Interest and identity in network formation:

Who do smallholders seek out for information in rural Ghana?

Paulo Santos
Ph D Student, Cornell University
ps253@cornell.edu

Christopher B. Barrett
International Professor, Cornell University
cbb2@cornell.edu

May 2004 First Very Rough Draft
Comments Greatly Appreciated
Please Do Not Cite Without Authors' Permission

Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, July 1-4, 2004

We thank Chris Udry for making the data available and seminar participants at Cornell University for helpful comments. Santos thanks the Fundação Calouste Gulbenkian (Portugal) and the Fundação para a Ciência eTecnologia (Portugal) for financial support. Barrett thanks The Pew Charitable Trusts and the Christian Scholars Program of the University of Notre Dame for their support. Any remaining errors are solely ours.

Interest and identity in network formation:

Who do smallholders seek out for information in rural Ghana?

Abstract: In this paper, we use an unusually rich data set from Ghana to explore the endogenous formation of information network linkages among farmers. We propose and test a new measure of social distance that accommodates possible asymmetries in social distance. Using this improved measure, we show that social distance plays a major role in shaping network structure, but that other factors related to the inherent costs and benefits of linkage matter significantly as well. Network interlinkages appear relatively modest. We are also able to corroborate the sociological "strength of weak ties" hypothesis.

Key words: agricultural production, identity, markets, social networks,

1. Introduction

The idea that social interactions influence individual decisions by means other than the impersonal play of the market has made its way into economics, typically through the discussion of how membership in some group (through peer or neighborhood effects, moral norms and the like) affects individual decision (see, for example, the discussion in Durlauf (1999) or Manski (2000)).

Since the intersection of such affiliations defines one's identity (Breiger, 1990) this is an explanation that, under a different light, gained relevance in the economics literature with the work of Akerlof and Kranton (2000). The concept of social distance (Akerlof, 1997) and the tendency to deal with those who are similar to us (homophily) ties both approaches, by feeding back into how groups are formed and maintained.

Most of the extant literature nonetheless takes social interactions as exogenous. As Arrow (2000) puts it, in perhaps the clearest defense of such position, "[t]he concept of measuring social interactions may be a snare and a delusion. Instead of thinking of more and less, it may be more fruitful to think of the existing social relations as a preexisting network into which new parts of the economy (…) have to be fitted". Such an approach, however, moves from an under-socialized perspective of human agency that so many criticize in economics, to an over-socialized one (Granovetter, 1985) that economists typically associate with other social sciences. The nonrandomness of individuals' group memberships makes inference problematic.[1]

This reaction is not only inconsistent with the core economic principle that people respond to material self-interest, albeit not to material self-interest exclusively, it is also

[1] Manski (1993) emphasizes three problems in particular: the reflection problem, the problem of omitted variables, and the endogenous nature of social networks. This paper focuses on the latter problem.

contradicted by much of the literature in the other social sciences. Not only can group-level ascription be a choice motivated by material interest,[2] as in the religious conversion of the Orma, in Kenya, studied by Ensminger (1997), or changes in ethnic identity among the Fur in Darfur, Western Sudan, analyzed by Haaland (1969), but, more generally, the establishment of personal links between people often follow an implicit cost-benefit analysis.

For example, in analyzing the process of agricultural change in Ghana, Nigeria, Kenya and Zambia, Berry (1993) shows that investment in social relations was essential to gain (and maintain) access to productive resources: Because "In most African societies, social identity and status may be achieved as well as ascribed" and "partly because of the continued importance of social networks as channels of access to the means of production, many farmers invested part of their income in maintaining or advancing their position within established networks and/or gaining entry into new ones. *As a result*, membership and status in social networks have influenced the organization of agricultural production, the level of agricultural output and sales, and the structure of social relations within rural communities" (pp 159-160, emphasis added).

