arrangements stand out for they are both ubiquitous and vary in the fashion in which members are connected to each other. In contrast, both *iddirs* and *equbs* operate as tightly-knit groups where every agent is directly linked to every other while oxen-sharing and crop-sharing arrangements typically involve pairs of farmers so that there is no variation in architectures in such arrangements. Labour sharing networks in these villages are classified in different ways in the local languages but two main types of labour sharing arrangements seem to dominate. We discuss this variation in type in detail in what follows: however, the fact that otherwise similar work parties are known under different local names according to the network architecture of the labour-sharing arrangements provides support to our view that the structure of the social network and not just its size matters when it comes to identifying informal institutions.

In this paper, we thus focus on labour sharing networks. We build a simple model of labour-sharing arrangements where farmers can decide whether and with whom to labour-share. Farmers differ in their quality or other productivity enhancing endowments. A better endowed farmer is a more appealing partner; however better endowed farmers have lower incentives to labour-share. As ubiqitous in the game-theoretic literature on endogenous network formation, the village economy that we model admits multiple equilibria: different social networks might emerge among the same farmers in the same village, under the same conditions. However, while this has the important implication that it is impossible to uniquely predict the social network that will form on the basis of the initial distribution of endowments, we nevertheless identify equilibrium characterisations that are shared by any stable network architecture. In particular we show that in any endogenous network symmetric labour-sharing arrangements (i.e. labour-sharing arrangements between farmers who have the same number of partners) are stable only among farmers who are of similar quality. On the contrary, asymmetric labour sharing arrangements (i.e. such that the farmers involved have a different number of partners) are stable only among farmers who differ greatly in quality and endowments, with the better farmer having more links than his less endowed neighbour. Multiplicity of equilibria prevents a simple direct test of our network formation model: any initial distribution of endowments could lead to multiple equilibrium configurations. We can nevertheless take the features of our equilibrium characterisation to the data and on doing so find that the observed social network structures in rural Ethiopia conform to the properties that our model predicts. In accordance with our theoretical characterisation, we find empirically that symmetric networks are more likely to emerge among households who do not differ in quality; heterogeneity in quality, on the other hand, is more likely to be observed among participants in asymmetric network structures.

These results have important consequences for the impact of networks on economic performance. In our data, we illustrate this by focusing on agricultural output and the impact of labour-sharing networks. The theoretical framework in this paper emphasises the key role of architecture in the characteristisation of equilibrium networks. Consequently we take the view that correcting for the endogeneity of networks cannot be based solely on instrumenting for the number of connections, but must also account for the network architecture in order to obtain accurate estimates of the impact of networks on outcomes.

The structure of the paper is as follows. In section 2 we summarise the existing literature. In section 3, we present a simple endogenous network formation model to explain the emergence of particular architectures for labour-sharing groups. Section 4 describes the data. Sections 5 provides an empirical test of the equilibrium characterisation while Section 6 examines the implications of endogenous links and architecture for performance. Section 7 concludes the paper.

2 Related Literature

2.1 The Theory of Endogenous Network Formation

Bala and Goyal [1] and Jackson and Wolinsky [28] consider a model of information transmission where network externalities are positive: individual payoffs are increasing in the number of agents accessed through the network, both directly and indirectly. If only the direct links are costly to the agents and if information transmission is frictionless, so that reported information is as good as information received first hand, the only stable network architecture which is not empty is the star network[2]: a very asymmetric network structure where there is a central individual to whom all other agents in the population are linked. Hence the whole population is connected (there is a path linking any pair of individuals) and it is also minimally connected (there is only one such path, through the central agent). No individual is isolated.

Not surprisingly the results for negative network externalities are quite different. Jackson and Wolinsky [28] look at a model where the payoff to each agent in a given network is increasing in the number of direct links but decreasing in the number of indirect links. In such a setup, the stable network architecture is very symmetric. The equilibrium network has several fully interconnected components where all agents have the same number of links.

[2] Clearly when the cost of link formation is sufficiently high, the only stable configuration is the empty network.

The fact that from the very complex problem of endogenous network formation, extremely simple network architectures emerge as stable is certainly intriguing. However it does rest on some heroic assumptions.

A factor which plays a big role in these results is that all agents in the population are assumed to be identical and to face the same incentives towards link formation. The few results on network formation with heterogeneous agents are very promising in that they point at directions which are also different from a qualitative point of view. Hence the homogeneity assumption does not just simplify the analysis but has a much more crucial role.

In some recent work, Galeotti and Goyal [20] prove that the set of network architectures that result as the equilibrium outcome of an endogenous network formation game with heterogeneous agents is extremely large. The requirements that Nash equilibrium alone imposes on the structure of the network are minimal: almost any network structure can result as a Nash equilibrium outcome, for an appropriate set of costs and benefits of forming links.

It would seem that the greatest challenge that the treatment of heterogeneity poses is that some structure is needed in the way in which the incentives to form links are modelled. Ideally one would like to have a framework where both the network formation process and the incentives that underlie its formation are endogenised.

Johnson and Gilles [29] make a first step in this direction by endogenising the cost of forming a link within spatial social networks. They consider a population of agents who are located on a geographical space, which could also be interpreted as a space of characteristics. They assume that the cost of forming a link between any two agents depends on their distance within this space. They find conditions such that the simple network architecture given by the chain is stable. The chain is clearly a network architecture that is drastically different from the star, obtained under the assumption of homogeneous costs within the population.

Haller and Sarangi [26] provide results that seem to challenge the claim that, when network externalities are positive, the equilibrium network is connected, so that there is a path between any two agents in the population and no individual is isolated. They consider a network where there is heterogeneity in the (exogenous) probability that each link may fail. What they find is that the result in Bala and Goyal [2] that Nash networks are either connected or empty is only true when the probabilities of failure or success of different links are not very different from each other. Otherwise partial connectedness may occur in equilibrium. As Galeotti and Goyal [20], Haller and Sarangi [26] fail to provide more accurate predictions on what the stable network architecture might look like under heterogeneity. More structure is needed on the way in which the heterogeneity itself is modelled.

The lesson that can be drawn from this recent theoretical literature on endogenous network formation is that simple stability requirements are extremely successful in pinning down specific network architectures under the assumption of homogenous agents. As soon as we depart from this assumption and introduce the possibility that different agents might face different incentives of forming and/or deleting links, predictive power is lost and much more structure is needed on the heterogeneity of the incentives that we believe to be in place. This implies that while it is might not be feasible to produce a general and all-encompassing theory of network formation with heterogeneous individuals that has any predictive power, it ought to be possible to provide models of endogenous network formation where the incentive structure is fitted to the particular application at hand.

With this qualification in mind, we focus on a theory of endogenous network formation in labour-sharing, building on the observed characteristics of institutions of labour exchange in rural Ethiopia.

2.2 Empirical Literature

The evidence suggests that many individual outcomes from school attendance to adoption of new technologies are correlated with the behaviour of the social group or neighbourhood to which the individual agent belongs. In the context of rural economies, the focus thus far has been on learning about technologies and on informal insurance. The few papers on this topic (Foster and Rosenzweig [19], Conley and Udry [9], Bandiera and Rasul [4], Fafchamps and Lund [16], De Weerdt and Dercon [13]) all find that social effects matter. Measuring social effects is problematic for two main reasons. First, the network needs to be identified. Second, even if the network is correctly identified, distinguishing endogenous social effects from other correlated phenomena is very hard to do (Manski [30], Brock [7]). Network members may just behave in similar ways, not because they affect each other but because, for instance, they live in similar environments and use identical technologies.

The empirical literature on networks has largely taken the network as exogenous. In work on developing countries, often villages are taken as the unit of study (Townsend [38]) and where they are not, exogenous characteristics such as ethnicity or caste are thought to structure the network (Fafchamps and Lund [16], Grimard [24] , Munshi and Rosenzweig [33], Luke and Munshi [34]). But even if channels of communications are inherited, the decision of whether to maintain them (and at what level of activity) does remain. Foster and Rosenzweig use village-level averages to proxy the information set, while Bandiera and Rasul have information on the number (but not the identities) of people who might be part of the network. Conley and Udry [9] do better, for they have information on the actual identities of individuals who might represent the social network for information flows. De Weerdt and Dercon [13] examine networks formed to offer informal insurance: again, they use information on all the individuals that can be relied upon to provide transfers. However, none of these papers seeks to model the formation of networks nor do they use any information on the structure of the network: for instance, are all individuals in the social group connected to every other or are they linked in some other fashion? Conley and Udry [9] acknowledge that "The next step in this research programme is to model the choices of farmers regarding the formation of information links in these villages [.....] but we know of no empirical work in economics that examines the formation of decentralised networks".

In some recent work Udry and Conley [39] address this point, by offering evidence on the determinants of link formation among farmers in Ghana. They map information, financial, labour and land networks and estimate the probability of link formation between any two farmers as a function of own and partner's characteristics. In particular they find that network interactions are more likely between farmers that are located near each other, that are in the same matrilineage. Information links are more likely amongst those who are different in wealth and who share the same soil characteristics. De Weerdt [12] presents a similar empirical analysis for Tanzania, investigating the probability of links for informal insurance based on a full village census. He finds that kinship, geographical proximity, clan membership, religion and wealth all determine participation in networks and in particular, the poor have fewer informal insurance links than the rich. He does not investigate why this is so or indeed its effect on outcomes.

Udry and Conley [39] discuss an equilibrium model of network formation, where each agent's decision to link up also depends on his expectations on what everyone else will do, and argue that estimation of such a model would be problematic because of multiplicity of equilibria. This is an obvious difficulty that any empirical work on network formation must face and this paper is no exception. We therefore take an alternative approach: rather than testing equilibrium predictions, we identify equilibrium characterisations, i.e. properties that all possible equilibria in our model must share, and we test whether the data conform to these properties.

3 A Theory of Labour-Sharing Arrangements

The Ethiopian economy is dominated by agriculture which contributes about 45% of GDP and employs 80% of the labour force. The rural economy has long been subject to strict controls which have strongly affected the structure of the agricultural labour market. The 1975 land reform declared illegal all private ownership of land, as well as transfers of land by lease, sale or mortgage. Since then all land is state owned and allocated to farmers by the local Peasants Association (PA)[3]. The 1975 land reform also prohibited tenancy and wage labour, although the legal code allowed female-headed households with dependents, landholders who were ill and soldiers and their dependents to hire labour or lease out their land.

In 1990 the government started a program of economic reforms: land tenancy and wage labour (but not land sales) were made legal[4]. In the last twenty years markets for agricultural labour have gradually emerged, but in most villages hired labour is still very uncommon. Wage-labourers are often villagers who have lost their entitlement to land and for this reason there is a clear social stigma attached to them[5]. Supply of hired labour is extremely thin. Households draw most of the labour they require from household members. However the seasonal nature of some of the activities makes household labour periodically insufficient[6]. Labour shortage particularly occurs during the harvesting, ploughing and weeding seasons. To meet the high labour demand during different agricultural operations, people depend heavily on labour-sharing arrangements. In the sample used here, only 4% of the labour used per hectare is hired against cash or other payments in kind. More than half of the households reported having used labour in the form of a labour-sharing arrangement and about a fifth of the total days worked were supplied in this way.

