Social Networks in Ghana[1]

Christopher R. Udry
Yale University
udry@yale.edu

Timothy G. Conley
University of Chicago
tim.conley@gsb.uchicago.edu

May, 2004

[1]This research has been funded by The Pew Charitable Trusts through its project on Moral and Social Dimensions of Microeconomic Behavior in Poor Communities. The underlying field research received financial support from the NSF (grants SBR-9617694 and SES-9905720), International Food Policy Research Institute, Institute for the Study of World Politics, World Bank Research Committee, Fulbright Program, ESRC and Social Science Reseach Council. We thank members of the Pew network for many helpful comments.

Abstract

In this chapter we examine social networks among farmers in a developing country. We use detailed data on economic activities and social interactions between people living in four study villages in Ghana. It is clear that economic development in this region is being shaped by the networks of information, capital and influence that permeate these communities. This chapter explores the determinants of these important economic networks. We first describe the patterns of information, capital, labor and land transaction connections that are apparent in these villages. We then discuss the interconnections between the various economic networks. We relate the functional economic networks to more fundamental social relationships between people in a reduced form analysis. Finally, we propose an equilibrium model of multi-dimensional network formation that can provide a foundation for further data collection and empirical research.

I Introduction

The design of economic policy in Africa has been, and continues to be, hampered by dependence upon misleading units of analysis inherited from venerable models that were developed in a radically different social context. We think about 'firms' that produce, interacting through markets with 'households' that consume, and about small enterprises in which the roles of the 'household' and the 'firm' are merged. It is increasingly clear, however, that the very definition of the core units of the 'household' and 'firm' are problematic in many African contexts and that the facile use of these concepts can lead to misleading judgments about the process of economic development.

In Africa, the boundaries of 'household,' 'firm' and 'market' are mobile and permeable. Things resembling market transactions occur within households and family-like interactions take place across households and firms. Rather than conceptualizing membership in a household or firm as a zero-one event, it is fruitful to think of individuals as participating in numerous social relationships of varying qualities and intensities with a variety of different people. Important insights about the structure of African economies and the design of appropriate economic policy are gained by the recognition that people are embedded within social networks.

Some of these implications are apparent in the Eastern Region of Ghana. The region is one of the key foci of Ghana's development strategy of expanding non-traditional exports. It is experiencing a transformation of its farming system from the production of local foodstuffs to intensive cultivation of pineapple for export as fresh fruit to Europe. The salient aspects of this transformation include learning about a set of new technologies, alternations in established patterns of land resource management, a dramatic shift to more capital-intensive production and important new risks. On each of these dimensions, patterns of change have been fundamentally conditioned by the composition of social networks.

For example, in Conley and Udry (2001, 2003), we find that information about the proper use of the new technology passes informally between farmers. Farmers experiment with varying levels of input intensity (particularly fertilizer) and they discuss the results of their experiments with a restricted set of peers. Farmers learn from these discussions. Our data show, for example, that a given farmer will begin to use more fertilizer after a neighbor with whom he is linked in an information network uses high amounts of fertilizer and achieves surprisingly high profits. As a consequence, people who are strongly connected with many other pineapple farmers (call them central) have learned the contours of this radically new technology more rapidly than others. We hypothesize that, in turn, other farmers value information connections with these central farmers because of the value of the knowledge that central farmers have obtained from others. The particular structure of the information network in these villages has had a dramatic effect on the diffusion of these new techniques.

Similarly, we find that access to land is secured via informal channels of political power. Plots in Ghana, as in much of Africa, are controlled by individuals, not households. Goldstein and Udry (2004) show that profits on women's farms are much lower than those of their husbands. In these villages, land fertility is maintained by periodic fallowing, and we show that the reason for the much lower profits on wives' plots is the fact that they are left fallow for shorter periods than are the plots of their husbands. This, in turn, is a consequence of the fact that the ties of mutual obligation and political influence that undergird land tenure security – and thus permit extended fallow periods for men – largely bypass their wives. In southern Ghana, access to land is based on negotiation, status and identity within a corporate group. Those with relatively weak ties to the group leadership are reluctant to fallow their land for fear of finding it reallocated to a rival when it is time to re-establish cultivation. Women are disproportionately likely to inhabit the margins of this political network and so fallow relatively little. What is particularly striking about our findings is the extent to which the "household" is inconsequential for cultivation decisions. The political influence of the husband seems to be entirely irrelevant for the security of tenure of the wife. It is in those households in which the husband holds a political office and the wife does not that the gap in fallowing (and thus fertility and profits) between the husband's plots and the wife's plots is largest.

Political influence and thus security of access to land flows through networks that cleave households. Similarly, husbands and wives tend to participate in separate insurance networks. Goldstein (2002) shows that risk-sharing between husbands and wives is highly imperfect. Women pool risk with other women in their villages more than with their husbands, and men share risk with male members of their extended families. Other financial flows show a similar pattern. Rates of return on capital invested in this export-led rural economy are astonishingly high (Goldstein and Udry 1999a). However, net financial inflows are small and largely follow well-established paths between individuals with deep social connections: sometimes between husband and wife, but more often between siblings, long-term friends and members of the extended family. It is clear that credit constraints apply at the level of the individual, not the "household": there were numerous instances during our fieldwork in which a wife was unable to make potentially profitable investments (say, in pineapple cultivation or a non-farm enterprise) as the husband was expanding his on-going enterprise. Once again, it is the individual's position in the relevant social network that is the crucial determinant of access to a valued resource.

Our work thus far has made it clear that economic development in this region is being shaped by the networks of information, capital and influence that permeate these communities. The first goal of this chapter is to describe the patterns of information, capital, labor and land transaction connections that are apparent in these villages. We discuss the interconnections between the various economic networks: how closely related are the functionally different connections between individuals? We then relate these functional connections between individuals to more fundamental spatial, kinship, migration, gender and religious relationships between people. Finally, we propose an equilibrium model of multi-dimensional network formation that can provide a foundation for further empirical research. Our hope is that this chapter can serve as a resource for future data collection on social networks guided by economic theory.

II Data and Setting

The research was conducted in four clusters of villages in the Eastern Region of Ghana. Commercial agriculture is no recent innovation in the study area. Historically, the farming systems of the study area have undergone a series of significant changes. In the 19th century, oil palm production sparked the first inflow of migration to the area. This district was at the heart of Ghana's cocoa revolution at the turn of the 20th century (Hill 1963). In 1930, swollen shoot disease devastated cocoa production and farmers adopted a system based on intercropped cassava and maize. Most recently (since the early 1990s), farmers in the area have moved towards intensive pineapple production for export.

The southern Ghanaian forest-savanna transition zone has seen a dramatic reduction in forest cover since the 1970s (as evidenced by aerial photography (Gyasi et al 1994) and satellite imagery (Hawthorne and Abu-Juam 1995), important reductions in fallow lengths over the same period (Gyasi et al 1994, Amanor 1994) and increased evidence of soil deterioration and infestation by pests, particularly the virulent weed known locally as akyeampong) (Gyasi 1990; Amanor 1994). Land rights, however, have not been transformed into anything close to individualized freehold tenure.

Berry (2001) notes that land in southern Ghana "is subject to multiple, overlapping claims and ongoing debate over these claims' legitimacy and their implications for land use and the distribution of revenue" (2001, p. xxi). Goldstein and Udry (2004) provide evidence that fallowing decisions are strongly influenced by the cultivator's political position in the village and lineage.

Within each village cluster we selected 60 married couples (or triples - a bit more than five percent of husbands have multiple wives) for our sample. Each member of the pair or triple was interviewed 15 times during the course of the two years. Every interview was carried out in private, usually by an enumerator of the same gender.

The survey was centered around a core group of agricultural activity questionnaires (plot activities, harvests, sales, credit) that were administered during each visit. In addition, about 35 other modules were administered on a rotating basis. We also administered (once per field) an in-depth plot rights and history questionnaire and mapped each plot using a geographical information system. We supplemented this with data on soil fertility: the organic matter and pH of each plot was tested each year.

Much of the analysis of this paper is based on data we collected on the social and economic interactions of our sample members with other people, and in particular with other people in our sample. We recorded data on the contacts that each of our respondents had with other individuals (in the data archive, this is the "individual roster"). 'Contacts' include learning interactions, credit and gift transactions, sales/purchases of farm inputs and outputs and labor and land market interactions. In the second year of the survey, for each such transaction we asked the respondent to provide certain information about the other party to the transaction. Data was recorded on the relationship and frequency of contact between the respondent and the contact, the residence and occupation of the contact and the identification number of the contact if he/she is in the sample. Every contact has a unique identification number so we can trace them throughout the different questionnaires.

These contacts are reasonably well-defined for credit, gift, purchase/sale, land and labor transactions. However, the nature of learning interactions is more subtle and subjective. We asked a long series of questions about specific tasks, ideas and decisions. For example, "Did anyone advise you which particular crops to plant on your farm?", or "From whom did you first learn about the idea of using fertilizer?" Our list of information links for any individual includes anyone named in response to these questions. In principle, all pairs of respondents who learned about farming from each other should be listed. In fact, it is impossible to record the names of all individuals with whom respondents discuss farming, so enumerators were instructed to request the names of individuals with whom more than casual conversations took place. Two important issues arise as a consequence. First, different respondents interpret ‘casual’ differently, and second, important information might be transmitted during quite casual conversation.