In recounting the 19th century history of the Lake Turkana region, Kenya, Sobania (1991) similarly describes how different relationships formed in response to a clear cost-benefit analysis of each social interaction, in particular through the potential

[2] The ascription to race underlied South Africa's apartheid system. The codification of such ascription is not always clear. Consider, for example, the passage from the official gazette once read on stage by Pieter-Dirk Uys, a South African comedian, "In terms of the Population Registration Act and in answer to a question from the Member of Parliament from Houghton, Mrs. Helen Suzman, five hundred and eighteen Coloreds were reclassified as Whites, fourteen Whites became Colored, seven Chinese became White, two Whites became Chinese, three Malays became Whites, one White became an Indian, fifty Indians became Colored, fifty seven Coloreds became Indian, seventeen Indians became Malay, four Coloreds became Chinese …" (*New Yorker*, May 10, 2004, p.75). Note that the reclassification is mainly into White, that is, in direction to the privileged position in South African society during the apartheid era.

role such relations played as part of a safety net to be mobilized in times of stress: "the gift of a head of stock was not an impulsive action but was rather both given and received as a compliment calculated to extend and individual's sphere of supportive relationships. Although unable to garner support in political matters or settlement of disputes from those friends who lived in neighboring societies, as would be the case with partners from within his own society, a herdsman's inter-societal partnerships greatly enlarged his knowledge of the region and his options in the economic sphere … when confronted with the risks and perils brought on by natural or men made disasters, such as droughts, disease and raids, the herdsmen could turn to his bond partners in addition to his kin and affines. When the difficulties of his intra-society coincided with his own, the individual who had invested in partnerships beyond the bounds of his own society continued to have options of assistance open to him." (p. 135-136).

The common thread of these and many other anthropological, historical and sociological accounts is the focus on a set of links between agents (a social network) that are purposefully chosen by the individuals involved . As Wellman and Berkowitz (1988a) put it, "social networks are the strings that simultaneously constrain our freedom and provide us with opportunities to take initiatives" (p. xii). The appeal of explicitly incorporating the role of human agency in the design and evolution of observable networks of human relations is that, again quoting Wellman and Berkowitz "[i]t immediately directs analysts to look at linked social relations, and frees them from thinking of social systems as collections of individuals, two-person dyads, bounded groups or simple categories" (idem, 1988b, p.4).

A small but growing empirical literature in development economics[3] addresses this issue of endogenous network formation relying heavily on the concept of social distance (Akerlof, 1997) and the tendency people have to deal more with those who are similar to us (homophily), rather than those who are different from us (heterophily). For example, Dercon and deWeerdt (2002) test for risk-insurance using data on the complete set of personal networks in a Tanzanian village while deWeerdt (forthcoming) explains the process by which these networks form. In a similar spirit, Conley and Udry (2002) study the process of learning about fertilizer application in Ghana while Udry and Conley (forthcoming), address the question of the formation of the networks that shape the access to information, credit, land and labor in that same region.[4]

This paper breaks from the literature in the way that we conceptualize social distance. In the economics literature to date, distance has been measured simply as a Euclidean norm to capture the magnitude of differences in any of several observable characteristics between network partners, with the most commonly used characteristic being physical location. Geographic distance has thus been the primary measure used. We employ a simple modification that allows for the possibility that the direction (or sign) of these differences also matters, that one's ordinal position with respect to a potential network partner can affect culturally defined norms of behavior and one's subjective evaluation of the benefits of establishing or maintaining a social link.

---

[3] With this choice, we miss the theoretical literature on models of network formation (see Jackson (forthcoming) for a review) and the (mostly US-based) literature on neighborhood effects, especially upon education outcomes (see Akerlof and Kranton (2003) for a review). Although both of those literatures provide some insights into this problem, space limitations prevent us from addressing them.

[4] To this list, one could add Fafchamps and Lund (2003), that study risk-sharing networks in the Philippines and Behrman, Kohler and Watkins (2001), that analyze the diffusion of family planning and AIDS worries in Kenya. Although with similar objectives, their approach to the endogenous nature of such networks makes no use of the concept of social distance and, as such, is quite apart from the discussion in this paper.