Labour-sharing arrangements involve a group of people working together for a particular task, typically but not exclusively for agricultural activities, such as harvesting and weeding. The tasks involved are those requiring many hands, benefiting from team labour. The exchange of labour is symmetric in terms of duration and task - a household that is invited to help weed a field for a day, expects to be reciprocated for the same task, for a similar length of time. Bevan and Pankhurst [6] emphasise that farmers value labour exchanges for the synergy generated in working with each other. In all cases, calling a work party implies willingness to reciprocate, either virtually immediately or in the future. Enforcement is obtained through repeated interaction: work party participants are usually from the same village and involve both relatives and friends (on average, 24% of parties involve at

[3] The rights to transfer land remained highly restricted and despite some attempts at liberalisation, remain so. Transfer through lease, sale, exchange or mortgage was completely prohibhited and inheritance allowed only to immediate family members, with some risk of re-distribution upon inheritance. The use of land is still contingent upon physical residence making even temporary migration difficult.

[4] One year later, the Derg government was forcibly removed when EPRDF-forces won the civil war. The new government has confirmed lease rights in land and wage labour to be legal.

[5] Reasons for losing one's entitlement to land include own land neglect and tax evasion.

[6] The production technology in most highland areas is an ox-plough, with the main crops being teff (a grain particular to highland Ethiopia), barley and maize. Modern inputs, particularly the use of fertilisers have expanded since 1996; however, for the period under study, the use of inputs beyond land, labour and oxen was minimal.

least one relative). Social sanctions are also quite harsh: farmers who have called a work party but have then failed to reciprocate are socially ostracized with all their family and nobody would enter labour-exchange arrangements with them in the future. There is a large variety of specific labour-sharing arrangements in Ethiopia. While they all share these common elements, there are some clear differences between the observed work parties. For example, in Amhara[7] regions, a *wonfel* is often used to describe a work party performed by a group who work in rotation for each group member, within the same season. Rotation is strict. A *debo* involves a work party whereby it is expected that the household calling the party will reciprocate if called upon by any of the participants some time in the future. In a *debo,* the household organising the work party provides drink (and/or food) for the day but this does not occur in a *wonfel.*

In practice, the main difference between a *wonfel* and a *debo* seems to lie in the architecture of the labour-sharing arrangement. Our data suggest that reciprocation of labour largely took place within the same season, whether the household participated in a *debo* or a *wonfel. Wonfels* are tightly-knit groups, similar in their structure to *equbs* (rotating savings and credit associations), where every participant is symmetrically linked to everyone else in the group. Each member of a *wonfel* relies on the same set of households for labour-sharing. The network of labour-sharing arrangements that results from this is dense and heavily clustered. *Debos*, on the other hand, have a more sparse and unclustered structure, where each of the participants to the work party typically relies on a different set of households. A clear implication of this is that while *wonfels* have a symmetric structure where all the members of the work party have the same number of links, *debos* typically take asymmetric structures, where some members have more links than others.

The fact that otherwise similar work parties are known under different local names according to the network architecture of the labour-sharing arrangements provides support to our view that the structure of the social network and not just its size matters when it comes to identifying informal institutions. Moreover this observation suggests that for labour-sharing relevant aspects of the network architecture are the symmetry or asymmetry of the relationship and the level of clustering of the local set of ties.

Finally it is worth emphasising that, irrespective of the network architecture that emerges from the union of ties, labour-sharing arrangements are bilateral agreements. The pattern of labour-sharing arrangements observed in these villages mirrors those remarked on elsewhere, particularly

[7]The Amhara have traditionally been the politically and culturally dominant ethnic group of Ethiopia. They are located primarily in the central highland plateau of Ethiopia and comprise the major population element in the provinces of Begemder and Gojjam and in parts of Shoa and Wallo. In terms of the total Ethiopian population, however, the Amhara are a numerical minority.

Africa and Asia. A large number of accounts of such arrangements (called cooperative labour or exchange labour in the anthropological literature) are explored by Erasmus [15] and summarised by Moore [32]. They suggest that exchange labour arrangements are bilateral in the main: "reciprocal arrangements may be made amongst individuals: A and B may work six days for C; C in return works six days for each of them, but A and B do not work for one another at all". Note that this does not exclude the possibility that C has bilateral arrangements in place with each of A and B but the important point is that the reciprocity is to the individual and not to the group. This is emphasised by the pattern observed in these Ethiopian villages where the custom is for each participant to the work party to be invited individually. Moore [32] highlights several different economic advantages offered by labour exchange arrangements in rural areas of developing countries, covering most of the reasons offered here including the lack of hired labour market (perhaps precipitated by the lack of landless labour), unpredictable requirements for larger labour parties, economies of scale and higher motivation in working at tedious tasks.

For the purpose of this study we consider a very stylised model for the formation of labour-sharing arrangements. In particular we build on a model by Jackson and Wolinsky [28] which we modify by introducing heterogeneity.

Consider a set N of $n \geq 3$ farmers, indexed by $i = 1, 2, ..., n$ and a set G of (potential) labour-sharing arrangements $g_{ij} \in 0, 1$, where $g_{ij} = g_{ji} = 1$ represents that i and j have a labour-sharing arrangement and $g_{ij} = g_{ji} = 0$ represents that i and j do not labour-share. Given that labour-sharing arrangements are reciprocal we assume that $\forall ij, g_{ij} = g_{ji}$. The resulting (undirected) graph $g = \{g_{ij}\}_{i,j \in N}$ represents the network of labour-sharing arrangements.

Individual harvest of farmer i is a function of own effort, effort of all the other participants in the labour party and a synergy component that represents the fact that more can be achieved by working together than individually. Assume now that farmers differ in quality and denote by $q_i \geq 1, \forall i$ some index of the quality of farmer i. Denote by n_i the number of partners that farmer i has and assume that each farmers allocates his quality adjusted time (or effort) equally across himself and all partners, so that the amount of effort that farmer i exerts on his own field and in each collaboration is equal to $q_i/(n_i + 1)$.

The payoff (the harvest) of farmer i, given his position in the network g, is assumed to be equal to:

$$h_i(g) = \frac{q_i}{1 + n_i} + \sum_{j:ij \in g} \left[\frac{q_j}{1 + n_j} + \frac{q_i q_j}{(1 + n_i)(1 + n_j)} \right] \tag{1}$$

where

$$\frac{q_i}{1+n_i} + \sum_{j:ij\in g} \left[\frac{q_j}{1+n_j}\right]$$

is total effort exerted by farmer i and his work party; and the term

$$\sum_{j:ij\in g} \frac{q_i q_j}{(1+n_i)(1+n_j)} = \frac{q_i}{1+n_i} \sum_{j:ij\in g} \frac{q_j}{1+n_j}$$

is the synergy component: more is achieved by working together than alone. The presence of such synergy implies that entering a labour-sharing arrangement, although costly to the farmer (it reduces own effort) may nevertheless be strictly profitable. As in [28], a feature of this model is that there is an implicit cost of link formation in that, through the synergy component, any additional link dilutes the synergy with all existing partners.[8]

Farmers maximise total harvest and can rationally form new links or sever existing ones to this aim. Links can be deleted unilaterally but for a link to be formed both farmers involved have to agree. This element of mutual consent implies that it is difficult to try to use any off-the-shelf noncooperative game theoretic solution concepts. "In whatever game one specifies for link formation, requiring the consent of two players to form a link means that either some sort of coalitional equilibrium concept is required, or the game needs to be an extensive form with a protocol for proposing and accepting links in some sequence. Another serious challenge to the off-the-shelf noncooperative game theoretic approach is that the game is necessarily ad hoc and fine details of the protocol (e.g., the ordering of who proposes links when, whether or not the game has a finite horizon, players are impatient, etc.) generally matter. [...] A different approach to modeling network formation is to dispense with the specifics of a noncooperative game and to simply model a notion of what a stable network is directly" (Jackson, [27], p. 21). This is the approach we take as well, in line with the argument cited above.

Jackson and Wolinsky [28] characterise *stable* networks, by introducing the notion of pairwise stability. A pairwise stable network is such that no agent has an incentive to exit from existing collaborations and no pair of agents have any incentive to form a new link. This notion of stability allows that agents review their relationships with network members, one at a time[9]. Given the long-standing and bilateral nature of labour-sharing arrangements, we believe this to be the appropriate stability concept here.

[8] One might compare this to someone who hosts a large party, where each of the friends invited contributes to the fun, but when the number of invited friends increases, the quality of the interaction with each of them decreases. Reduced ability to monitor might have a similar impact in labour-sharing arrangements for agriculture.

[9] For example, a pairwise stable network is not necessarily robust to deviations that involve an agent deleting his links to more than one partner at a time; or deleting a link with one partner and simultaneously initating a new link with another.

More formally we state the following:

Definition 1 (Pairwise stable network) *A network g is pairwise stable if*

1. $\forall ij \in g \Rightarrow h_i(g) \geq h_i(g - ij)$ *and* $h_j(g) \geq h_j(g - ij)$
2. $\forall ij \notin g$, *if* $h_i(g) < h_i(g + ij) \Rightarrow h_j(g) > h_j(g + ij)$

where $g - ij$ is the shorthand for the network that is obtained by deleting the link between i and j in network g; $g + ij$ is the shorthand for the network that is obtained by adding the link between i and j in network g. Moreover we use the notation $ij \in g$ to represent the fact that $g_{ij} = g_{ji} = 1$, so that i and j are directly linked in the network g; the notation $ij \notin g$ to represent the fact that $g_{ij} = g_{ji} = 0$.

In this setup, farmers of different quality will face different incentives to link formation. Farmers who are of higher quality (or better endowed) are more appealing partners. Farmer i will be more willing to link to farmer j when q_j is high: labour-sharing partners of better quality contribute more to the harvest. However, higher quality farmers are less willing to form new links: the increase in the number of links has a larger negative marginal impact on the harvest of a better farmer. Finally, farmers whose existing partners are of better quality have lower incentives to form new links for a very similar reason: through the synergy component of their payoffs, the increase in the number of their own links has a negative marginal impact on the level of harvest and this is largest when the pool of current partners is better endowed.

We can prove the statements above as follows. Consider $ij \notin g$: farmer i will (strictly) want to form a link to farmer j if:

$$h_i(g + ij) > h_i(g)$$

where $h_i(g)$ in (1) can also be rewritten as:

$$h_i(g) = \frac{q_i}{1 + n_i} + \left(1 + \frac{q_i}{1 + n_i}\right) \sum_{k:ik \in g} \frac{q_k}{1 + n_k}$$

$$h_i(g + ij) = \frac{q_i}{2 + n_i} + \left(1 + \frac{q_i}{2 + n_i}\right) \left[\sum_{\substack{k:ik \in g \\ k \neq j}} \frac{q_k}{1 + n_k} + \frac{q_j}{2 + n_j} \right]$$

One can easily check that $h_i(g + ij) > h_i(g)$ iff

$$\frac{q_j}{n_j + 2} > \frac{q_i}{1 + n_i} \cdot \frac{1}{n_i + 2 + q_i} \left[1 + \sum_{\substack{k:ik \in g \\ k \neq j}} \frac{q_k}{1 + n_k} \right] \tag{2}$$

We can interpret the left hand side and the right hand side of (2) respectively as the marginal benefit and cost that i faces when linking to j. The marginal benefit represents the fact that j contributes his effort to the harvest of i: this is increasing in j's quality (q_j) and decreasing in the number of links of j (n_j). The marginal cost represents the fact that both directly, through costly reciprocation, and indirectly, through the synergy component, the increase in the total number of links of i has a negative impact on farmer i's harvest: this marginal cost is increasing in the quality of the existing partners of i (i.e. it is increasing in q_k); moreover the marginal cost of linking to j is also increasing in q_i. In fact differentiating the rhs of (2) with respect to q_i one obtains:

$$\frac{n_i+2}{\left(1+n_i\right)^2\left(n_i+2+q_i\right)^2}\left[1+\sum_{\substack{k:ik\in g\\ k\neq j}}\frac{q_k}{1+n_k}\right]>0$$

In this setup links have to be mutually agreed, i.e. for a link to be formed between i and j we will require both i and j to be willing to make that link. Hence we can ask how the sorting takes place: will strong farmers always link up with strong, and the weak with weak, or is it the case that poor (in terms of quality) and better farmers will want to link? The answers to these questions crucially depend on the equilibrium network architecture. In particular whether or not two farmers are willing to form a labour-sharing arrangement will depend not only on the quality of the two potential partners, but also on the number of links that each of them already has in place, and on the average quality of the existing neighbourhoods of each of them. It would seem, then, that any question on how the strong and weak farmers sort cannot be answered without first characterising the equilibrium network architecture.