Therefore, we also collected complementary data on information interactions from questions about randomly selected *pairs* of respondents. We collected two such datasets. One was focused particularly on learning about pineapple. The other was more general. Consider the latter. Each sample respondent was questioned about a random sample (without replacement) of seven other individuals in the same village, and with three other predetermined individuals who appear to be focal in the village. Links for each respondent are defined according to his responses to specific questions about the relationship with the selected persons. For example, “Have you ever gone to _ _ _ for advice about your farm?” For each pair of individuals we also record information on the frequency of contact, the period over which they have interacted and information on familial relations. This type of data has proven particularly useful for understanding the process of social learning through network connections. It is less useful for understanding the overall shape of social networks in the villages because it provides no information on interactions between individuals outside the random sample of seven for each respondent. Therefore, for the remainder of the chapter we utilize data from the ‘individual rosters’ of each of the respondents.

In addition to the data on social and economic interactions, we will make use of data on each respondent’s family background, wealth and education. Finally, we rely on our data on the spatial relationships between the plots of our respondents derived from the village and plot maps.

III Networks of Information, Finance, Land and Labor in Southern Ghana

The data on network links provides us with a random sample from which we can infer some characteristics of the social and economic networks in our research villages. The data are ‘egocentric’ in the parlance of network theory: we know (in principle) about the links from our sample individuals to other people. If one of these links happens to be to another individual in our sample then we know a great deal about both nodes of that link, and about further connections along the network. However, if the individual at the other end of the link is not a member of our sample, then our information is quite limited. In particular, we know nothing about further links along the network.

A simple consequence of our fixed sample size across villages is that the probability that both nodes of a given link are included in our sample is lower in a large village than in a small village. This fact does not hinder inference about the relative densities of networks across villages. Let L_v^t be the number of links of type t in village v. The density of the t network in v is defined as

$$\Delta_v^t = \frac{2L_v^t}{n_v(n_v - 1)} \tag{1}$$

(Wasserman and Faust 1994). Clearly, Δ_v^t ranges from 0 to 1 (for a complete network in which each individual is connected to each other individual). Consider two villages v and u with $\Delta_v^t = \Delta_u^t$. However, the sizes of the villages could differ ($n_v \neq n_u$). Since the densities are equal,

$$L_v^t = \frac{n_v(n_v - 1)}{n_u(n_u - 1)} L_u^t. \tag{2}$$

A link is observed in our data if *both* involved individuals are in our sample. Therefore, if S is the fixed sample size across villages, the probability of observing any given link in village v is $\left(\frac{S}{n_v}\right)^2$. The expected number of links observed in village v is

$$\left(\frac{S}{n_v}\right)^2 L_v^t = \left(\frac{S}{n_v}\right)^2 \frac{n_v(n_v - 1)}{n_u(n_u - 1)} L_u^t \approx \left(\frac{S}{n_u}\right)^2 L_u^t \tag{3}$$

(for large n_u, n_v), which is the expected number of links observed in village u.

Our sampling was based on villages, and hence captures links only if both individuals reside in the village. It is commonly assumed that the village is the natural domain for the analysis of social and economic networks in rural West Africa, but that need not be true in all places and for all types of social network. These villages, in particular, have long been tightly integrated into the regional, national and global economy. An analysis of connections entirely within the village may be profoundly misleading if the most important interactions generally cross village boundaries.

The data permit us to identify the relative extent of interactions within and across village boundaries by network type. For most of the interactions we consider the majority of activity occurs within the village. Most of the individuals named by our respondents as sources of information about farming reside within the village. 66% of the recorded cases of learning activity occurred with other people in the village. This proportion varies across villages, from a low of 53% in the smallest village to between 70% and 80% in the remaining three villages.

Financial interactions are similarly concentrated within villages. 59% of credit transactions and gift exchanges occur within village boundaries; virtually all the rest with relatives and friends outside the village. Only 13 of 719 loans were taken from formal or semi-formal sources. Again, the proportion rises with village size, from 54% in the smallest village to 65% in the largest. 62% of the individuals from whom our respondents obtained land reside within the same village. This proportion ranges from 47% in the smallest village to 74%.

Labor transactions appear to be even more concentrated within villages. 76% of identified non-household labor used on sample plots comes from within the village. However, this is likely to be an overestimate of the degree of concentration, because 40% of non-family labor is identified in our data simply as anonymous "no relation." For these transactions, we do not record the residence of the laborers, so we do not know if they originate from the village. Informal discussions and observation during the field research make it apparent that a substantial proportion of this kind of labor is brought in as daily wage labor from nearby towns. Moreover, most within-village labor transactions occur when a household head or spouse hires dependents in other households. When this is the case, we do not have information on the household identifier of the hired worker, so the link can not be included in much of the analysis that follows.

In dramatic contrast, only 7.6% of sales of farm output occur between residents of the village. Virtually all output is sold directly in regional markets, to itinerant traders or, in the case of pineapple, to specialized commercial exporters.

The predominance of intravillage interactions for information transfer, financial flows and land exchange makes salient the analysis of their village-level networks. We are less confident of the importance of the within-village labor network and place less emphasis on this network. Within-village sales of output/purchases of farm inputs are unimportant. Therefore, we do not examine further the shape or determinants of this network.

All of the networks that we consider are sparse in terms of direct connections. Less than 2% of pairs of respondents are connected by any sort of direct link (this corresponds to a network density of approximately .04). Table 1 presents summary statistics on the frequency of the different kinds of links by village. In each case, the denominator of the proportion is the number of pairs of individuals in our sample in each village and the numerator is the number of observed links of a given type between sample individuals. There are few strong differences across villages in the density of the individual functional networks. Village 2 has a noticeably sparse network of labor connections, while the network of land transactions in village 3 is particularly dense.

There is a strong spatial element to these networks, and it is useful to be able to visualize their shape in real space (Faust et al 1999). Figures 1 through 4 plot the geographic location and the various network links between our sample individuals. The geographic location of each individual is defined by the average position of the plots (s)he cultivates and is indicated by a node of the graph. Links between sample individuals in the various networks are indicated by the edges of the graphs.

There are individuals in each village for each network who appear isolated in these graphs. That appearance is a misleading consequence of the strategy of constructing these graphs based on 'ego-centric' data from a random sample of the population. In fact, for each of these functional networks there is virtually no one in any of these villages who has no interactions with anyone. Virtually everyone in our sample has learning contacts, exchanges credit and/or gifts, hires labor and has obtained land from someone. If none of those other parties happens to be in our sample, the individual appears isolated in the graphs.

The graphs do point in certain interesting directions. First, there is a substantial, but by no means complete, overlap between the different networks in a given village. Individuals connected along one dimension appear more likely to be connected along the others. This impression is strongly supported by the statistical evidence that follows. Second, information networks appear to be more characterized than the other networks by the presence of focal individuals, who have multiple links in our sample. Third, it is apparent that connections are more likely between nearby farmers, but that links between more distant individuals are common, particularly for credit networks.

Links between individuals in the different functional networks are by no means independent. Tables 2 $A-F$ show that the presence of any kind of link between two individuals is much more likely conditional on the existence of any other kind of link. Credit and information links are the most closely associated: nearly half of the pairs connected by an information link are also connected by a flow of credit or gifts, compared to an unconditional probability of 1.25% for the existence of a credit link. The same general pattern, however, holds for all pairs of networks.

Careful analysis is required, therefore, to distinguish the consequences for behavior or outcomes of different kinds of connections between individuals. A finding, for example, that adoption of a new technology is particularly likely for an individual who shares an information connection with someone who has recently had success with that technology might be attributed to social learning effects. However, because information links are correlated with financial links, this same observation might arise if credit markets are imperfect and the success of the first farmer lowered the cost of capital to the second. Appropriate attention to the entire set of relevant networks may make it possible to distinguish between the different kinds of social interaction effects that influence behavior.

Alternatively, it might be appropriate to think of a single network of basic connections between individuals that can be mobilized for multiple purposes depending upon circumstances. Over some period, a connection might yield flows of finance; at others it might involve a land transaction. Recall how our data on links are constructed: the credit, land and labor links are based on the existence of an exchange over the relevant period. Information links are defined somewhat differently because they look back at the whole history of learning interactions over a respondent's farming career. We believe that these are better-defined measures than some loose notion of "who could you approach for credit," land or labor. However, it obviously has its costs because it provides no direct information about the potential (rather than realized) uses of a given tie between individuals.

Finally, close correlation between links of various types raises the possibility that the underlying determinants of network connections are similar for different types of links. If the different functional links really do have different purposes, this pattern tends to focus our attention on the costs of constructing and maintaining connections, which could well be similar for different types of links. We turn now, therefore, to an examination of the correlations between current ties in the various networks and a set of arguably deeper social connections between farmers.