In the next section, we briefly discuss the data we use. These publicly available data were collected in the Eastern Region of Ghana by Chris Udry and Markus Goldstein and have been used by them and their collaborators in several important recent papers on the role social networks play in rural Africa. Section 3 then presents our econometric results, which support three major conclusions. First, both material interest and sociocultural identity help explain individuals' decision to contact an acquaintance to obtain information relevant for the solution of specific agricultural production or marketing problems. Second, we test the measure of social distance proposed in Section 2 against the standard, Euclidean norm approach and find that the former outperforms the latter in explaining three of the four networks we study and is statistically equivalent in the fourth. Third, we find that frequency of contact has an inverted-U effect on the probability of contacting someone for information, giving some support to the "strength of weak ties hypothesis" suggested by Granovetter (1973).

2. Data

We use data on economic activities and social interactions between people living in four villages in southeastern Ghana. The publicly available data[5] are discussed at length in Goldstein and Udry (1999).

This region has a long tradition of commercial agriculture and, in the early 1990s, initiated a process of conversion from cassava and maize production, directed towards domestic urban consumer markets, into pineapple production, directed to European export markets. The transition in crops and markets brought with it new inputs (primarily,

[5] The data, and the survey instruments, are available at http://www.econ.yale.edu/~cru2/ghanadata.html.

inorganic fertilizers) and a process of differentiation between those farmers who adopted the new crop (mainly, male and wealthier farmers) and those who did not.

Udry and Conley (forthcoming, p.3) summarize this dynamic by saying that "economic development in this region is being shaped by the networks of information, capital and influence that permeate these communities." They are able to quantify the importance of social networks – for example, in learning about fertilizer application (Conley and Udry (2002) – because the survey collected very detailed data on respondents' patterns of personal interaction. We use those data to study endogenous network formation among these Ghanaian farmers.

Social networks information was collected in two ways. First, respondents were asked to identify those with whom they had significant discussions on agricultural matters. Second, respondents were matched with seven individuals randomly chosen from the sample[6] and asked about the possibility of addressing them when faced with some specific problem, through the following questions:

"Could you go to ___ if you had a problem with unhealthy crops?"

"Could you go to ___ for advice about when to apply a new kind of fertilizer?"

"Could you go to ___ if you wanted to discuss changing your method of planting?"

"Could you go to ___ if you wanted to find a buyer for any of your crops?"

We use the answers to these questions to indicate the possibility of a link between the two individuals, i.e., the answers reflect the potential network of each respondent, not

[6] The respondents were also non-randomly matched with three individuals considered to be focal in their village (a farmer in the survey and two other persons not previously surveyed). We do not use data on those three prospective matches, both due to its non-random nature and the fact that information is generally not available on most of the focal individuals, making the sort of analysis we perform impossible.

the actual one. Given that we care about the determinants of these networks (and not their actual benefits), this does not seem to be a limitation[7]. There were three possible answers, "yes", "no" and "yes, but he wouldn't know", although the last choice was never reported. Table 1 presents the answers to these questions, disaggregated by whether or not the respondent knows his or her match.

[TABLE 1]

Three key facts emerge from this Table. First, not everyone knows everyone else, even in a small, rural village setting. This calls into question the widespread practice of using common village membership as a proxy for a social connection. Second, not knowing the matched individual effectively prevents respondents from addressing the matched individual in order to gain access to information. Pre-existing social ties are plainly a necessary condition to obtaining information through informal channels. Third, knowing someone does not mean that one can or would go to them to ask for information. Combining the second and third points, prior knowledge of someone is a necessary but not a sufficient condition to approach them for information. Thus the shape of the network through which rural Ghanaian farmers obtain information is plainly social yet determined by factors other than just prior association.