In order to be able to state our results formally, we need some definitions. First, the notion of a (network) component:

Definition 2 (Component) *Given a network g a component $C(g)\subset N$ is a set such that $\forall i,j\in C(g)$, i and j are directly or indirectly connected and $\forall i,k$ such that $i\in C(g)$ and $k\in N\backslash C(g)$, i and k are not connected either directly or indirectly.*

For a network g, let $m\geq 1$ be the number of components in g. The partition in components is denoted by

$$P\left(g\right)=\left\{C_1(g),C_2(g),...,C_m\left(g\right)\right\}.$$

Next we distinguish between components which are symmetric and components which are not: when a component is symmetric, any pair of agents in the component who are directly linked have the same number of total links. A fully interconnected component is a special case of a symmetric component: in particular a fully interconnected component of size m is a symmetric component where all agents have $(m-1)$ links. We call a component asymmetric whenever it is not symmetric.

Definition 3 (Symmetric Component) *A component $C(g)$ is called symmetric if $\forall i, j \in C(g)$ such that $ij \in g$, $n_i = n_j$.*

Definition 4 (Asymmetric Component) *A component $C(g)$ is called asymmetric if $\exists i, j \in C(g)$ such that $ij \in g$ and $n_i \neq n_j$.*

We are now ready to state our first characterisation of pairwise stable labour-sharing networks: if a pairwise stable network admits a component where all agents are of identical quality, then that component must be symmetric. Conversely, any asymmetric component in a pairwise stable network is given by a set of agents with differing qualities. This is the key testable implication of our theory that will later on be taken to the data. More formally:

Proposition 1 *Given any pairwise stable network of labour-sharing arrangements g, if there is a $C(g) \in P(g)$ such that $\forall i \in C(g)$, $q_i = q$, then $C(g)$ is symmetric.*

Proof. See appendix.

Hence whenever agents who are equally endowed enter in a labour-sharing arrangement, they do so with farmers who have the same number of total links as they have.

We can go further and show that homogeneous components of pairwise stable networks of labour sharing arrangements are fully interconnected. In a component of farmers of the same quality, everyone is directly linked to everyone else.

Corollary 1 *Given any pairwise stable network of labour-sharing arrangements g, if there is a $C(g) \in P(g)$ such that $\forall i \in C(g), q_i = q$, then $C(g)$ is complete, i.e. $\forall i, j \in C(g)$ $g_{ij} = 1$*[10].

Proof. See appendix.

The result in proposition 1 provides us with a characterisation of asymmetric components: if farmers with different numbers of total links agree to link up, it must be the case that they are heterogenous.

Corollary 2 *Given any pairwise stable network of labour-sharing arrangements g, if $\exists C(g) \in P(g)$ such that $C(g)$ is asymmetric, then $\exists i, j \in C(g)$, such that $ij \in g$ and $q_i \neq q_j$.*

[10] We would like to thank Natalie Quinn for pointing this out.

Proof. See appendix.

Proposition 1 gives us a sufficient condition for a symmetric component, not a necessary one. Hence we might still have components of pairwise stable networks that allow for heterogeneity of labour quality. However, if a component of a pairwise stable network is symmetric, the heterogeneity cannot be very large.

Proposition 2 provides us with a characterisation.

Proposition 2 *Given any pairwise stable network of labour-sharing arrangements g, if $C(g) \in P(g)$ is symmetric, then, for each agent $i \in C(g)$, the deviation of the quality of the weakest neighbour of i from the average quality of i's neighbours, is bounded above by*

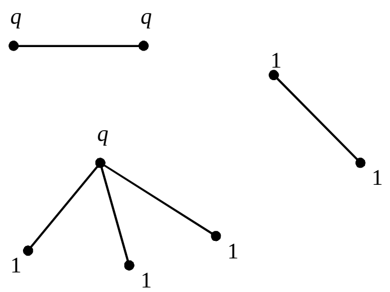
$$\overline{q}_i - q_{\widehat{j}} < n + \frac{q_{\widehat{j}} - 1}{n}$$

where n is the degree of the symmetric component and $\widehat{j} = \arg\min_j \{q_j\}_{j:ij \in g}$.

Proof. See appendix.

Through propositions 1 and 2 we have shown that pairwise stability is only compatible with the following: homogeneity gives us symmetric components; on the other hand, for a component to be symmetric, quality cannot be too heterogeneous. This characterisation also implies (corollary 2) that *if* a pairwise stable network admits asymmetric components, these must be such that participants are sufficiently heterogenous. The following example establishes that asymmetric components can be part of a pairwise stable network architecture. Hence it shows that the case of corollary 2 is not just hypothetical.

Consider the following network where 8 agents form three separate components: a star and two pairs.



Assume that quality can only take two values so that $q_i \in \{1, q\}$. In the diagram above each node is labelled with the corresponding quality.

It is easy to show that such a network, which comprises both symmetric and asymmetric components, is pairwise stable as long as $\frac{20}{3} \leq q \leq 12$. The upper bound on q is needed so that the hub in the asymmetric component agrees to be linked at all; the lower bound guarantees a high level of heterogeneity among participants to the asymmetric component. It can be showed that if the hub is sufficiently stronger than each of the spokes, these are not only happy to maintain their links with the hub, but also are not inclined to link up between them, so that asymmetry in the number of links is indeed pairwise stable.

We should emphasise that equilibrium (here pairwise stable) networks are typically inefficient. This should not be surprising given that there are negative network externalities in place. When individually optimising in deciding whether or not to form a new link, farmers do not internalise the loss that their connection causes for their existing partners. In equilibrium we will observe more labour-sharing than it is optimal. For example: among homogenous farmers, the aggregate payoff is maximised whenever labour-sharing occurs in pairs and it decreases monotonically whenever more farmers join in the work-party. Individual incentives to labour-share, however, do not work in the direction of minimising the size of the labour-sharing group and no farmer has any incentive to unilaterally deviate from an arrangement, no matter how large the symmetric component. Such tension between pairwise stability and efficiency is somewhat mitigated within asymmetric components, where the willingness of the spokes to join in the work-party decreases rapidly with the size of the star. Hence for asymmetric components, stability requirements pose some bound on the level of overconnectedness (and inefficiency) that one would observe in equilibrium.

In summary, the theoretical framework and results from this section provide a number of testable implications. We have characterised labour sharing as the outcome of (pairwise stable) network formation with negative externalities, given our assumption about the value of output for farmers in a network. Consequently, the amount of *debo* (or *wonfel)* labour a farmer is able to generate is increasing in the number of his own links but decreasing in the amount of links his partners have. The model allows for multiple equilibria so a simple direct test of our network formation model is not possible: any initial distribution of endowments could lead to multiple equilibrium configurations. The route taken is to consider the features of the equilibrium characterisation and take these to the data. Propositions 1 and 2 lead us to a key testable hypothesis: in equilibrium, asymmetric networks are necessarily heterogenous while symmetric networks tend to be more homogenous in the quality of endowments. Increased heterogeneity in the quality of the network partners is only compatible with asymmetric structures. Also, as suggested by corollary 1, homogeneous groups are not only symmetric but also highly clustered.

Finally, notice that our model allows for the fact that some farmers may not link at all. When this happens, it does not depend uniquely on the endowments and characteristics of the isolated farmer, but also on the specific network architecture that emerges among those who link; in particular, the same farmer may be isolated in some equilibria and not in others. For this reason we do not conduct an empirical test of who is likely to belong to a labour-sharing network and who is not, and we instead focus on testing for those who belong to a network whether the observed network architecture conforms to those properties that our model predicts for equilibrium networks.

4 The Data

The data used in this paper come from the second round of the Ethiopian Rural Household Survey, conducted in 1994. This is a panel data survey conducted by the Economics Department of Addis Ababa University in collaboration with the Centre for the Study of African Economies at Oxford University. It covers 15 villages, representative of different areas and agro-ecological zones across the country, and a total of 1477 households were interviewed. The survey has detailed socio-economic information on households, including consumption, incomes, productive activities, assets and demographics. The second round of this survey is particularly useful for our purposes since households were asked specific questions about the local networks they participate in. Surveys in these areas were supplemented by anthropological data on the character of these institutions. Since the questions on local networks were posed only once, we cannot exploit the panel nature of the data. However, we do use information on endowments and household structure obtained in the first round of data collection, about 6-8 months previous to the second round. The data collected pertain to the harvest in the main season, the *meher*, and hence the arrangements in place also refer to this period. In the sample, labour-sharing (*debo/wonfel*) is very common in all but the two most Northern (Tigrayan) villages. In these two villages and the surrounding region, labour-sharing and most of the other informal institutions, such as rotating savings and credit associations, funeral groups and forms of sharecropping, common throughout rural Ethiopia, were not at all prevalent and were of forms not comparable to the other areas of Ethiopia[11] These two villages were excluded from the analysis. Since the focus of our analysis is the formation of particular groups within well-defined communities, excluding these villages is not a problem[12].

[11] The main reason for the absence of work parties is ascribed to the very small land holdings in these areas and the perception that family labour is sufficient. These areas also report that labour sharing is rare in general. They also confirm that the lack of labour sharing arrangements is not particular to the year surveyed.

[12] The data sections on labour sharing were completed for 1323 households of the remaining sample, which forms the basic sample for identifying matched networks.

The data were suitable for further analysis of networks because households were asked whether they had hosted a labour-sharing party in the main (*meher*) season, the number of people invited, and the identities of the people called. The names thus recorded were later matched against the sample roster. If the names matched a sampled household, the name was coded with the appropriate sample identification. If they were not part of the sample they were given a generic code. This in turn means that we have information on the total number of links of every household in the sample but not all the identities of their links. The sample per village is usually about 30% or more of the total population, and this has allowed us to obtain information on one or more of the network partners of most of the sampled households. The striking advantage of this matching process is that in large part we have complete information on the endowments, household structures, production and consumption of a sample of the partners in the labour sharing network[13]. This is the information that will prove crucial in describing network formation for labour sharing. We discuss the implications for our classification of network structures and for the econometric tests further below but first turn to a discussion of the important role that labour-sharing plays in production in these villages.