IV Predicting Networks

Table 3 provides summary statistics about the social connections between (non-married) pairs of individuals in the villages. The sample is the set of all within-village pairs of sample individuals, except for pairs linked by marriage.[1] Approximately one-fifth of pairs are members of the same religion. Again, one-fifth of pairs are members of the same extended matrilineage (each village has two to three matrilineages that account for a collective $70 - 80\%$ of the population, and five or six other, smaller matrilineages). For a further fifth of the sample pairings, at least one individual in the pair holds a traditional village or lineage office (for example, is village elder). The two individuals have the same soil type (described as one of three major groups of soil types) in a third of the pairings. The average absolute value of the age difference between the paired individuals is a decade, while the average absolute nonland wealth difference is $750,000$ cedis[2] (compared to average wealth of $670,000$ cedis). The average distance between the plots of pairs of individuals is about $1\frac{1}{4}$ kilometers.

Logit estimates of the probabilities that a link exists in a pair are presented for each of the four networks (information, credit, land and labor) in Table 4.[3] Let $l_{ij}^t = 1$ be an indicator variable that there is a link of type t between individual i and individual j. Let the notation $\Pr\{l_{ij}^t = 1\}$ refer to the probability that $l_{ij}^t = 1$ conditional on observable (to the econometrician) information. We estimate

$$\Pr\{l_{ij}^t = 1\} = \Lambda(X_{ij}\beta), \tag{4}$$

where Λ is the logistic CDF. Coefficients are presented as odds ratios (e^{β_k}), along with z-statistics for the test against the null hypothesis that the odds ratio is unity. The standard errors are heteroskedasticity-consistent.

The covariates and samples are identical across the different regressions, and the predicted probabilities of a link in the middle of the distribution of the covariates is very similar (always near $1 - 2\%$), so direct comparisons of the odds ratios are sensible.

In each case, gender is a crucial determinant of the likelihood of a link between individuals. It is very rare for non-married individuals of opposite genders to interact directly in any of these four dimensions. It is also the case that membership in the same matrilineage is strongly associated with the existence of any of these network links. This is particularly so in the case of credit links, and relatively less so in the case of information and land links. Given the informal enforcement mechanisms that can be brought to bear within the matrilineage, it is not surprising to find that credit and mutual

[1] Married pairs are excluded from the analysis samples in Tables 4 and 5 as well.

[2] The exchange rate during the survey period ranged from 1700 to 2300 cedis per US dollar.

[3] See Anderson et al (1999) for a discussion of logit models in network analysis.

gift exchanges are more common within lineages (Klingelhofer 1972). More surprising is the finding that among the four networks, land transactions are least positively affected by membership in the same lineage. Land rights in southern Ghana are very strongly associated with the matrilineage (Goldstein and Udry 1999b, Berry 2001). It is possible that the relatively small increment to the likelihood of a land link associated with membership in the same matrilineage reflects the potential ambiguities of plot rights when land is transferred within the matrilineage. "The process of acquiring and defending rights in land is inherently a political process based on power relations among members of the social group" (Bassett and Crummey 1993, p. 20); the transaction may be less problematic when made on a more commercial basis across lineage boundaries.

The finding that membership in the same church is strongly associated with the presence of a credit link again may reflect the informal enforcement mechanisms that are available within religious groups. Religion is not strongly associated with the likelihood of any other type of network connection. Similarly, we find that credit links are much more likely when both individuals share a family origin in the same region of Ghana. Region of origin can cross-cut matrilineages in these villages, so this may reflect the social enforcement mechanisms available to communities formed of migrants.

Farmers who share similar soil characteristics are much more likely to share an information link than otherwise. This is naturally related to the potential gains from the information link, if optimal farming practices are variable across soil types in a non-systematic (or at least non-obvious) manner. We also find that land links are much more likely across farmers who share soil types; the simultaneity of this relationship makes even tentative interpretations problematic. In column 5 of Table 4 we remove this covariate and distance, which is subject to the same worry, and show that the remaining odds ratios are generally stable.

Differences in wealth are very strongly associated with the presence of information connections. Looking more closely at the individuals in the network maps of Figure 1 we find that several of the apparently nodal individuals are much wealthier than average, and they tend to be connected to individuals who are relatively less wealthy. There appears to be a surprising element of hierarchy in the information network. This is otherwise apparent only in the land network, where it is to be expected. Credit and gift exchange, on the other hand, appears to be more horizontally oriented. These exchanges do not often occur between rich and poor. The hypothesis that the provision of insurance is an important element of credit and gift exchange is consistent with this pattern.

Information and land links are much less likely between individuals who are members of the first generations of their families to reside in the village. For land links this effect is sufficiently strong that there are no instances of such links in our data.

All of these network interactions are more likely between farmers who are located near each other. One would expect the cost of social interactions to rise with distance, hence this is a pattern to be expected. In fact, it is this general pattern that underlies the decision by many researchers to use geographic location as a proxy for social connections. The relationship between proximity and the likelihood of a network linkage is sufficiently strong that location may indeed be a reasonable proxy for network membership for certain applications. However, it is clearly not an exact relationship, and the impact of distance on the likelihood of interaction varies in an interesting way across the networks. In particular, distance has a much less strong negative impact on the likelihood of a credit tie than on any of the other network connections. We have shown in other work that there is a strong positive spatial correlation in agricultural shocks (Conley and Udry 2003, Goldstein and Udry 2004). This finding, therefore, points again to an important insurance motivation for the credit and gift exchanges.

Deeper social connections between individuals are strongly associated the presence of connections in these four social networks. It is possible that information about these deeper relationships between people for predicting network connections is less useful for predicting relatively transitory connections. Unfortunately, our data is constructed in such a way that we can examine only one side of this question. We have no information regarding lapsed ties in these social networks. If two individuals once shared a connection and no longer do, this connection will not appear in our data. However, we do know for how long two individuals who do share a connection have interacted.[4] The final two rows of Table 4 show that the probability of correctly predicting a link to exist when the two individuals have known each other for less than 5 years is lower than when the two have been acquainted for longer periods. This pattern is less evident for credit links, but is quite strong for each of the other types of connections. While the more fundamental social relations between people are important predictors of the likelihood of functional network connections, they are less useful in distinguishing more recent connections and will also be less useful

[4] The precise question is "How long have you known ____?" This is not necessarily the same as knowing the length of time the two individuals have shared this particular interaction.

to distinguish transitory links more generally.

IV.1 Robustness

The simple logit estimates presented in Table 4 leave open the possibility that there are unobserved characteristics of individuals that influence the probability that they are linked with others. Put simply, it is plausible that certain individuals are more likely to build connections in a network than others. Consequently, we explore the possibility that the appropriate probability model takes the form

$$\Pr\{l_{ij}^t = 1\} = \Lambda(X_{ij}\beta + \lambda_i + \lambda_j), \tag{5}$$

where λ_i and λ_j are unobserved individual fixed effects. This model is obviously strongly related to the conditional logit model (Chamberlain 1980) but the standard conditioning approach is not available when there are two dimensions of fixed effects. Instead, we adopt a maximum likelihood approach to estimate β and λ_i. When there is a fixed number of observations per individual the conventional ML estimator of β is inconsistent. However, the median number of pairs including any individual in our data is 114, so we rely on the asymptotic consistency of the ML estimator. Estimates of equation 5 are presented in Table 5.

This estimation strategy implies that any individual who has no link of a particular type (that is, any i such that $l_{ij}^t = 0 \ \forall j$) contributes no information to the estimate of $\hat{\beta}$. Such individuals and all of the pairs of which (s)he is a part are dropped from the estimation sample. The precise numbers are reported in Table 5. In addition, the variable "at least one of pair holds office" is colinear with the individual effects and is dropped.

The remarkable feature of Table 5 is the overall similarity of the results to those reported in Table 4. The strong effect of gender is equally apparent in the Table. Also, distance between plots is once again strongly negatively associated with the likelihood of a linkage in any of these dimensions. Once again, however, this effect is much less strong for credit links than for any of the other types of connections between individuals. As in Table 4, there is a very strong positive association between two individuals sharing the same soil type and the existence of an information link between them.

The strong positive relationship observed in Table 4 between the absolute value of the wealth difference between individuals and the likelihood that they share an information link is not replicated in Table 5. The result in Table 4 appears to reflect the fact that certain quite wealthy individuals are focal in the information network. Once this fixed effect is taken out, the wealth difference is negatively related to the likelihood of a linkage.

The very strong positive effect of membership in the same matrilineage for the likelihood of a credit link that we observed in Table 4 is not as apparent in Table 5. Instead, there is a much stronger effect of being a member of the same religion.

V Modeling Social Networks

The long term goal of this research program is an enriched understanding of individuals' behavior in the context of multiple and overlapping social networks and, in particular, to acknowledge the agency of the individuals embedded in social networks. The constituent links of the networks are chosen. They confer benefits, imply responsibility and require effort to create and maintain. The choice to invest in social connections within and across conventional household boundaries is not so much a simple process of accumulation as of developing an appropriately complementary set of ties – 'composition' is Belinga and Guyer's felicitous term (Belinga and Guyer 1995). While the shape of the networks moulds the pattern of economic development, economic change alters the value of particular network links and thus the shape of the networks themselves. Our goal is to understand this dynamic of network formation.