[7] The other source of information on network structure in this data set, the question on with whom did the respondents' had significant discussions on agricultural matters, does not provide a better guide, for two reasons. First, not all the listed individuals were part of the sample, hence no information on their identity is available. Second, the period when each of the listed individuals was contacted is not known (making impossible the estimation of differences in characteristics that change over time, such as experience or wealth). One other possibility could be to use the answers to the question "Have you ever gone to X for advice about your farm?", where X is one of the random matches. This possibility poses similar problems to the one just described, to which one may add that a negative answer may just indicate that there were some other source of information that the respondent valued more and on which we may not have any information (including its existence). This may account for the small number of positive answers (around 10% of the total number of matches), which adds the additional problem that with such an imbalance in the structure of the dependent variable, the estimates of the parameters of a probit model, like the ones that will be estimated, become quite unreliable.

In order to explain the decisions summarized in Table 1, we use data on variables such as clan membership, gender, age, formal education, non-land wealth,[8] experience with different crops,[9] sources of income other than farm production (non-farm wage or salaried employment and self-employment) and occupation. We take these variables as indicators of the costs and benefits of establishing the information link, both as attributes of the respondent and as measures of distance between the respondent and the match.

In the literature on social distance, the Euclidean norm appears to be the only measure used. Implicitly, its use imposes symmetry on the effect of differences between the two parties to a link. For example, taking wealth as the relevant dimension, the use of Euclidean distance assumes that wealth ordering is irrelevant to the incentives to establish a link, that a rich farmer faces the same costs and benefits of linking to a poor farmer as the same poor farmer does to the same richer man.

In order to avoid such an unnecessarily restrictive assumption, we measure social distance as a simple modification of the Euclidean norm, through the definition of a pair of indicator variables. Let X be any of the non-categorical variables (age, wealth, agricultural experience, non-agricultural sources of income) on which information is available. We then measure the distance between the respondent i and the match j by the following two variables:

---

[8] Non-land wealth was estimated as the sum of the values of the following assets: foreign currency, cash in bank accounts, bonds, susu and esusu, livestock, crops, seeds, chemicals and farm equipment. The definition of individual wealth is made difficult by the number of missing answers in the value of some assets, specifically jewelry (244 missing answers), cash on hand (66) and clothes (45), even when we use the inventory of assets from round 1, when all respondents were interviewed. Non-farm equipment was enumerated only in rounds 9 and 15, while stocks for trade were only captured in round 2, and with a high number of missing values (360). We therefore omit these latter categories of assets from the non-land wealth estimates.

[9] In the survey, an important number of respondents indicated their experience with some crops as "more than x years". We approximated such information by taking their experience with such crop to be "x years".

$$I_{(X_i - X_j < 0)} * |X_i - X_j| + I_{(X_i - X_j \geq 0)} * |X_i - X_j| \qquad (1)$$

where $I_{(\bullet)}$ is an indicator function taking value one if true, zero otherwise. Clearly, one of the two indicator variables in the distance definition (1) equals zero.

For the categorical variables (gender, occupation, migrant status), distance is defined by a set of dummy variables that consider the several possible characterizations of the match. Hence, for clan, we define only the variable "share_clan" (that takes the value 1 if both respondent and match belong to the same clan) but a complete characterization of the effect of gender requires the definition of the following four variables:

mm = {1 if i = male and j = male; 0 otherwise}

ff = {1 if i = female and j = female; 0 otherwise}

mf = {1 if i = male and j =female; 0 otherwise}

fm = {1 if i = female and j =male; 0 otherwise}.

We explore the effects of differences in formal education by defining a dummy variable designated "literacy" that accounts for the case when i is illiterate and j is literate.[10] Table 2 presents descriptive statistics on the set of explanatory variables.

[TABLE 2]

As in the extant literature, we interpret our measures of social distance (that is, differences in gender, age, wealth and education), conditioned on differences in experience and importance of non-agricultural activities, as reflecting the cost of establishing a link. Hence, as a rule, we would expect negative coefficient estimates on these variables reflecting the fact that greater differences between people tend to discourage individuals from investing in establishing interpersonal links.

[10] Illiterate is defined as never having attended school; literate is its complement.