Table 1 provides some descriptive statistics on labour-sharing in the sample. We find that 64% of farmers were involved in a work party in this season. *Debo/wonfel* labour constituted about 16% of the labour supplied (in person-days) on the farm. In contrast, even though about a third of households used some hired labour, labour against cash or in-kind payment constituted only about 6% of total days supplied. The table also gives details by land-labour ratios (in terms of endowments). Here we notice that these labour transactions are not simply driven by endowment differences in terms of land-labour ratios. While there is some difference in participation and use by those with relatively less land, *debo/wonfel* labour shares in total labour are not very different across the land-labour distribution. Being reciprocal in nature, a *debo/wonfel* is not suitable for balancing land-labour ratios. This is reflected in our labour-sharing data: 83% of those who called a work party also worked for other farmers during the same season[14]. In the sample we found that the median size of a labour-sharing group is 6 people, although some are much larger. They mostly consist of other farmers in the village, often relatives (45% were relatives of the farmer).

[13] The villages in the study are often in remote settings and mimic island economies. However, even within villages that are closer to towns in this sample, labour-sharing (and other) transactions are almost entirely organised within the village so that the probability of capturing the nodes of the network within the village is high.

[14] A further inspection of the data suggested that actual factor use intensities were much more similar across the distribution of land-labour endowments, as one would expect with factor market transactions. Labour market and labour sharing are not responsible for it: it is largely achieved by sharecropping arrangements and adjustment in own family labour supply.

Table 1: Labour sharing and contribution to labour in agriculture

	0 - 0.19 ha/adult	0.19-0.62 ha/adult	$\geq$ 0.62 ha/adults	All
Involved in labour parties (%)	59	70	68	64
Shared labour (% of tot. labour)	13	18	16	16
Hired labour (% of tot. labour)	7	5	6	6

Network architectures are defined as symmetric if the number of links of one's partners is (approximately) equal to the number of links one has. We classify 128 networks as symmetric, of which 28 were perfectly symmetric in that the number of links of each of the partners was identical. Networks were also classified as symmetric if the difference between the total number of links of the household and the total number of links of each of its partners was systematically small. In particular, we include those households as symmetric if the highest difference in links with any of its partners was lower than the lowest number of links in that network, provided the network consisted of at least 4 partners[15]. Below, we report the impact of changing this definition. All other networks are treated as asymmetric[16]. Within asymmetric networks, we further distinguish hubs where households have more links than their neighbours; and spokes, who are linked through hubs and so have far fewer links than their (hub) neighbours. We contrast these households to those who reported that they had not called a labour sharing party. Table 2 provides a summary of average characteristics for the 1323 households in the sample. This includes 551 households for whom we have sufficient (matched) information from within the sample to identify their network architecture; 477 households with no labour-sharing links; 295 households with labour-sharing links (22% of the households) but for whom no partners could be identified within the sample and so remain unclassified.

Recall that our model predicts that homogeneous components are not only symmetric, but also complete (i.e. fully interconnected). For each household, we obtain a measurement of the level of interconnectedness among the household's neighbours through the clustering coefficient. This is defined as the ratio of existing links connecting each household's neighbors to each

[15] The limit of 4 partners is to avoid counting as symmetric small networks with unequal number of partners and therefore clearly asymmetric.

[16] We also attempted to map this to the type of arrangement specified by village but this was made more difficult by the plethora of local names and missing information on type. The local names for these arrangements vary by village and the classification is rendered particularly difficult because in non-Amhara villages, the local names are often transcribed in Amharic as *debo* even though, when asked to describe the structure, the description corresponds to the Amhara *wonfel*.

other to the maximum possible number of such links. The clustering coefficient ranges between 0 and 1: it is equal to one when all of the household's neighbours are also linked to each other, so that the local neighbourhood of the household is dense (or clustered); if the clustering coefficient is small, this implies that the local neighbourhood of the household is rather sparse, with only few neighbours linked to each other (if at all). A clustering coefficient of zero implies that the network is minimally connected, and there are no direct links between any two neighbours of the same household.

We find that households belonging to networks that we classify as symmetric have a very high clustering coefficient (0.90 on average); on the other hand, households belonging to networks that we classify as asymmetric present a significantly lower clustering, with an average of 0.5.

This finding suggests that symmetric structures also tend to be heavily clustered: households who belong to symmetric components rely on the same set of neighbours for labour-sharing. While providing additional evidence in support of our theoretical framework, this result also constitutes a further robustness check on our classification between symmetric and asymmetric structures: we are indeed classifying households in the direction suggested by the theory.

Beyond the number of links of the household, the characteristics included are possibly reflecting differences in own labour quality and other endowments affecting productivity. Age and sex of the household head are included, as well as the number of adults (above 15 years of age) and the average education of the adults in the household. The latter is measured in years, and as can be seen in table 2, education levels of adults are still extremely low - across the sample below 2 years per adult. A key characteristic for our purposes is a direct measure of labour quality and strength, based on questions related to activities of daily living (functionings). Each adult in the household was asked whether they could perform five basic tasks, scoring them from 1 (easily) to 4 (not at all). The tasks were: standing up after sitting down; sweeping the floor; walking for 5 km; carrying 20 liters of water for 20 meters; hoeing a field for a morning. A simple average score was rescaled to a score between 0 and 1 (best). The score used here is the average score per adult in the household[17]. Finally, other endowments are included, such as land per adult and livestock values per adult.

A simple comparison of the characteristics reported in table 2, using pairwise t-tests and 5 percent significance levels, reveals that those without links are different from all the other categories in that the household head is older, they have fewer adults and lower labour quality in terms of strength. Given that labour-sharing is a reciprocal arrangement, these households may simply not be able to participate in work parties. The differences between households across architectures are less pronounced, except in terms

[17] Gertler and Gruber use a similar index [21].

the number of links. Symmetric networks tend to be significantly smaller than asymmetric networks, but hubs have systematically more links than members of symmetric components, while spokes have less than either. It does appear that households in symmetric networks have significantly less land per adult than both hubs and spokes. A further check is to distinguish those classified as exactly symmetric and other symmetric. (Recall that the latter group was defined on the basis of those for whom the highest difference in links with any of its partners was lower than the lowest number of links of the household and of any of its partners). The former group is small - only 28 households, but in comparing their labour characteristics to the others, it was found that they were not significantly different for any reasonable significance levels. In the regression analysis below, further robustness tests are reported, to check whether the classification rule adopted here matters substantially in interpreting the results.

Network architectures are also not particularly linked to substantially higher labour contributions in particular tasks. Labour sharing networks in these villages were employed largely for weeding and harvesting with the share of *debo/wonfel* labour called for ploughing or construction being negligible. Labour from network members contributes slightly more labour as a proportion of total labour for symmetric networks (21 percent versus 18 percent), and the percentage of households using labour sharing groups for weeding is slightly higher in symmetric than in asymmetric groups (59 versus 49 percent), but the differences are not significant at 5 percent. Both types of architectures, symmetric and asymmetric are associated with similar labour allocations, and with both types of tasks (weeding and harvesting). The type of task does not seem to drive the choice of network architecture.

A further critical issue is the degree to which information missing on links of neighbours (where they are not part of the household sample) affects the classification into architectures. Note that architectures are determined by total number of links of the household and that of its partners in the arrangement. In most villages, over 30% of the village forms the sample and in some cases, about three-quarters of the village was surveyed, which allowed us to map at least two or more partners. Overall, we managed to identify network partners for more than two-thirds of the labour sharing groups. Due to the sampling procedure, small networks are more likely to be under-sampled, since the probability that at least one of the partners is in the sample is higher for larger work parties. This appears to be confirmed in the data, with a low mean (and median) of 3 partners for those networks for whom we have no information on partners' characteristics. Important for our purposes, t-tests of differences in mean characteristics in table 2 show that those households lacking information on network partners are not systematically different from any of the classified households. In terms of the number of links and land owned per adult, they are similar to symmetric groups (and different to the other groups), but not systematically similar to one of the groups in terms of other characteristics, suggesting (at least on the basis of observables) that no systematic bias is introduced by their exclusion[18].

In any case given the scope of our analysis, the issues with only having a (possibly non-random) sample of networks for analysis is less serious than may seem at first. We do not aim to predict which network will form in a village, not least since our theoretical framework suggests that multiple equilibria are possible and therefore makes no particular predictions about which particular network should be observed in a village, among a (random) sample of households. Our interest lies in testing predictions about the characterisation of any equilibrium that may have emerged, by examining the architecture of networks and the correlates of its membership.

5 Empirical Characterisation of Equilibria

The aim of this section is to examine the empirical implications of the theory and ask whether the characterisation of the equilibrium in Section 3 is borne out by the data. We have characterised labour sharing as the outcome of (pairwise stable) network formation with negative externalities, given our assumption about the value of output for farmers in a network. The return to *debo* labour a farmer is able to generate is increasing in the number of his own links but decreasing in the number of links his partners have. Recall again that the model allows for multiple equilibria so that it is not possible to perform a direct test of our network formation model: for any given initial distribution of endowments, multiple equilibrium configurations could emerge. Nevertheless, in this section, particular features of the theoretical characterisation of equilibria can be tested. The key testable result is that asymmetric networks are more heterogeneous in the quality of endowments, since heterogeneity in the quality of the network partners is only compatible with asymmetric networks.

We begin by exploring the link between heterogeneity in labour quality (and other related characteristics) and the architecture of these groups. Table 3 provides basic descriptive statistics about the partners in symmetric and asymmetric networks. We use a number of definitions of 'heterogeneity', to show the robustness of any interpretation on the basis of descriptive statistics, before exploring them further through multivariate analysis in the next section. In particular, we report the results using the following possible notions of heterogeneity: the standard deviation of each characteristic among network members; the coefficient of variation; the mean absolute deviation of each member's characteristics; the difference between the mean of

[18] There are also no significant differences in the tasks for which these groups are used nor the extent of labour allocation between the unidentified groups and the networks with matched data.

a characteristic across the network partners less the value of the characteristic for the 'least endowed' member (as implicit in the functional form in proposition 2). For comparison, the first column presents the mean of each characteristic across network partners. For each of these measures of heterogeneity, we conduct a simple t-test of whether symmetric and asymmetric networks differ in the level of heterogeneity: differences significant at 5 percent are marked by an asterisk. The table demonstrates that there are no systematic differences in the mean characteristics between the two types of architecture, except that symmetric networks appear to consist of members with less land on average. There is no straightforward explanation for this, but it suggests that any multivariate exploration should control for mean land areas per adult. In terms of the different definitions of heterogeneity, the results are strong and robust. First, consider two of the characteristics reflecting labour quality: education and functionings (strength). For all measures, symmetric networks have lower heterogeneity, exactly as predicted by the theory. Furthermore, in all cases, the heterogeneity in the age of the head and whether the network consists of male headed households, is lower for symmetric networks, and mostly significantly so. Finally, the heterogeneity in two 'wealth' related endowment characteristics, land and livestock per adult, is also lower for symmetric networks. These could possibly reflect past higher quality of labour, or, as in nutrition-productivity relationships, feeding into current higher quality of labour, and therefore again consistent with our narrative. Only in the case of adults in the household, is the heterogeneity (just about) significantly larger in symmetric networks, but then only when using the standard deviation as a measure and not in any of the other measures of heterogeneity. Overall, the results are strongly suggestive of the theoretical setup and this is explored further in the regression analysis below.

The regressions in table 4 explore network architecture, assessing the role of heterogeneity amongst partners, controlling for other endowments and characteristics. Furthermore, since each village may have converged to a particular equilibrium out of many, it appears important to assess these correlations controlling for village fixed effects. Given that the theory suggests a trade-off between heterogeneity in quality and endowments on the one hand, and symmetry of the network architecture on the other, we first present a simple probit regression model, denoting by 1 that a household is in a symmetric network, and zero if it is in an asymmetric network, to check whether this trade-off is observed. The sample consists of the 551 households with matched data on partners. The advantage of exploring this further in a multivariate framework is our ability to offer some controls for a number of household characteristics and community fixed effects, thereby ensuring that the correlations observed in table 3 are not simply driven by these observables, and then possibly unrelated to issues of heterogeneity.