The reduced form analyses of the previous section provide clues as to the motivations that might be driving individuals' decisions as they compose a set of network connections. The results are consistent with the notion, for example, that people are willing to pay a higher price for longer-distance credit connections than they are for information connections. However, in the absence of a model that incorporates both individuals' motivations for shaping the eventual structure of a particular social network and an appropriate equilibrium concept, these hints must remain just hints.

There is a large and useful literature that examines network efficiency, with the goal of characterizing the network configuration that maximizes a value function (Bolton and Dewatripont 1994; Hendricks et al 1995; Economides 1996). This is appropriate for a planner (such as a telecommunications monopoly), but not for the decentralized process that governs the formation of social networks like those in the sample villages. Instead, we focus on the incentives of the individuals who build the links of the network (Granovetter 1973, 1992; Coleman 1966; Fafchamps and Minten 1999; Fafchamps 1999).

Consider for now just two networks: the network for information and that for credit/gift exchange. The links in these networks are bidirectional: if i and j converse about farming, then information flows in both directions. This is less obviously true of credit links, but to the extent that insurance motivations are central to these transactions, a bidirectional characterization is appropriate. Of course, this does not imply that the value of the link is identical to i and j. Bala and Goyal (1999) model the situation in which l_{ij}^t is costless to j, while in Jackson and Wolinsky (1996) both parties bear the cost and either can sever the link. It is not entirely clear what approximation is appropriate in these villages. In fact, costs are borne by both parities. But they are also asymmetric; the individual who approaches the other is likely to bear a larger proportion of the cost.

Let $N = \{1, 2, ...n\}$ be the set of agents in the village. Agent $i's$ strategy is a vector $l_i = \{l_{i,1}^I, l_{i,2}^I, ...l_{i,n}^I, l_{i,1}^C, l_{i,2}^C, ...l_{i,n}^C\}$ where $l_{i,j}^t = 1$ if i forms a type t link with j. Let the payoff to i for a direct connection of type t to j be V_{ij}^t. This represents the value to i of the information exchange or insurance contract between i and j.[5] V_{ij}^t varies across ij pairs.

The literature assumes exponential network decay (parameterized by δ), so that an indirect connection between i and j provides less value than a direct connection. The appropriate model for network decay would likely be very different depending upon the specific context and the type of network. Information, for example, might be garbled when transferred between people, in which case exponential decay could be a natural assumption. The value of an indirect link in a financial network might be very different, however, depending upon the nature of the bilateral exchanges. At one extreme, there might be no decay so that everyone in a connected network shares complete insurance. One intermediate case would involve a limited commitment model, which might involve complete insurance for small shocks and incomplete insurance for larger shocks (Ligon 1998). It is clear that even if the decay in the value of indirect connections in a financial network can be approximated by an exponential process, δ would vary according to the type of network. We omit that notation for simplicity.

In a given network g, if the shortest path of type t between i and j is d, the value to i is $V_{ij}^t \delta^d$. The cost of forming links (l_{ij}^I, l_{ij}^C) is $c_{ij}(l_{ij}^I, l_{ij}^C)$,

[5] Obviously, in general, V_{ij}^t depends upon the entire network configuration. In virtually any learning model, the value of an additional bit of information declines as the volume of information increases. The situation is even more complex with respect to credit, in which V_{ij}^C will depend upon the covariance of $j's$ income with everyone else connected directly or indirectly with i. We assume that these effects are small so that we can approximate the value of the connection in this linear fashion.

where $0 < c_{ij}(0,1), c_{ij}(1,0) \leq c_{ij}(1,1) < c_{ij}(0,1) + c_{ij}(1,0)$. The value of a network g in which $N(i;g)$ is the set of individuals with direct or indirect connections of any type to i in g is

$$\Pi_i(g) = \sum_{j \in N(i;g)} \left\{ V_{ij}^I \delta^{d(i,j;I,g)} + V_{ij}^C \delta^{d(i,j;C,g)} - c_{ij}(l_{ij}^I, l_{ij}^C) \right\} \tag{6}$$

The obvious Nash equilibrium of this network formation game implies a series of inequality constraints that must hold. Let g_{-ijt} denote network g without a type t link between i and j, and g_{+ijt} denote the same network with the addition of this link. If $l_{ij}^t = 1$, then

$$\begin{aligned} \Pi_i(g) &> \Pi_i(g_{-ijt}) \\ \Pi_j(g) &> \Pi_j(g_{-ijt}), \end{aligned}$$

while if $l_{ij}^t = 0$ then obviously enough

$$\begin{aligned} \Pi_i(g) &< \Pi_i(g_{-ijt}) \\ \Pi_j(g) &< \Pi_j(g_{-ijt}). \end{aligned}$$

Heterogeneity in the characteristics of individuals (and their deeper social relationships to one another) can be introduced. Let $V_{ij}^t = V^t(x_i, x_j, \varepsilon_{ij})$ and $c_{ij} = c(l_{ij}^I, l_{ij}^C; z_i, z_j, \upsilon_{ij})$

Bala and Goyal (1999) provide a characterization of the equilibria of a related model. The network in a village will be connected, and in the symmetric case ($V_{ij}^t = V^t$ and $c_{ij} = c(l_{ij}^I, l_{ij}^C)$) there are many Nash equilibria. This immediately raises very important questions about the feasibility of estimation (Jovanovic 1989). The problem is less severe in our application: since V and c vary across pairs of individuals in our model, the set of equilibria is made smaller. Nevertheless, it should not be expected that there would be a unique Nash equilibrium of this game, which complicates estimation of the parameters of V_{ij}^t and c_{ij} (Tamer 1999).

The most important intuition of this model for empirical work on network formation is that the expected benefit to i of a link with j depends upon i's opinion about the links that j has. Hence, i is more likely to pay the cost of linking to j if he thinks that j is linked to many people relevant to i. If j has characteristics (such as a common extended family or a common church) which make it likely that he will have links with individuals whose experience is valuable to i (with people who have plots similar to i for an information link, for example), then i is more likely to pay for a link with j. The costs of a link between j and, say, k depend on the interaction between aspects of their social backgrounds. Hence, i's prediction of the network resources available to j (and thus V_{ij}^t) need not be collinear with the costs of forming the link with j.

This model can in principle be estimated. However, the obstacles to be overcome are daunting. The amount of independent information available across individuals within a given network is limited because of the interdependence of each individual's actions. Data may be required on a large number of networks in order to distinguish the different equilibria that exist in the different villages (Conley and Topa 2003). The need for a broad cross-section of data is in tension with the simultaneous need for rather intensive collection of data from each individual, as described below. We believe that the more promising direction is to study the dynamics of the networks directly, by observing changes in the composition of individuals' network ties. The problem then becomes one of understanding the *change* in l_i conditional on the current existence of a particular equilibrium network g. It remains the case that plausible sources of exogenous variation in c_{ij} or V_{ij}^t are required in order to interpret any observed changes in the set of network links in the village. However, this appears to be the most promising strategy and it has important implications for the design of data collection.

VI Lessons for future research

We have argued that patterns of economic change in rural Africa are being shaped by the configurations of social networks in these communities. In turn, individuals' incentives to make and break network connections are influenced by economic transformation. The process through which individuals compose their various network connections, and the implications of these connections for economic and social activities is an important frontier for empirical research. The data that are generated by conventional surveys, however, are not well-suited for this task.

Equation 6 makes explicit the importance of a complete enumeration of any given social network within a community. It is not possible to calculate the value of the network – or of any of its constituent links – in its absence. Therefore, if an analysis of the dynamics of social networks is a goal of the research the basic strategy of ego-centric analysis based on a random sample is not appropriate. If the village is sufficiently closed to make it the relevant domain of the network, then a complete enumeration of the network connections will be required to make progress on understanding the determinants of network formation. This is a costly decision to make in planning data collection, because for most other purposes a complete enumeration would not be the optimal sampling strategy.

Data on the creation and dissolution of network connections is particularly useful to overcome some of the central technical challenges of analyzing the determinants of the shape of social networks. Therefore, collection of panel data might be an appropriate strategy. Alternatively, it might be possible to collect retrospective data on the history of each individual's links to other individuals in the community.

The potential usefulness of retrospective data on network links is limited by the need for variation in the returns to or costs of links between individuals that are driving the changes in the composition of people's networks. Data on network connections has to be integrated with more conventional socioeconomic data to begin to understand the implications of the network for behavior and outcomes, and thus to model the incentives individuals face for link formation and dissolution. It is not generally possible to collect retrospectively such data.

It is clear that functional network connections, like the information, credit, labor and land interactions described in this chapter, are strongly influenced by background variables like geography and family history. In addition to these underlying social variables, any analysis of transitions requires data on exogenously changing factors that influence the costs or benefits of connections. For example, new market opportunities, new technologies, or changes in the returns to different kinds of assets could provide sources of variation to provide insight into the creation or collapse of network links.

A key advantage of the Ghana dataset discussed in this chapter is the relative precision with which it defines network connections. It is possible, for example, to distinguish between social interactions that are focused on the exchange of farming information and those that are associated with credit or gift exchange. It is plausible that the incentives people face to develop and maintain these different types of connection are different, and in fact we have provided reduced form evidence that they shaped by different underlying social connections between people. The weakness of data on realized interactions is that they do not necessarily reveal the potential utility of a given network link.