Nonetheless, it is not obvious whether the variables associated with non-agricultural activities should encourage or discourage information links related to agricultural production and marketing. Can someone with a non-farming occupation (say, a teacher) be contacted in order to provide information on some problem related with agricultural production? In principle, one would expect the teacher's direct experience would not be especially helpful to a farmer. However, the teacher's capacity to access information that may not be available to a farmer could also be thought valuable *a priori*. A similar argument can be made about the value of farming experience, which may be valuable in dealing with commonplace problems, but which may not be so valuable in dealing with disequilibria, in which case, formal education (that we included, in this context, only as a variable that signals social position) could be more important.[11]

We analyze two further aspects of the structure of these information networks, the first suggested by the data itself and the second suggested by the literature on social networks. First, we explore the interlinkage between different networks and the effect of the existence of an information link between both individuals on some other matter on the probability of asking for advice on a different question.

Table 3 summarizes the possible combinations of answers to the set of information link questions asked of each respondent with respect to each randomly chosen match individual.

[TABLE 3]

The data show that more than 82 percent of the responses are all or nothing, either don't ask the match for advice on any of the problems (58.0 percent) or ask the match for

[11] See Barrett et al. (2004) on the importance of education in dealing with disequilibria in west African farming systems.

advice on all of them (24.2 percent). Let us focus just on the positive answers. No matter the question, more than two-thirds of the positive responses are associated with positive responses to each of the other questions as well, signaling interlinkage between different networks.

It is also apparent that although respondents rarely choose to contact a match for advice on only one of the three problems related to agricultural production (unhealthy crops, fertilizer application and planting technique), they are considerably more likely to make a link solely for advice on crop marketing. Networks interlinkage would appear to depend in part on the degree to which the explicit purposes of the links are closely related.

Second, we explore Granovetter's classic "strength of weak ties" hypothesis, that the most valuable connections people have are those that they exercise infrequently. The strength of weak ties hypothesis can be tested using respondents' answer to the question

"In a normal month, how often do you talk with ___?"

where we take frequency of contact as the measure of the strength of the link. Figure 1 presents a smoothed frequency distribution of answers to this question, where several focal points plainly emerge, around 0 times per month, weekly, semi-weekly, and daily contact.

Figure 2 displays the smoothed frequency distribution of "yes" answers to each of the link questions conditional on frequency of contact. The frequency with which respondents indicate they would ask for advice from the random match seems to increase at a rapidly diminishing rate in frequency of contact. This suggests that the marginal effect of a weak tie – as signaled by very infrequent contact (e.g., weekly or less) with the match – on the likelihood of establishing a link is far higher than the marginal effect of increased contact with a strong tie.

[FIGURE 1]

[FIGURE 2]

3. Econometric estimation

To explain the decision to establish a link, we estimate the model

$$l_{ijk}^{*} = X'\beta + u_{ij} \qquad (2)$$

where l_{ijk}^{*} denotes i's propensity to establish a link with j in order to get access to information on problem k, X'is the matrix of explanatory variables described above, β is the corresponding parameter vector and u_{ij} is a normally distributed error term. We cannot observe the latent variable l_{ijk}^{*} but can observe the dichotomous variable l_{ijk} defined as

$$l_{ijk} = \{ 1 \text{ if } l_{ijk}^{*}>0, 0 \text{ otherwise}\} \qquad (3)$$

and that takes the value 1 when the answer to the k^{th} question is "yes". We further assume

$$E(u_{ij}) = 0 \text{ , } Var (u_{ij}) = 1 \qquad (4)$$

$$E(u_{ij}, u_{ih}) \neq 0 \text{ if } j \neq h \text{ , } E (u_{ih}, u_{ih}) = 0 \text{ if } i \neq j \qquad (5)$$

$$E(X,u_{ij})=0 \qquad (6)$$

These assumptions imply a probit model estimated by clustering the observations on the identity of the respondent and taking the explanatory variables as exogenous.