The variables used are household characteristics as in table 2: characteristics of the head (age and whether male), labour quantity and quality (number of adults in the households, the average number of years of education among the adults and the average functionings score per adult) and wealth characteristics, in particular land per adult and livestock holdings per adult. The latter is by far the most important liquid asset while other assets such as durables constitute only a very small proportion of total wealth. The same characteristics, but now in terms of the heterogeneity among all the network partners, are also included (as in table 3). Note that all these variables may have a bearing on labour quality, directly or (as in the case of wealth) indirectly (for example, wealth accumulated linked to being a 'stronger' farmer). However, the most direct measure of labour quality is the functionings score, which is thus the main variable of interest. Finally, the regressions also control for mean characteristics of the network partners in terms of land per adult, thus purging any technology effects of land-labour ratios. Including other mean characteristics of network partners left the results qualitatively unaffected.

The regressions (in Table 4) were run for a number of different measures of heterogeneity as in table 3. Alternative regressions, also including levels and heterogeneity in other characteristics of the head (education, functionings) or measures of adult labour, separating males from females, gave similar results. The regressions clearly confirm the findings from table 3. There is significant negative correlation between the heterogeneity in labour quality, measured in terms of functionings among all network partners, and the probability of being in a symmetric network, irrespective of the measure of heterogeneity. A similar negative correlation is observed for heterogeneity amongst network partners in land and livestock holdings, which, as was argued before, could be a reflection of labour quality differences. The regressions find significant effects for some of the controls, most notably own livestock holdings and mean land holdings of network partners (as well as the community fixed effects which are not reported). There is some positive correlation with heterogeneity in terms of number of adults in each partner household but this is insignificant in all the specifications. However, most important for our purposes is that the negative impact of heterogeneity in labour quality is maintained despite these controls.

An issue of concern is whether the classification into symmetric and asymmetric groups is measured with error, since some larger groups were classified as symmetric if the number of links between partners is approximately the same, rather than exactly so. Two robustness tests are presented in Table 5. The first check is to look at the determinants of having differences in links, effectively studying the correlates of increasing deviations from symmetry. The left hand side is the mean difference in number of links between the household and its partners - a measure of increasing asymmetry, bounded at zero, with pure symmetric groups having no difference at all. A tobit model is used in the estimation. Under the null that heterogeneity among partners reduces the probability of symmetry, this would suggest a positive correlation between the left hand side and different measures of heterogeneity. The second robustness check, in the last 3 columns of Table 5, compares the relatively small group (28 networks) of strictly symmetric networks where linked farmers have exactly the same number of links with the rest of the sample. We use a probit model where the value 1 denotes strictly symmetric networks. In both regressions, the right hand side variables are as the same as before.

These results appear to confirm the earlier findings, with the signs of the variables of interest as before, even though not always as significant. Nevertheless, using mean difference in links, heterogeneity in labour quality characteristics such as education and functionings as well as male headship and land areas among network partners are positively correlated with higher asymmetry, showing the robustness of the results in Tables 3 and 4. Even running a probit regression, using a base group of 28 in a sample of close to 400 groups still gives significant negative correlations between symmetry and heterogeneity, related to male headship and land, while negative (but insignificant) coefficients of similar order of magnitude as in table 4 can be found for functionings and education[19].

This section has shown that some predictions can be made from theory about network architectures and the composition of networks, beyond issues of network size. The underlying idea that network architectures matter is next used to illuminate discussions of the impact of networks on economic performance.

6 Empirical Illustration: The role of architecture in performance

So far, we have described the relationship between network structures and the distribution of endowments within these structures. But does network architecture matter for outcomes? Given that households have incentives to form labour sharing links, it is useful to explore the impact of networks on observed economic performance (in terms of output generated by these farmers). In this section, we aim to show, in an empirical examination of the relationship between outcomes and networks, not only that network links matter but more particularly that network architecture matters. The main lesson from this is that a narrow focus on the number of links (as the approach taken in Putnam [37]) may lead to biased results. The sample suitable for this analysis contains those households classified as in either symmetric or asymmetric networks for which we have information (i.e. the

[19] The sample size is smaller than in table 4 because in some villages no symmetric groups defined as in these regressions can be observed so that these observations were dropped given the community fixed effects formulation used.

sample used in the previous section).

The key issue of interest is to show that in this particular sample, network architecture is correlated with performance, controlling for the number of links. Furthermore, we show that inference on the impact of the social networks will be incorrect if network architecture is not included in the regression. The regression is a standard (Cobb-Douglas) production function augmented for land quality, human capital, further household characteristics and network variables. Our focus here is on the network variables. The total value of the harvest in the main season is regressed on land and adult labour, all expressed in logarithms. Adult labour is included both as levels of labour available in the household and as a weighted average of adults with the weights reflecting their functioning scores, thus providing simple labour "efficiency" units. In particular, recall that the functionings index is the score, scaled between 0 and 1, with one if the individual can perform all tasks of daily living without any problem, and zero if none can be performed, providing a simple weight for counting adults. Other labour characteristics are also included, such as the age of the head and the average education (in years) of the adults in the household. Given the crucial role of oxen in the farming system and the possible contribution of manure, the value of livestock is also included in the regression. Land quality is different across farmers, and the survey data contains detailed information on both slope and quality of land, using well-defined local characterisations. Thus the share of land of each quality and slope is included in the specification. In all regressions (see table 6), land and quality weighted labour are strongly significant. Once quality corrected adult labour is included, the total number of adults is not significant. Male headship increases returns, as does higher land quality and both are significant (with higher returns for 'lem' land, which denotes the best land using local self-reported characterisations). Finally, all regressions control for village fixed effects. All these effects are robust across the different specifications.

Turning to the impact of networks, we use two different specifications. The first specification includes the (log of the) number of links and whether the architecture is symmetric or asymmetric to examine the impact of networks. In the second specification we explore the impact of refining architectures further by distinguishing hubs from spokes among the asymmetric networks, by including a dummy for each of these two, and defining 'symmetric' as the base group.

When addressing the role of networks on outcomes, the issue of endogeneity is vital. Until recently, the empirical literature on the impact of social capital (or more narrowly, network membership) on outcomes has largely assumed that networks are exogenous. For networks defined by ethnicity or gender this might be appropriate. However, in other contexts, individuals do choose the agents they want to interact with, so that the network structure is part of the outcome that we aim to explain. To the extent that the formation of networks is endogenous to its function, it is clear that empirical studies might be misleading about the impact of being in a network on outcomes.

From an empirical point of view, our interest lies not only in reconciling the particular network structures that emerge with a model of endogenous network formation but also in examining whether the endogeneity of the network formation process has bite in determining outcomes of interest. To put it simply, the issue is not only whether one believes that those with higher productivity, for instance, choose to be in a particular network but also whether the unobservable components of such a decision are correlated strongly enough with outcomes for this to have an econometric impact.

The main econometric concerns in estimating the effects of membership in a labour sharing network on productivity are akin to the general problem of estimating the effects of belonging to some 'group' on individual outcomes. For instance, estimates of the effects on educational attainment due to peers are bedevilled by the difficulty of identifying how much of the variation in attainment is actually due to peer effects and how much is driven by differences in individual characteristics. The main critique is that group (or network) characteristics are endogenous (or correlated with the errors) in such regressions. In particular, if members of a labour sharing network have chosen to come together precisely because they generate a certain synergy in weeding or harvesting as a group, then taking the amount of shared labour as exogenous will obtain biased estimates of the impact of such labour on output.

Instruments used for networks are a number of variables which are unlikely to influence productivity directly, but that may matter for network formation. We include first the number of close blood relatives living in the village (on average each household has about 2 blood relatives in the village), whether the household head was born in the village (just over half of the heads were born in the village), the average number of years of residence of the head of household in the village[20]. These variables may all suggest how strongly the household is embedded in the village and reflect its relative role and position in the village. Furthermore, since network formation ought to be influenced by the pool of available partners for local networks, we also include locational fixed effects based on the neighbourhood. In particular, in each village, we identified (village-defined) geographical neighbourhoods, and checked in which quarters/neighbourhoods each household lived. We identified about 70 neighbourhoods in total across the 15 clusters in our sample. These neighbourhood fixed effects may also reflect local land quality, but this is controlled for directly at the household level using the land qual-

[20] In rural Ethiopia, mobility is low because rights in land are user rights granted by the Government. Leaving the village will lose one one's entitlement to land, so in general years of residence differ from age mainly because of exogenous circumstances.

ity variables[21]. Note that there are exactly two endogenous variables and so our minimal set consists of the number of blood relatives and the neighbourhood effects. The variables on birth and years of residence are used for the overidentification tests but it must be emphasised that the results are unaffected by this small expansion in the number of instruments.

Obtaining credible instruments is not an easy task, and we test whether the instruments used are both relevant and valid. We present 3 regressions in Table 6: the first column provides OLS estimates of parameters, using links and architecture. This can be compared to the last two columns that provide the estimates for the impact of networks, where both links and architecture are instrumented for. Column 2 thus presents IV estimates for links and whether the architecture is asymmetric, while Column 3 examines the impact of describing architectures more precisely as hubs and spokes, relative to being symmetric. The instrumentation employs a two-step efficient GMM[22] estimator, in preference to the standard 2SLS, mainly because the endogenous variables are discrete in nature and while the 2SLS estimator remains consistent, one is concerned with the likelihood of heteroscedasticity in this situation. If heteroscedasticity is present, the GMM estimator is more efficient than the simple IV - and if not, is no worse. The tests suggest that heteroscedasticity is a concern and hence the GMM results are preferred, but the standard IV estimates provide similar point estimates of the coefficients. The tests for heteroscedasticity report the Pagan-Hall [36] test statistics and suggest heteroscedasticity in all specifications.

Are the instruments relevant? The F statistics on the first stage regressions (presented in Table 6A and 6B) are highly significant with p-values of 0.03 for both the number of links and architecture. Since, we have multiple endogenous regressors, we present the Shea partial correlation coefficients that take the intercorrelations between instruments into account, as well as the standard R^2 and F-test of the joint significance. If both the Shea and the standard partial correlation coefficients yield similar results, then instruments are more likely to be relevant. The Shea and the standard R^2 are respectively 0.13 and 0.12 for the number of links, 0.18 and 0.19 for symmetric architectures, and 0.15 and 0.14 for both spokes and hubs, suggesting that the instruments are undoubtedly relevant. For comparison, the first stage-F statistics are significant for the number of links with p-values of 0.03 and for architectures, with p-values of 0.004.

Are the instruments valid? Are the instruments orthogonal to the error term? The results of Hansen's J-test in the context of the GMM estimates

[21] The total sample size reduces to 404 because some households could not be fully identified in terms of neighbourhoods. However, the uninstrumented regressions on the full sample yielded virtually identical results so the loss of these observations is not likely to be a cause for major concern.

[22] This estimator is also referred to as the heteroscedastic two-stage least squares (Davidson and Mackinnon, [10] p. 599).

(the value of GMM objective function evaluated at the efficient estimator) in both sets of regressions suggest that the null hypothesis that the instruments are valid cannot be rejected, with p-values of 0.20.