References

Amanor, K.S. *The New Frontier: Farmers' Response to Land Degradation, A West African Study*. London: Zed Books, 1994.

Anderson, C.; Wasserman, S.; and Couch, B. "A p* Primer: Logit Models for Social Networks." *Social Networks* 21 (1999): 37-66.

Bala, V., and Goyal, S. "A Non-Cooperative Theory of Network Formation." *Econometrica* 68, 5 (September 2000): 1181-1229.

Bassett, T., and Crummey, D. *Land in African Agrarian Systems*. Madison: University of Wisconsin Press, 1993.

Berry, S.S. *Chiefs Know their Boundaries: Essays on Property, Power and the Past in Asante, 1896-1996*. Portsmouth: Heinemann, 2001.

Bolton, P., and Dewatripont, M. "The Firm as a Communication Network." *Quarterly Journal of Economics* 109, 8 (1994): 809-840.

Chamberlain, G. "An Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47 (1980): 225-238.

Coleman, J. *Medical Innovation: A Diffusion Study*. New York: Bobbs-Merrill, 1966.

Conley, T., and Topa, G. "Estimating Dynamic Local Interactions Models." Manuscript. Chicago, IL: University of Chicago; New York, N.Y.: New York University, 2003.

Conley, T., and Udry, C. "Social Learning Through Networks: The Adoption of New Agricultural Technologies in Ghana." *American Journal of Agricultural Economics* 83, 3 (2001): 668-73.

Conley, T., and Udry, C. "Learning About a New Technology: Pineapple in Ghana." Manuscript. Chicago, IL: University of Chicago; New Haven, CT: Yale University, 2003.

Economides, N. "The Economics of Networks." *International Journal of Industrial Organization* 14, 6 (1996): 673-99.

Fafchamps, M. "Ethnicity and Credit in African Manufacturing." Manuscript. Oxford, UK: Oxford University, 1999.

Fafchamps, M., and Minten, B. "Relationships and Traders in Madagascar." *Journal of Development Studies* 35, 6 (1999): 1-35.

Faust, K.; Entwisle, B.; Rindfuss, R.; Walsh, S.; Sawangdee, Y. "Spatial Arrangement of Social and Economic Networks in Nang Rong District, Thailand." *Social Networks* 21, 2

(1999): 311-337.

Goldstein, M. "Chop Time, No Friends: Intrahousehold and Individual Insurance Mechanisms in Southern Ghana." Manuscript. London, UK: London School of Economics, 2002.

Goldstein, M., and Udry, C. "Agricultural Innovation and Resource Management in Ghana." Manuscript. London, UK: London School of Economics; New Haven, CT: Yale University, 1999a.

Goldstein, M., and Udry, C. "Gender and Land Resource Management in Southern Ghana." Manuscript. London, UK: London School of Economics; New Haven, CT: Yale University, 1999b.

Goldstein, M., and Udry, C. "Gender, Power and Agricultural Investment in Ghana." Manuscript. London, UK: London School of Economics; New Haven, CT: Yale University, 2004.

Granovetter, M. "Strength of Weak Ties." *American Journal of Sociology* 78, 6 (1973): 1360-1380.

_____. "Economic Institutions as Social Constructions–A Framework for Analysis." *Acta Sociologica* 35, 1 (1992): 3-11.

Guyer, J., and Belinga, S.L.E. "Wealth in People as Wealth in Knowledge: Accumulation and Composition in Equatorial Africa." *Journal of African History* 36 (1995): 91-120.

Gyasi, E.A.; Agyepong, G.T.; Ardayfio-Schandorf, E.; Enu-Kwesi, L.; Nabila, J.S.; and Owusu-Bennoah, E. "Environmental Endangerment in the Forest-Savanna Zone of Southern Ghana." Technical report. Tokyo, Japan: United Nations University, 1994.

Hawthorne, W.D., and Abu-Juam, M. *Forest Protection in Ghana With Particular Reference to Vegetation and Plant Species*. Cambridge, UK: IUCN/ODA/Forest Department Republic of Ghana, 1995.

Hendricks, K.; Piccione, M..; and Tan, G. "The Economics of Hubs: The Monopoly Case." *Review of Economic Studies* 62, 1 (1995): 83-101.

Hill, P. *Migrant Cocoa Farmers of Southern Ghana*. Cambridge, UK: Cambridge University Press, 1963.

Jackson, M.O., and Wolinsky, A. "A Strategic Model of Social and Economic Networks." *Journal of Economic Theory* 71, 1 (1996): 44-74.

Jovanovic, B. "Observable Implications of Models with Multiple Equilibria." *Econometrica* 57, 6 (1989): 1431-37.

Klingelhofer, A. "Agriculture." In Brokensha, D., editor, *Akwapim Handbook*. Accra, Ghana: Ghana Publishing Corporation, 1972.

Ligon, E. "Risk Sharing and Information in Village Economics." *Review of Economic Studies* 65, 4 (1998): 847-64.

Tamer, E. "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria." *Review of Economic Studies* 70, 1 (1999): 147-167.

Wasserman, S.; and Faust, K. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press, 1994.

Table 1: Incidence of Network Connections

Fraction of Pairs of Respondents within Villages with Network Connections

	1 Information Links	2 Credit/Gift Links	3 Land Links	4 Labor Links	5 Village Population
Overall	0.015	0.013	0.005	0.006	
By Village:					
1	0.013	0.011	0.002	0.008	1250
2	0.016	0.015	0.004	0.001	2000
3	0.018	0.012	0.010	0.009	450
4	0.013	0.013	0.004	0.003	1000

Table 2: Incidence of Network Connections

Fraction of Pairs of Respondents within Villages with Network Connections
Conditional on the Existence of Alternative Links

A

Information Link		Credit Link: no	Credit Link: yes
no	n	47860	274
	(row pct)	99.43	0.57
yes	n	406	338
	(row pct)	54.57	45.43

Pearson chi2(1) = 1.2e+04 Pr = 0.000

B

Information Link		Land Link: no	Land Link: yes
no	n	48034	100
	(row pct)	99.79	0.21
yes	n	602	142
	(row pct)	80.91	19.09

Pearson chi2(1) = 5.3e+03 Pr = 0.000

C

Information Link		Labor Link: no	Labor Link: yes
no	n	48042	92
	(row pct)	99.81	0.19
yes	n	556	188
	(row pct)	74.73	25.27

Pearson chi2(1) = 8.1e+03 Pr = 0.000

D

Credit Link		Land Link: no	Land Link: yes
no	n	48134	132
	(row pct)	99.73	0.27
yes	n	502	110
	(row pct)	82.03	17.97

Pearson chi2(1) = 3.8e+03 Pr = 0.000

E

Credit Link		Labor Link: no	Labor Link: yes
no	n	48176	90
	(row pct)	99.81	0.19
yes	n	422	190
	(row pct)	68.95	31.05

Pearson chi2(1) = 1.0e+04 Pr = 0.000

F

Labor Link		Land Link: no	Land Link: yes
no	n	48444	154
	(row pct)	99.68	0.32
yes	n	192	88
	(row pct)	68.57	31.43

Pearson chi2(1) = 5.5e+03 Pr = 0.000

Table 3: Summary Statistics of Link Characteristics

Variable	mean	std
At least one of pair holds office	0.22	0.41
Same religion	0.23	0.42
Same matrilineage	0.22	0.42
Same gender	0.52	0.50
Same soil type	0.32	0.47
Absolute value of age difference	10.61	11.61
Absolute value of wealth difference	0.75	1.29
Distance between plots (km)	1.27	0.78
Both members of 1st generation in village	0.05	0.21
Families trace origin to same region	0.55	0.50

Table 4: Predicting Network Links

Logit Regressions -- Dependent Variable = 1 if Link Exists

	1		2		3		3B		4	
	Information Links		Credit/Gift Links		Land Links		Land Links		Labor Links	
	odds	z	odds	z	odds	z	odds	z	odds	z
At least one of pair holds office	1.17	1.26	0.85	-0.78	1.46	1.92	2.30	4.09	2.17	2.62
Same religion	1.06	0.46	1.78	2.92	0.74	-1.19	0.74	-1.19	1.47	1.37
Same matrilineage	1.43	2.93	2.17	3.86	1.36	1.44	1.47	1.85	2.08	2.65
Same gender	7.29	11.57	3.95	5.75	2.54	4.42	3.41	5.56	3.89	4.41
Same soil type	3.09	8.44	1.36	1.55	5.40	6.75			3.39	3.70
Absolute value of age difference	1.00	-0.79	1.02	3.43	1.04	7.34	1.04	5.81	1.02	1.86
Absolute value of wealth difference	1.23	8.66	0.96	-0.62	1.29	5.22	1.29	6.05	1.08	1.25
Distance between plots (km)	0.56	-5.63	0.79	-1.68	0.30	-4.64			0.57	-1.90
Both members of 1st generation in village	0.37	-2.37	1.04	0.10	***	****	***	****	0.78	-0.33
Families trace origin to same region	1.09	0.65	2.20	3.22	1.26	1.01	1.31	1.21	1.26	0.75
Predicted probability of link	mean	std	mean	std	mean	std	mean	std	mean	std
if link=1	0.027	0.026	0.005	0.003	0.022	0.033	0.01	0.005	0.006	0.006
if link=1 & less than 5 years	0.014	0.014	0.005	0.004	0.004	0.007	0.003	0.003	0.002	0.002

standard errors heteroskedasticity consistent
n = 40960

***Both families migrants perfectly predicts no land link

Table 5: Predicting Network Links, with 2-way fixed effects

Logit Regressions -- Dependent Variable = 1 if Link Exists

	1		2		3		3B		4	
	Information Links		Credit/Gift Links		Land Links		Land Links		Labor Links	
	odds	z	odds	z	odds	z	odds	z	odds	z
Same religion	1.38	2.09	2.78	3.67	1.33	0.84	1.22	0.69	2.78	2.59
Same matrilineage	2.10	4.65	1.55	1.34	0.26	-2.60	0.34	-2.31	2.97	2.46
Same gender	6.02	9.15	4.01	5.05	1.89	1.99	1.80	1.89	2.01	1.44
Same soiltype	2.14	3.23	2.23	1.76	11.69	3.33			0.29	-1.35
Absolute value of age difference	1.00	-0.19	1.00	-0.34	1.09	4.02	1.07	3.92	1.06	2.02
Absolute value of wealth difference	0.87	-1.90	0.93	-0.26	1.19	1.03	1.19	1.16	1.43	1.07
Distance between plots (km)	0.35	-9.25	0.55	-3.02	0.13	-8.34			0.31	-4.06
Both members of 1st generation in village	2.66	1.73	0.53	-1.12	***	***	***	***	1.75	0.28
Families trace origin to same region	1.98	3.22	1.38	0.63	2.49	1.51	1.09	0.18	9.73	2.82
number of individuals dropped	247		338		335		335		358	
number of observations dropped	24696		33711		33463		33463		35786	