Note that we have taken care to restrict the instrument set to variables that capture one's own position in the village and general characteristics of the neighbourhood from where potential partners might be drawn rather than specific features of the network one belongs to. Hence we believe they are plausible.

We now turn to an examination of the coefficients in each regression. The results in table 6 provide interesting reading on the relevance of both number of connections and network architecture. The first column shows that links (size) matter: more links provide a higher return, with one percent more links giving 16 percent more returns while the measure of architecture as in whether the network is asymmetric suggests that it has a positive and significant effect, with similar returns. However, after instrumenting, these effects stay strongly significant, while the impact of links almost trebles to about 0.4. The impact of asymmetry is higher, offering about 43% higher returns in output: asymmetric networks in this sample (controlling for total links, and village fixed effects) lead to substantially higher returns. Architecture matters, over and above the number of connections. This is confirmed by exploring this further, by distinguishing hubs from spokes among asymmetric networks (with symmetry the base category as before). Without instrumenting, the pattern is that only hubs are significant. But after instrumenting, both the hub and spoke dummies are significant, as are total links. The size of the coefficients on hubs and spokes are similar and indeed not significantly different (a chi-square test of difference fails to reject the null hypothesis of no difference, at a p-value of 0.7), suggesting that the specification in column 3 is the parsimonious specification: asymmetry or not appears to matter most in this sample.

From a theoretical perspective it is very hard to compare the payoff that a farmer would get in symmetric versus asymmetric networks. It is so because of the very large number of asymmetric networks which are feasible. However we note that a household who belongs to a work-party that is totally asymmetric (a star) would be strictly better off as a spoke in the star rather than as a member of a similarly sized symmetric network, as long as the hub is sufficiently well endowed. Also, as emphasised earlier, the tendency to connect more than it is optimal is mitigated in asymmetric networks by the fact that the spoke's willingness to join in the work-party decreases rapidly with the size of the component. As a result, the inefficiencies caused by the presence of negative network externalities may well be lower in asymmetric compared to symmetric networks.

From an empirical perspective, it is clear that instrumentation raises the coefficients on the network variables, suggesting that the unobservable components of the decision to sort into particular kinds of networks are negatively correlated with the network variables. It is interesting to speculate (if hard to establish) what these unobservable components might be. One aspect of network structures that is tradionally seen as difficult to capture is the notion of social cohesion[23]. Definitions of this in the sociological literature seem to share the common intuitive core resting on how well a group is held together Moody and White [31] define a group as cohesive to the extent that multiple independent social relations among multiple members of the group hold it together. They go on to say that group cohesion might be captured by the number of independent paths linking each pair of agents in the group. Hence it appears that this sociological notion of cohesion might be related to the degree of connectedness of a graph. A connected graph (i.e. a component within a network) can range from being completely connected, where all agents are directly linked to every other agent, to minimally connected, where all agents are indirectly connected through the minimum number of links and there is a single path connecting any two agents. An example of minimally connected graph is the star network. Hence the star can be thought of as possessing a low level of social cohesion, since if any of the links fail, the network breaks up into more than one component. On the other hand, the complete network, where all the agents are connected to each other, displays very high social cohesion since the group is held together even if more than one link fails. More generally, symmetric network architectures (of which the complete network is an example) are never minimally connected, (with the notable exception of the pair) - and hence possess a higher level of social cohesion than most asymmetric architectures. Futhermore, the homogeneity displayed within symmetric structures may also be thought as consonant with the notion of cohesion. In short, we speculate that cohesion, which is an intrinsic attribute of network structure might well be the unobservable element, negatively correlated with asymmetry and possibly driving the under-estimation of the impact of links and asymmetric structures on outcomes before instrumentation.

The main conclusion from this section is in line with the rest of the paper. The results confirm strongly that the architecture of a social network, and not just number of links, has an important role to play in understanding network formation, and the role of social networks on economic performance.

7 Concluding Remarks

We provide a theoretical framework of endogenous network formation that yields testable predictions for the network architectures generated by labour-sharing groups in village economies. We use data from rural Ethiopia to test the consistency of the empirical evidence with the equilibrium characterisa-

[23]As Durkheim points out [14]"...social solidarity is a wholly moral phenomenon which by itself is not amenable to exact observation and especially not to measurement".

tions predicted by our model.

The hypothesis is that network formation for the purpose of labour sharing is driven by negative network externalities: popular network partners are also less likely to be available when you need them. We allow for heterogeneity amongst agents and capture this empirically by using a direct measure of physical ability and strength based on questions related the ease of performing five basic farming tasks.

As predicted, we find that heterogeneity in quality is associated with asymmetric network architectures, while homogeneity is associated with symmetric structures where network partners have similar number of links. Furthermore, we examine the impact of networks on output and, in contrast to the current empirical literature, we demonstrate the critical role of both number of links and architecture in determining the impact of social networks on performance.A narrow focus on links serves to underestimate the impact of labour-sharing on output. Social capital matters certainly, but its impact differs by the architecture of the network to which one belongs.

Appendix

Proof of Proposition 1. By contradiction. Suppose $C(g)$ is not symmetric. Without loss of generality, assume that $i \in C(g)$ has the largest number of links of all members of the component: $n_i = \max_{i'} \{n_{i'} \mid i' \in C(g)\}$. If $C(g)$ is not symmetric, then there exists a $j \in C(g)$ such that $ij \in g$ and such that $n_j < n_i$. This implies that $\exists k \in C(g)$ such that $ik \in g$ and $jk \notin g$. Since by assumption g is pairwise stable and $ik \in g$, then

$$h_k(g) \geq h_k(g - ik)$$

which implies a fortiori that

$$h_k(g + jk) > h_k(g)$$

in fact if k values his link to i, he should value even more so a link to j, given that $n_i > n_j$. More in detail:

$$h_k(g) \geq h_k(g - ik) \Longleftrightarrow \frac{n_k + 1 + q}{n_i + 1} \geq \frac{1}{n_k} \left[1 + \sum_{\substack{h:hk \in g \\ h \neq i}} \frac{q}{n_h + 1} \right]$$

while $h_k(g + jk) > h_k(g)$ requires:

$$\frac{n_k + q + 2}{n_j + 1} > \frac{1}{n_k + 1} \left[1 + \sum_{\substack{h:hk \in g \\ h \neq i}} \frac{q}{n_h + 1} + \frac{q}{n_i + 1} \right]$$

Since $n_j < n_i$ we know that

$$\frac{n_k+q+2}{n_j+1} > \frac{n_k+q+2}{n_i+1}$$

moreover

$$\frac{1}{n_k} + \frac{1}{n_k}\sum_{\substack{h:hk\in g \\ h\neq i}} \frac{q}{n_h+1} > \frac{1}{n_k+1} + \frac{1}{n_k+1}\left[\sum_{\substack{h:hk\in g \\ h\neq i}} \frac{q}{n_h+1} + \frac{q}{n_i+1}\right]$$

in fact $\frac{1}{n_k} > \frac{1}{n_k+1}$, moreover the expression

$$\frac{1}{n_k}\sum_{\substack{h:hk\in g \\ h\neq i}} \frac{1}{n_h+1}$$

is an average of n_k fractions and the expression

$$\frac{1}{n_k+1}\left[\sum_{\substack{h:hk\in g \\ h\neq i}} \frac{1}{n_h+1} + \frac{1}{n_i+1}\right]$$

can be interpreted as the average of (n_k+1) fractions, where the first n_k elements are the same as in the lhs and the extra element is the smallest and hence reduces the average. As a result:

$$h_k(g) \geq h_k(g-ik) \Rightarrow h_k(g+kj) > h_k(g)$$

Similarly, since g is pairwise stable by assumption and $ij \in g$, we know that

$$h_j(g) \geq h_j(g-ij)$$

We can show that, given that $n_k \leq n_i$ then

$$h_j(g) \geq h_j(g-ij) \Rightarrow h_j(g+kj) \geq h_j(g)$$

i.e. if j values his link to i he should value (at least as much) a link to k (the complete proof is very similar to the one above and it is therefore omitted). Hence $jk \notin g$ contradicts that g is pairwise stable. ■

Proof of Corollary 1. By contradiction. Suppose $C(g)$ is not complete. Then there exist $i, j \in C(g)$ such that $ij \notin g$. By pairwise stability this requires that at least one of the two following inequalities hold: either

$$h_i(g) > h_i(g+ij)$$

or

$$h_j(g) > h_j(g+ij)$$

The first inequality requires:

$$\frac{q_i}{1+n_i}+(1+\frac{q_i}{1+n_i})\sum_{k:ik\in g}\frac{q_k}{1+n_k} > \frac{q_i}{2+n_i}+(1+\frac{q_i}{2+n_i})\left(\frac{q_j}{2+n_j}+\sum_{k:ik\in g}\frac{q_k}{1+n_k}\right) \tag{3}$$

The second inequality requires:

$$\frac{q_j}{1+n_j}+(1+\frac{q_j}{1+n_j})\sum_{h:jh\in g}\frac{q_h}{1+n_h} > \frac{q_j}{2+n_j}+(1+\frac{q_j}{2+n_j})\left(\frac{q_i}{2+n_i}+\sum_{h:jh\in g}\frac{q_h}{1+n_h}\right) \tag{4}$$

By homogeneity of $C(g)$, $\forall i, q_i = q$. By proposition 1, $\forall i, n_i = n$. Hence both (3) and (4) can be rewritten as:

$$\frac{q}{1+n}+(1+\frac{q}{1+n})\sum_{k:ik\in g}\frac{q}{1+n} > \frac{q}{2+n}+(1+\frac{q}{2+n})\left(\frac{q}{2+n}+\sum_{k:ik\in g}\frac{q}{1+n}\right)$$

This inequality requires:

$$2n+3n^2+n^3+q<0$$

which cannot be verified. ■

Proof of Corollary 2. This is an immediate corollary of proposition 1. By contradiction, suppose that $\forall i, j \in C(g)$ such that $ij \in g$, $q_i = q_j$, then by proposition 1 it would follow that $C(g)$ is symmetric. ■

Proof of Proposition 2. $C(g)$ is symmetric by assumption, so that $\forall i, j \in C(g)$ such that $ij \in g$, $n_i = n_j = n$. Fix $\widehat{j}$ such that $\widehat{j} = \arg\min_{j\in C(g)}\{q_j\}$ and consider some i such that $i\widehat{j} \in g$. Pairwise stability of g, requires that $\forall j : ij \in g$

$$h_i(g) \geq h_i(g-ij) \Longleftrightarrow \frac{q_j}{n_j+1}+\frac{q_i}{n_i+1}\left[\frac{q_j}{n_j+1}-\frac{1}{n_i}\left(1+\sum_{\substack{k:ik\in g\\ k\neq j}}\frac{q_k}{n_k+1}\right)\right] \geq 0 \tag{5}$$