standard errors heteroskedasticity consistent

n = 40960

***Both families migrants perfectly predicts no land link

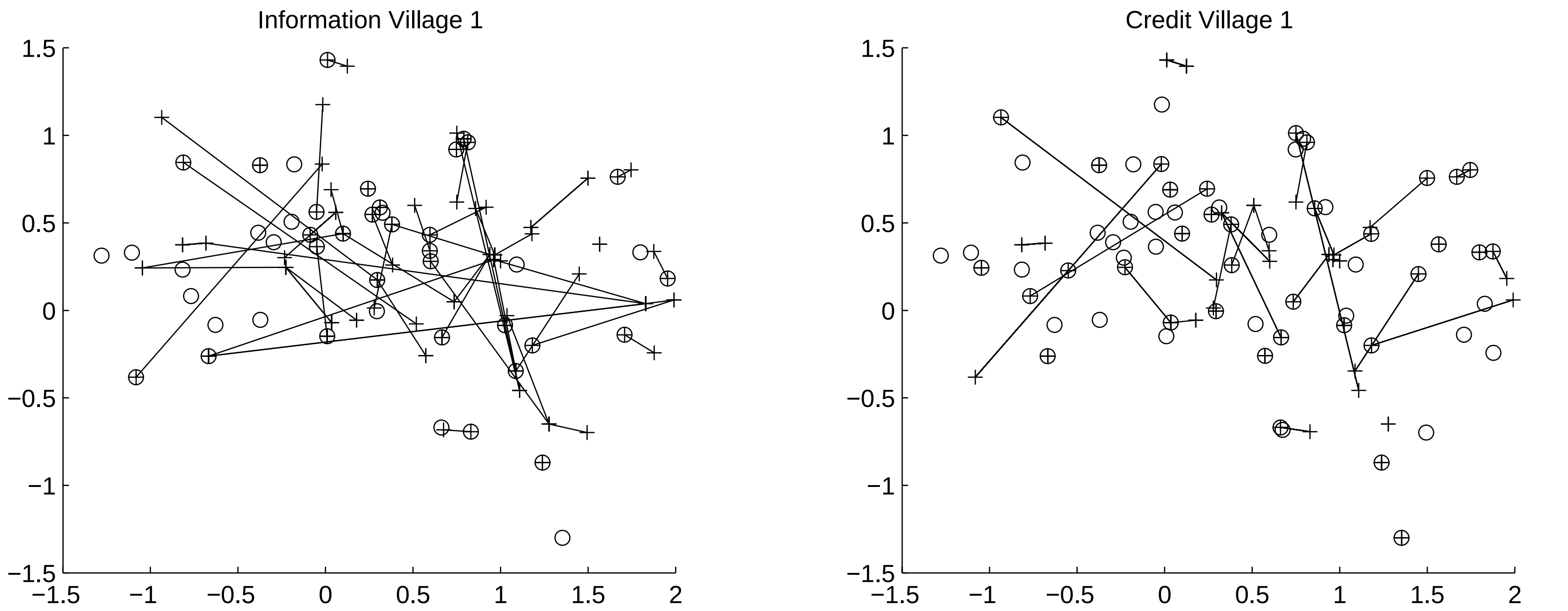

Figure 1A: Information and Credit Networks in Village 1

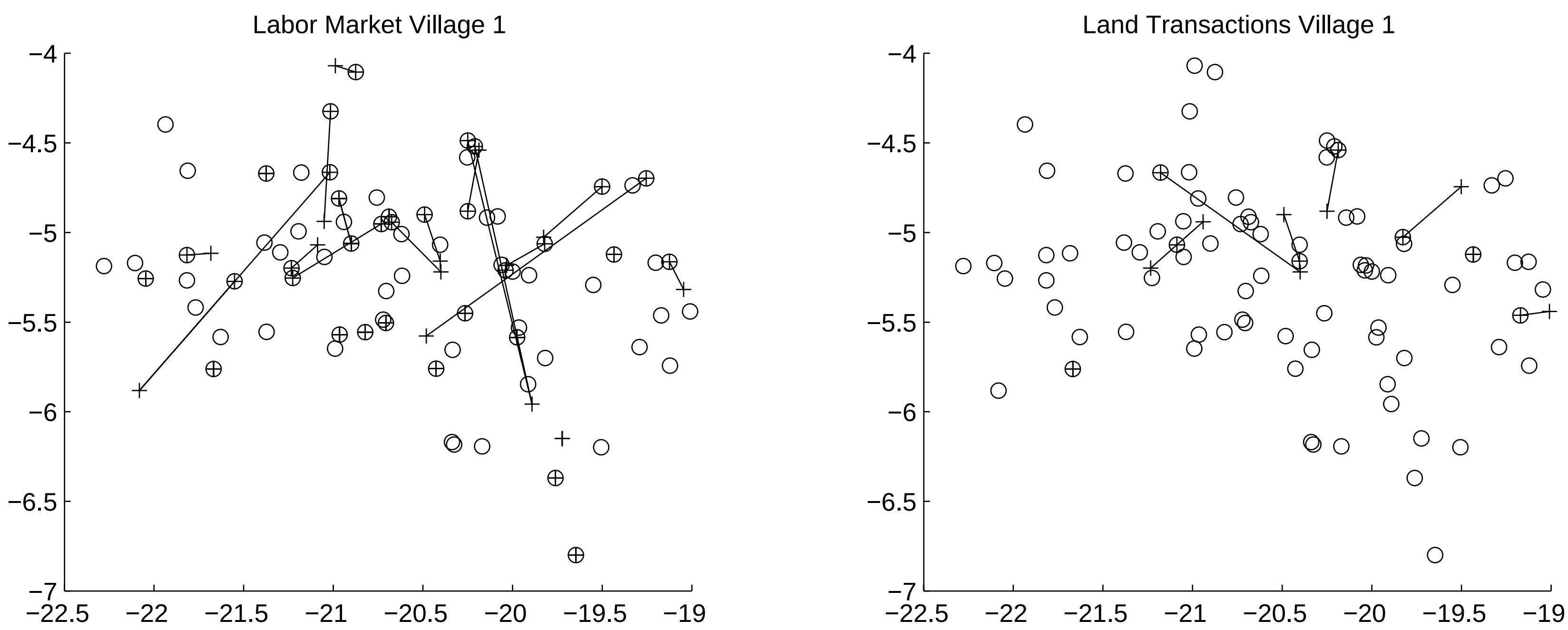

Figure 1B: Labor and Land Networks in Village 1

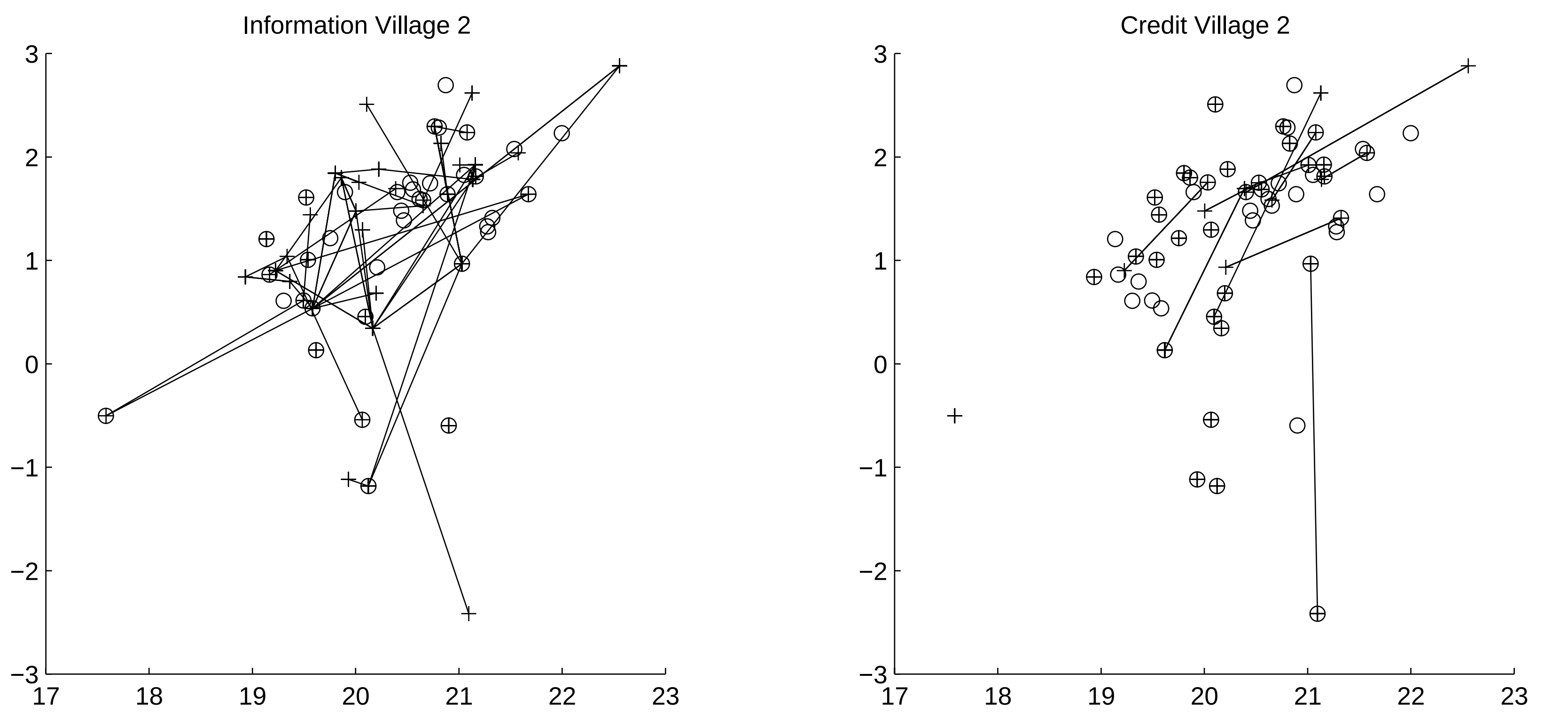

Figure 2A: Information and Credit Networks in Village 2

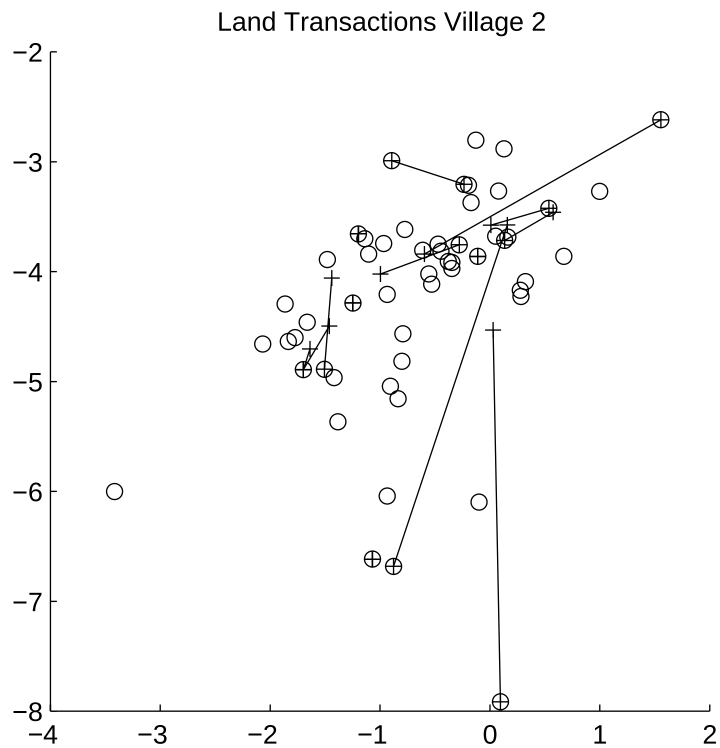

Figure 2B: Labor and Land Networks in Village 2

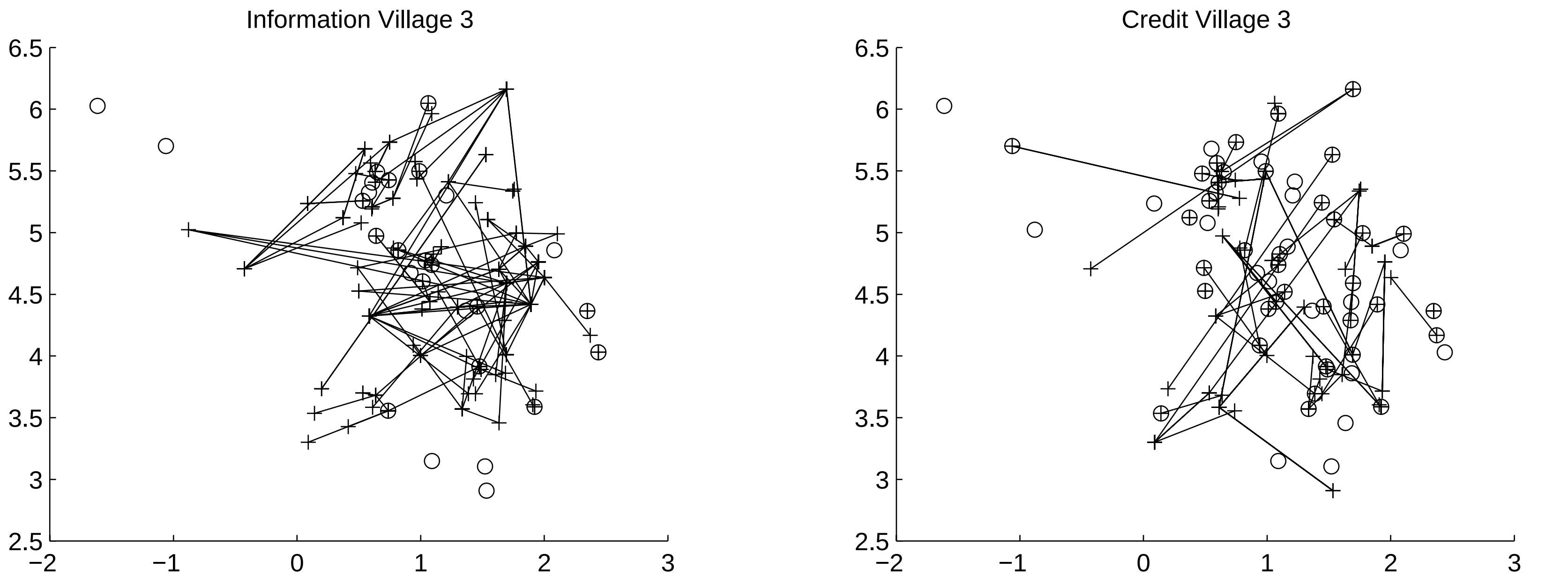

Figure 3A: Information and Credit Networks in Village 3

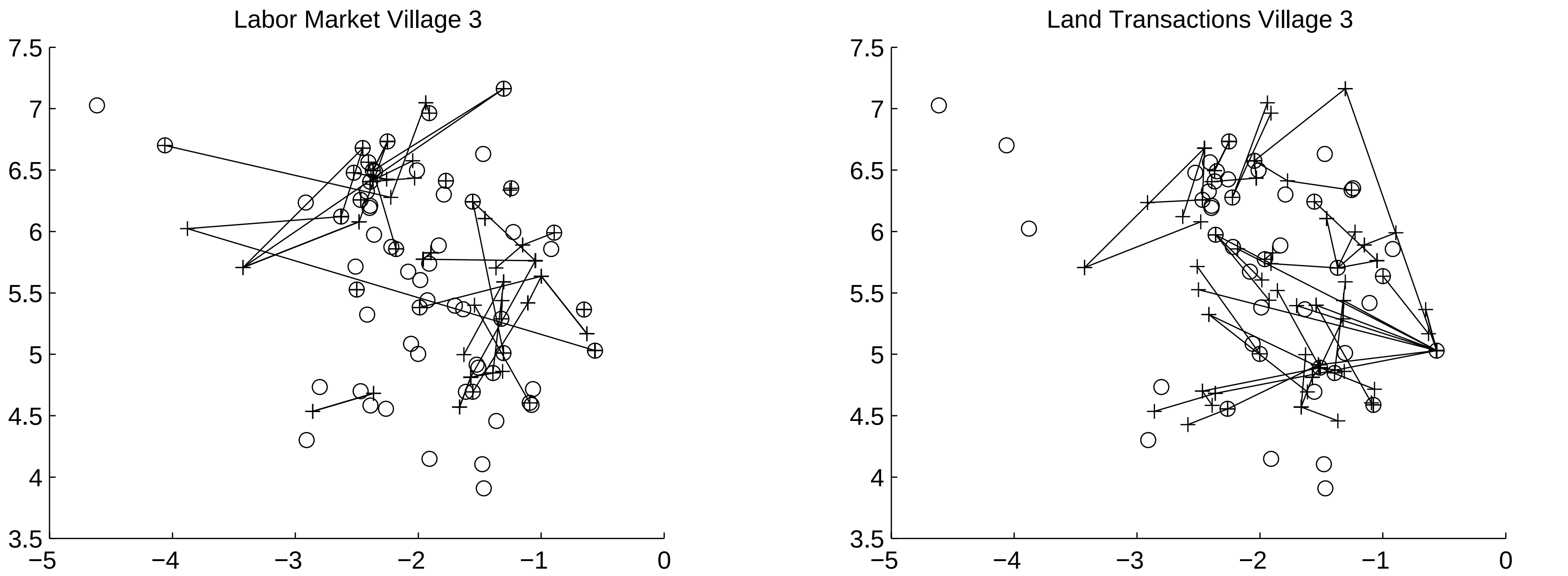

Figure 3B: Labor and Land Networks in Village 3

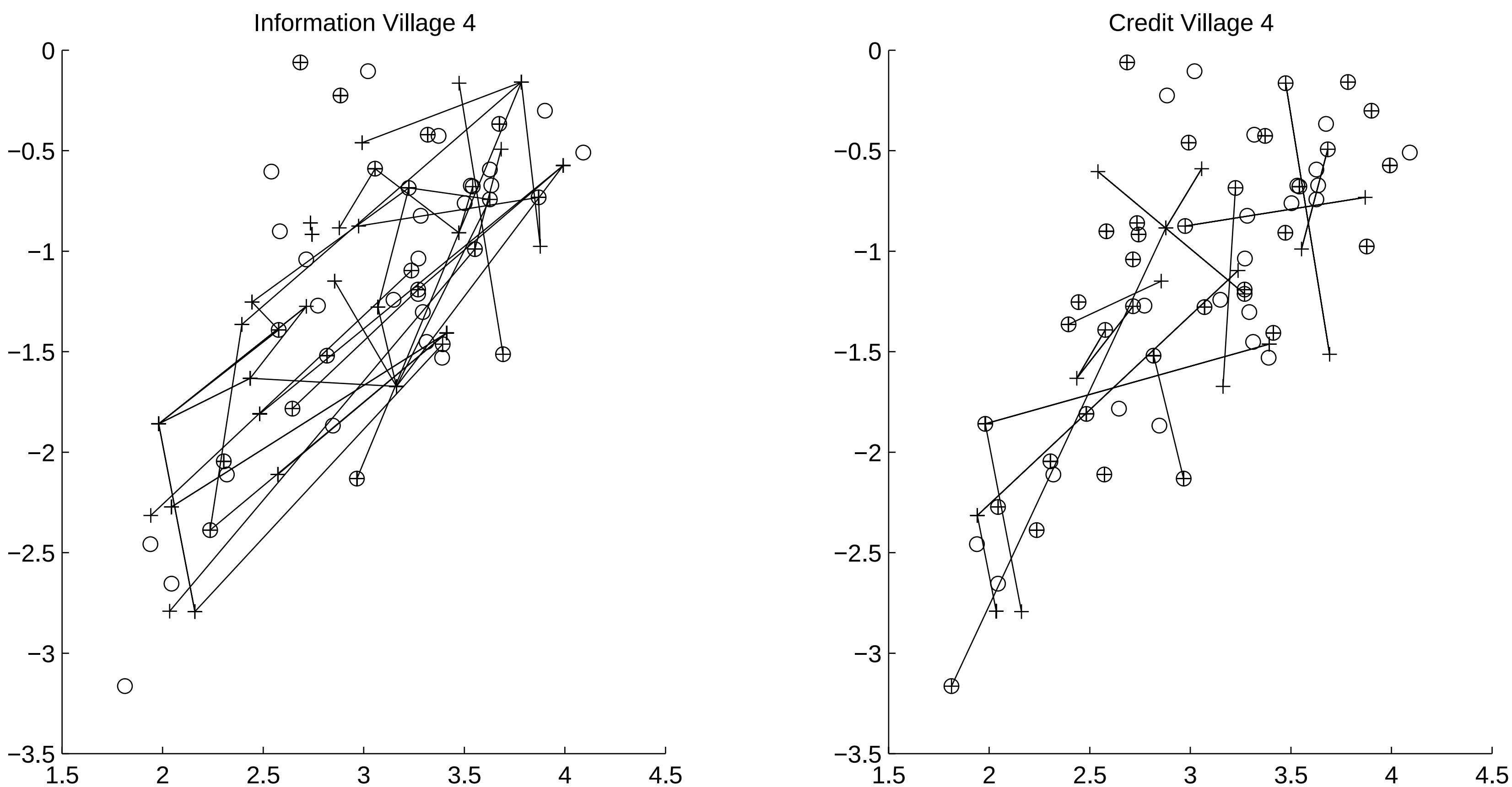

Figure 4A: Information and Credit Networks in Village 4

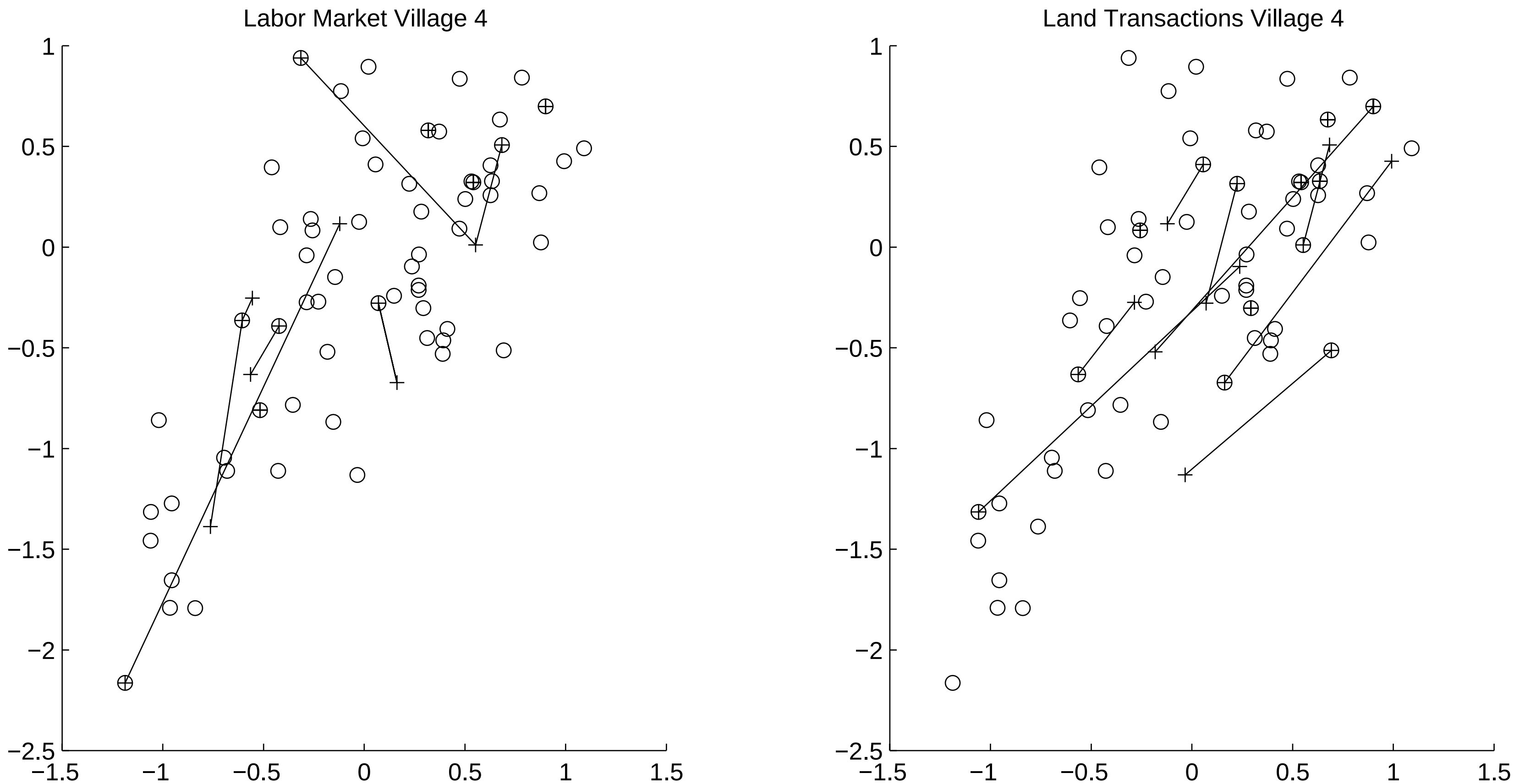

Figure 4B: Labor and Land Networks in Village 4