By symmetry of $C(g)$, the inequality above becomes:

$$q_j+q_i\left[\frac{q_j}{n+1}-\frac{1}{n}\left(1+\sum_{\substack{k:ik\in g\\ k\neq j}}\frac{q_k}{n+1}\right)\right] \geq 0$$

which can be rewritten as

$$\frac{q_j}{q_i} \geq \frac{1}{n} + \frac{1}{n} \sum_{\substack{k:ik \in g \\ k \neq j}} \frac{q_k}{n+1} - \frac{q_j}{n+1}$$

and by adding and subtracting $-\frac{1}{n+1}\frac{q_j}{n}$:

$$\frac{q_j}{q_i} - \frac{1}{n} \geq \frac{1}{n+1} \left[\sum_{\substack{k:ik \in g \\ k \neq j}} \frac{q_k}{n} + \frac{q_j}{n} \right] - \frac{1}{n+1}\frac{q_j}{n} - \frac{q_j}{n+1}$$

which becomes

$$\frac{q_j}{q_i} - \frac{1}{n} \geq \frac{1}{n+1} \left(\overline{q}_i - q_j \right) - \frac{1}{n+1}\frac{q_j}{n} \tag{6}$$

where $\overline{q}_i$ denotes the average quality of agent i's neighbours (including j):

$$\overline{q}_i = \sum_{k:ik \in g} \frac{q_k}{n}$$

From (6) it follows that:

$$\overline{q}_i - q_j \leq (n+1) \left(\frac{q_j}{q_i} - \frac{1}{n} + \frac{q_j}{n\,(n+1)} \right)$$

$$\overline{q}_i - q_j \leq (n+1) \left[\frac{q_j}{q_i} - \frac{n+1-q_j}{n(n+1)} \right]$$

In particular for $\widehat{j}$:

$$\begin{aligned} \overline{q}_i - q_{\widehat{j}} &\leq (n+1) \left[\frac{q_{\widehat{j}}}{q_i} - \frac{n+1-q_{\widehat{j}}}{n(n+1)} \right] \\ &< n + \frac{q_{\widehat{j}} - 1}{n} \end{aligned}$$

Hence the (absolute value of the) deviation of the least endowed neighbour from the average neighbour of i (in the lhs) is bounded above by the degree of the symmetric component (n) and by how the endowment of the weakest neighbour of i compares to 1 (which is the minimum value that we assume the endowment of a farmer can take). Notice that given that $\widehat{j}$ is the weakest farmer in the component, a large value of $q_{\widehat{j}}$ is only found in high-quality components. ■

References

[1] Bala, V. and S. Goyal (2000): "A Non-Cooperative Model of Network Formation", *Econometrica*, vol. 68 (5), pp. 1181-1231.

[2] Bala, V. and S. Goyal (2000): "A strategic analysis of network reliability, *Review of Economic Design*, vol. 5 (3), pp. 205-228.

[3] Durlauf, S. and M. Fafchamps (2005): "Social Capital", in Handbook of Economic Growth, S. Durlauf and P. Aghion eds, Elsevier.

[4] Bandiera, O. and I. Rasul (2006): "Social Networks and Technology Adoption in Northern Mozambique", *Economic Journal,* forthcoming.

[5] Besley, T, .S. Coate and G. Loury (1993): "The Economics of Rotating Savings and Credit Associations", *American Economic Review*, 83(4), pp. 792-810.

[6] Bevan, P. and A. Pankhurst (1996): "Ethiopian Village Studies", Centre for the Study of African Economies, Oxford University.

[7] Brock, W and S. Durlauf (2001): "Interactions-Based Models", in Handbook of Econometrics, vol. 5, J. Heckman and E. Leamer eds.

[8] Coleman, J., E. Katz and H. Menzel (1957): "The Diffusion of an Innovation among Physicians", *Sociometry* 20, pp. 253-270.

[9] Conley, T. and C. Udry (2002): "Learning about New Technology: Pineapple in Ghana", Working Paper, Yale University.

[10] Davidson, R. and J.G. MacKinnon (1993): Estimation and Inference in Econometrics, Oxford University Press, 2nd Edition, Oxford.

[11] Deninger, K., Jin, S., Gebre-Selassie, S. and Mulat, D.(2003). "Market and Non-Market Transfers of Land in Ethiopia: Implications for Efficiency, Equity and Non-Farm Development", Working paper, World Bank. Washington, U.S.A.

[12] De Weerdt, J.(2002): "Risk-Sharing and Endogenous Network Formation", Discussion Paper 2002/57, WIDER, Helsinki.

[13] De Weerdt, J. and S. Dercon (2006): "Risk-Sharing Networks and Insurance Against Illness", *Journal of Development Economics*, forthcoming.

[14] Durkheim, E. (1893): The Division of Labor in Society. Trans: A.D. Halls, 1984. New York: The Free Press.

[15] Erasmus, G.J, (1956): “Culture, Structure and Process: The Occurrence and Disappearance of Reciprocal Farm Labor”, *Southwestern Journal of Anthropology* 12 (4).

[16] Fafchamps, M. and S. Lund (2003): “Risk Sharing Networks in Rural Philippines”, *Journal of Development Economics* 71, pp. 261-287.

[17] Fafchamps, M. (2006): “Social Capital and Development”, *Journal of Development Studies*, forthcoming.

[18] Fafchamps, M., and B. Minten (2002): “Returns to Social Network Capital Among Traders”, *Oxford Economic Papers* 54, pp. 173-206.

[19] Foster, A. and M. Rosenzweig (1995): “Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture”, *Journal of Political Economy*, 103(6), pp. 1176-1209.

[20] Galeotti, A. S. Goyal and J. Kamphorst.(2006) "Network formation with Heterogeneous Players", *Games and Economic Behavior*, 54(2), 353-372..

[21] Gertler, Paul, and Jonathan Gruber (2002) "Insuring Consumption Against Illness", *American Economic Review,* 92(1), pp. 51-70.

[22] Glaeser E.L., D. Laibson and B. Sacerdote (2000): "The Economic Approach to Social Capital", NBER Working Paper No. 7728.

[23] Granovetter, M. (2005): "The Impact of Social Structure on Economic Outcomes", *Journal of Economic Perspectives,* vol. 19 (4), Winter. pp. 35-50.

[24] Grimard, F. (1997): “Household Consumption Smoothing through Ethnic Ties: Evidence from Côte d’Ivoire” *Journal of Development Economics*, vol. 53, pp. 391-422.

[25] Goyal, S. and S. Joshi (2005): “Unequal Connections”, mimeo.

[26] Haller, H. and S.Sarangi (2005): “Nash Networks with Heterogeneous Agents”, *Mathematical Social Sciences*, 72, pp. 181-201.

[27] Jackson, M.O. (2005) “The Economics of Social Networks,” July 2005. Proceedings of the 9th World Congress of the Econometric Society, edited by Richard Blundell, Whitney Newey, and Torsten Persson, Cambridge University Press, forthcoming.

[28] Jackson, M. O. and A.Wolinsky (1996): “A Strategic Model of Social and Economic Networks”, *Journal of Economic Theory*, 71 (1), pp. 44-74.

[29] Johnson,C. and R.P. Gilles (2000): “Spatial Social Networks”, *Review of Economic Design* 5 (3), pp. 273-299.

[30] Manski, C.(1993): “Identification of Endogenous Social Effects: The Reflection Problem”, *Review of Economic Studies*, 60(3), pp. 531-542

[31] Moody, J. and White, D.R. (2003): "Social Cohesion and Embeddedness", *American Sociological Review* 68:103-127

[32] Moore, M.P.(1975): “Co-operative Labour in Peasant Agriculture”, *Journal of Peasant Studies*, pp. 271-291.

[33] Munshi, K. and M. Rosenzweig (2006):"Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy" *American Economic Review* forthcoming.

[34] Luke, N. and K. Munshi (2006):"New Roles for Marriage in Urban Africa : Kinship Networks and the Labor Market in Kenya .” *Review of Economics and Statistics* forthcoming.

[35] Nigussie, T and D. Mulat (2003), “The Productivity and Profitability of Wheat and Teff Technologies: Comparison between Extension and Non-Extension Plots in Three Selected Villages of Ethiopia”, *Technological Progress in Ethiopian Agriculture, Proceedings of the National Workshop on Technological Progress oin Ethiopian Agriculture,* Department of Economics, Addis Ababa University, Nov 29-30 2001.

[36] Pagan, A.R. and D. Hall (1983): "Diagnostic Tests as Residual Analysis", *Econometric Reviews* 2(2): pp. 159-218.

[37] Putnam, R.D. (2000): Bowling Alone: The Collapse and Revival of American Community New York: Simon & Schuster

[38] Townsend, R. (1994): “Risk and Insurance in Village India”, *Econometrica*, 62(3), pp. 539-91.

[39] Udry, C.R. and T.G. Conley (2004): “Social Networks in Ghana”, mimeo.

Table 2: Characteristics of the household by network architecture

	no links		symmetric		asymmetric networks: spoke		asymmetric networks: hub		no information	
number of observation	477		128		272		151		295	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
own links	0.00	0.00	6.13	2.78	3.40	2.54	12.06	8.47	4.75	9.57
male headed?	0.14	0.15	0.16	0.11	0.17	0.14	0.15	0.10	0.14	0.12
age head?	49.73	16.96	44.62	13.65	43.26	14.01	45.60	15.18	45.43	15.82
adults	2.84	1.82	3.33	2.05	3.06	1.66	3.50	1.91	3.13	1.98
education in years per adult	1.47	2.20	1.30	1.89	1.38	2.09	1.51	2.14	1.73	2.37
functionings score per adult	2.76	0.43	2.84	0.27	2.83	0.33	2.79	0.30	2.82	0.28
livestock per adult (birr)	448.42	725.26	412.02	506.07	407.19	492.86	488.65	531.97	357.99	476.67
land per adult	0.33	0.42	0.31	0.28	0.39	0.34	0.38	0.32	0.29	0.32

Note: own links are the number of network connections of the household. Average education in years per adult in the household. The functionings score is based on the average score on activities of daily living questions, scaled between 0 and 1, with 1 denoting best. Symmetric networks are defined as those networks in which each partner has either exactly or approximately the same number of links as the other members of the network. See main text for details. Spokes are networks in which the network partners have more partners than the household, while hubs denote those households who have more links than their partners. No information households are those households involved in labour sharing networks whose partners could not be matched with other households in the sample.

Table 3 Characteristics of network partners in labour sharing arrangements

	Means across partners			Standard deviation across partners			Coefficient of variation across partners			Mean absolute deviation from mean across partners			Bound (deviation from mean of 'lowest' partner)		
	Symmetric?		Diff?	Symmetric?		Diff?	Symmetric?		Diff?	Symmetric?		Diff?	Symmetric?		Diff
	No	Yes		No	Yes		No	Yes		No	Yes		No	Yes	
male head?	0.16	0.16		0.09	0.07	*	0.55	0.49		0.06	0.05	*	0.07	0.06	*
livestock per adult	437.72	407.99		299.09	242.65	*	0.79	0.64	*	216.30	175.77	*	255.05	188.97	*
land per adult	0.40	0.32	*	0.19	0.13	*	0.53	0.45	*	0.13	0.09	*	0.16	0.11	*
adults	3.28	3.45		1.20	1.38	*	0.36	0.36		0.88	0.99		1.03	1.06	
Agehead	44.12	44.55		11.14	10.32		0.25	0.23		8.17	7.41		9.90	8.41	*
education years per adult	1.40	1.33		1.32	1.07	*	1.25	1.11	*	0.96	0.78	*	0.99	0.79	*
functionings score per adult	2.81	2.83		0.21	0.17	*	0.08	0.06	*	0.16	0.12	*	0.24	0.15	*

Note: each column gives the indicator for either symmetric ('yes') or asymmetric ('no') network components. The 'diff' column simply reports whether a simple t-test of the difference in the mean between the two groups is significant at 5 percent or less. Variables are defined on the same basis as in table 2.

T able 4 Multivariate correlations: probit model of network architecture for different measures of heterogeneity (n=491)
Dependant variable = 1 if structure is asymmetric, 0 otherwise.

	Standard Deviation	Mean Absolute Deviation	Bound of 'poorest' partners
male headed household?	1.349	1.404	1.433
	[1.84]*	[1.91]*	[1.53]
number of adults in household	-0.018	-0.013	0.004
	[0.44]	[0.31]	[0.09]
years of education per adult in own household	-0.017	-0.016	-0.016
	[0.36]	[0.34]	[0.32]
Functionings score per adult in household	0.132	0.111	-0.028
	[0.44]	[0.36]	[0.07]
land per adult in household	-0.815	-0.808	-0.742
	[1.96]**	[1.92]*	[1.74]*
livestock value per adult in household	0.000	0.000	0.000
	[1.83]*	[1.89]*	[2.13]**
heterogeneity in male headship	-2.054	-2.802	-2.369
	[1.78]*	[1.82]*	[1.26]
heterogeneity in number of adults	0.128	0.151	0.115
	[1.93]*	[1.59]	[1.62]
heterogeneity in education	-0.129	-0.185	-0.134
	[1.36]	[1.48]	[1.17]
heterogeneity in functionings	-0.624	-0.929	-1.152
	[1.77]*	[1.99]**	[3.21]***
heterogeneity in livestock values	-0.001	-0.001	-0.001
	[2.40]**	[2.48]**	[2.37]**
heterogeneity in land areas	-1.402	-2.138	-2.083
	[2.96]***	[3.35]***	[3.62]***
mean land among partners	1.796	1.862	1.945
	[2.46]**	[2.44]**	[2.36]**
Pseudo R-squared	0.12	0.12	0.12

Regression controls for village fixed effects and robust, clustered standard errors
Robust z stats in brackets - * significant at 10%; ** significant at 5%; *** significant at1%

Table 5 Robustness tests: Tobit model of mean absolute difference in links between household and partners, and Probit model on narrow symmetry definition for different definitions of hetereogeneity

	Mean absolute difference in links for different definition of heterogeneity (tobit model) (n=491)			Narrow definition of symmetry (29 symmetric groups=1) (probit model) (N=397)		
	Standard Deviation	Mean Absolute Deviation	Bound of 'poorest' partners	Standard Deviation	Mean Absolute Deviation	Bound of 'poorest' partners
male headed household?	1.391	1.296	0.896	0.53	0.55	0.348
	[0.33]	[0.31]	[0.21]	[0.40]	[0.40]	[0.25]
number of adults in household	0.443	0.412	0.307	-0.014	-0.014	-0.022
	[1.60]	[1.48]	[1.11]	[0.20]	[0.19]	[0.30]
years of education per adult in own household	-0.11	-0.111	-0.068	-0.101	-0.103	-0.117
	[0.43]	[0.43]	[0.26]	[1.47]	[1.48]	[1.71]*
functionings score per adult in household	0.804	0.904	0.846	0.353	0.327	0.134
	[0.49]	[0.56]	[0.54]	[0.81]	[0.75]	[0.30]
land per adult in household	1.412	1.264	0.475	0.013	0.06	0.27
	[0.61]	[0.55]	[0.21]	[0.02]	[0.08]	[0.30]
livestock value per adult in household	0.001	0.001	0.001	0	0	0
	[1.06]	[1.07]	[0.87]	[0.27]	[0.29]	[0.33]
Heterogeneity in male headship	10.834	14.172	13.479	-4.696	-6.655	-7.928
	[1.69]*	[1.62]	[1.58]	[2.09]**	[2.10]**	[2.58]***
Heterogeneity in number of adults	-0.148	-0.073	0.231	-0.108	-0.156	-0.088
	[0.33]	[0.12]	[0.41]	[1.04]	[1.09]	[0.66]
Heterogeneity in education	1.679	2.318	1.559	-0.194	-0.28	-0.187
	[3.75]***	[3.79]***	[2.63]***	[1.52]	[1.56]	[1.10]
Heterogeneity in functionings	3.562	5.352	3.618	-0.237	-0.445	-1.144
	[1.55]	[1.71]*	[2.00]**	[0.30]	[0.40]	[1.33]
Heterogeneity in livestock values	0.002	0.003	0.003	-0.001	-0.001	-0.001
	[1.18]	[1.23]	[1.28]	[1.00]	[0.99]	[0.84]
Heterogeneity in land areas	6.141	8.74	7.96	-2.328	-3.351	-3.562
	[1.71]*	[1.77]*	[1.90]*	[2.07]**	[2.22]**	[2.74]***
mean land among partners	-10.838	-10.948	-10.025	1.954	1.971	2.147
	[2.75]***	[2.77]***	[2.61]***	[1.70]*	[1.70]*	[1.66]*
Pseudo R-squared	0.07	0.07	0.07	0.22	0.23	0.27
Z statistics in brackets	* significant at 10%; ** significant at 5%; *** significant at 1%					

Table 6: Networks and performance – explaining value of agricultural output (in ln)

	1	2	3
	Links & Asymmetry	Links & Asymmetry	Links, Hubs & Spokes
	Not Instrumented OLS	Instrumented GMM	Instrumented GMM
share of land with lem quality	0.48	0.565	0.548
	[2.59]**	[3.27]***	[3.13]***
share of land with lem teuf quality	0.143	0.254	0.238
	[1.19]	[1.43]	[1.33]
share land on slope	0.083	-0.412	-0.415
	[0.73]	[1.68]*	[1.69]*
share of land flat	-0.261	-0.607	-0.598
	[1.29]	[2.36]**	[2.30]**
Ln of land (in ha)	0.431	0.41	0.425
	[3.57]***	[6.35]***	[6.28]***
Ln of number of adults	0.113	0.116	0.13
	[0.92]	[0.88]	[0.98]
Ln of funct score weighted adults	0.461	0.425	0.404
	[3.74]***	[2.76]***	[2.59]***
Male head?	1.388	1.203	1.198
	[1.97]*	[2.31]**	[2.31]**
Ln of age head	-0.126	-0.196	-0.177
	[0.92]	[1.41]	[1.24]
Ln of average years of education	0.069	0.004	0.006
	[1.74]	[0.13]	[0.19]
Ln of livestock value	0.041	0.04	0.042
	[2.24]**	[2.81]***	[2.83]***
Ln of number of links	0.162	0.403	0.425
	[2.79]**	[4.45]***	[3.05]***
Asymmetric? (yes=1)	0.14	0.428	
	[1.87]*	[2.88]***	
If asymmetric, whether hub?			0.428
			[1.98]**
If asymmetric, whether spoke?			0.509
			[2.56]**
Observations	472	404	404
R-squared	0.49		
Diagnostics related to instruments:			
Shea Partial R2 (relevance of instruments)		Links: 0.13	Links: 0.14
		Asymmetry: 0.19	Hub: 0.15
			Spoke: 0.15
Hansen J statistic (overidentification test): Chi-q(44)		53.23	51.17
		(p-value=0.17)	(p-value=0.18)

Robust t statistics in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

The regressions control for fixed effects at the village level (not reported). First stage regressions include whether born in the village, the number of blood relatives in the village, years of residence in the village as well as a full set of neighbourhood fixed effects.

Table 6A: First stage regressions for instrumentation of number of links and architecture

	1	2	3	4
	First stage regression (Links)	T-statistic	First stage regression (Asymmetry)	T-statistic
Exogenous variables				
share of land with lem quality	-0.249	-1.3	-0.080	-0.79
share of land with lem teuf quality	-0.076	-0.39	-0.072	-0.68
share land on slope	0.405	1.25	-0.029	-0.16
share of land flat	0.535	1.69	-0.020	-0.12
Ln of land (in ha)	0.130	1.62	-0.016	-0.33
Ln of number of adults	0.195	1.27	0.124	1.65
Ln of funct score weighted adults	-0.080	-0.47	-0.161	-1.84
Male head?	0.621	1.2	-0.365	-1.31
Ln of age head	0.155	0.83	0.005	0.05
Ln of average years of education	0.000	0	-0.003	-0.3
Ln of livestock value	0.040	2.63	-0.004	-0.47
Community fixed effects				
V1	1.332	4.25	-0.491	-1.56
V2	-1.156	-2.31	0.427	1.24
V3	-0.484	-0.84	0.514	1.73
V4	-0.081	-0.15	0.094	0.22
V5	0.521	0.99	-0.051	-0.1
V6	0.775	1.92	-0.438	-1.35
V7	-0.591	-1.79	0.266	0.87
V8	-0.824	-0.95	0.480	1.68
V9	0.564	1.88	0.532	1.81
V10	0.431	1.28	0.159	0.37
V11	0.155	0.41	0.438	1.35
Instruments				
No. of blood relatives in village	0.069**	1.98	-0.015***	-2.96
born in village?	-0.043	-0.34	-0.041*	-1.63
Years resident for head	-0.003*	-1.82	0.001	0.31
Neighbourhood fixed effects (65 fixed effects across villages, F-test of joint significance:)	p-value=0.00		p-value=0.02	
R-squared		0.13		0.19

Table 6B: First stage regressions for instrumentation of number of links and architecture (hubs & spokes)

	1	2	3	4		
	First stage regression (Links)	T-statistic	First stage regression (Hubs)	T-statistic	First stage regression (Spokes)	T-statistic
Exogenous variables						
share of land with lem quality	-0.273	-1.450	-0.077	-0.660	-0.163	1.570
share of land with lem teuf quality	-0.107	-0.560	-0.010	-0.090	-0.067	-0.630
share land on slope	0.420	1.280	-0.149	-0.740	0.130	0.710
share of land flat	0.551	1.700	-0.187	-0.940	0.160	0.880
Ln of land (in ha)	0.115	1.360	-0.029	-0.550	0.006	0.120
Ln of number of adults	0.163	1.050	0.024	0.250	0.117	1.330
Ln of funct score weighted adults	0.062	0.290	-0.191	-1.460	0.042	0.350
Male head?	0.906	1.200	-0.959	-2.070	0.435	1.030
Ln of age head	0.058	0.260	-0.139	-1.000	0.133	1.050
Ln of average years of education	-0.002	-0.090	0.005	0.320	-0.014	-1.060
Ln of livestock value	0.040	2.230	-0.015	-1.360	0.009	0.900
Community fixed effects						
V1	0.042	0.040	-0.081	-0.140	-0.144	-0.270
V2	0.406	0.620	-0.249	-0.620	0.353	0.960
V3	1.447	2.660	-0.778	-2.330	-0.083	-0.270
V4	0.990	1.210	-0.359	-0.710	0.613	1.330
V5	1.745	2.970	-0.676	-1.870	-0.148	-0.450
V6	0.839	1.560	-0.099	-0.300	0.214	0.710
V7	0.625	1.250	-0.203	-0.660	0.200	0.710
V8	-0.399	-0.390	0.915	1.470	-0.030	-0.050
V9	0.524	0.730	-0.178	-0.410	0.459	1.140
V10	0.900	1.490	0.473	1.280	0.020	0.060
V11	0.577	1.210	-0.053	-0.180	-0.144	-0.270
Instruments						
No. of blood relatives in village	0.065*	1.800	-0.039*	-1.770	0.017	0.860
born in village?	-0.071	-0.550	0.050	0.640	-0.077*	-1.760
Years resident for head	-0.003	-1.620	0.002	0.630	-0.001	-0.460
Neighbourhood fixed effects (65 fixed effects across villages, F-test of joint significance)		(p-value=0.00)		(p-value=0.01)		(p-value=0.01)
R-squared		0.15		0.16		0.14