Because we don't have information on all variables for all individuals/matches, we must drop approximately one-third of the observations. Nonetheless, a Kolmogorov-Smirnov test of the equality of the distributions for all dependent and independent variables does not support the rejection of the null hypothesis that the distribution of the variables for which we have complete information is from the same as the original one.[12]

We start by estimating the determinants of these networks, as revealed by the answers to each of the four prospective information link questions stated above. As shown in Table 1, the decision to address one's match is shaped by previous knowledge of the match, where a negative answer to "Do you know ___?" determines a negative answer to the following questions. The apparently sequential choice process leads us to estimate the determinants of these networks using only the subsample of those who know their match (Maddala, 1983, p 124).

But since the question of who knows whom is of interest in its own right, given the commonplace assumption of the village as a natural unit of analysis in much of the development literature, we start by addressing it.

a) Who knows whom?

Table 4, column A presents the results of the probit estimation for the dependent variable reflecting knowledge of the random match.[13]

[TABLE 4]

The model performs remarkably well, correctly predicting individuals' knowledge of their random matches in more than 90 percent of cases. Individuals with greater wealth and net business revenue are more likely to know others, although individuals are

[12] We repeated this test for subsequent exclusions of observations, namely when we only use the observations for which the respondent knows his match, and we could not reject the null hypothesis in any case.

[13] Given the large size of these results tables, we skip the usual presentation of the marginal effects of explanatory variables. The coefficients measure the impact of changes in the explanatory variables upon the latent (unobservable) variable and thus offer meaningful inference on the sign of the effect upon the binary variable and allow for comparisons of the relative effect of different explanatory variables. Detailed results on the marginal effects of each explanatory variable are available from the authors on request.

significantly more likely to know people wealthier or earning less than themselves, as reflected by the coefficient estimates on the wealthier/poorer and net_rev_more/net_rev_less. A teacher will more likely be known but was no special advantage in knowing other people. Migrants are less likely to know nonmigrants and women are less likely to know men.

One of the striking things about these results is the asymmetric effect of variables such as gender, migrant status, wealth, net revenue and occupation.

Table 4, column B presents the probit estimates of the same equation (2) with social distance now measured more traditionally, as the absolute value of the difference, rather than using the modification we propose to accommodate potentially asymmetric costs and benefits between individuals. Although the number of explanatory variables that differ between the two models is relatively small, the results in column A tell a richer and more compelling story, as the interpretation of the coefficients associated with the wealth and net revenue variables make clear. We can test these alternative approaches more formally by defining

$$l_{ijk}^{*} = X'\beta + Y'\theta + u_0 \qquad u_0 \sim N(0,1) \qquad (7)$$

$$l_{ijk}^{*} = Z'\gamma + Y'\theta + u_1 \qquad u_1 \sim N(0,1) \qquad (8)$$

where X is the vector of differences as defined in equation (1), Z is the vector of absolute values of the differences and Y is the vector of variables that is common to both specifications. We test the adequacy of each model by checking their capacity to explain over and above what is explained by the alternative specification: let l_{ijk}^{1} be the predicted value of l_{ijk}^{*} from equation (7) and let l_{ijk}^{2} have an analogous interpretation, from equation (8). After estimating the equation

$$l_{ijk}^{*} = \alpha l_{ijk}^{2} + X' \beta + v_0 \qquad v_0 \sim N(0,1) \qquad (7')$$

we can test whether the set of variables X has any explanatory capacity through a Wald test of the null hypothesis that the vector B is statistically equal to zero. The same can be done with the model in equation (8) through estimation of the equation

$$l_{ijk}^{*} = \alpha l_{ijk}^{1} + Z' \gamma + v_1 \qquad v_1 \sim N(0,1) \qquad (8')$$

and testing for the joint significance of D. If we can reject the null hypothesis that β is statistically equal to zero and cannot reject the null hypothesis that γ is statistically equal to zero, then our proposed approach to measuring social distance outperforms the standard method in these data. The two models will be statistically equivalent if we cannot reject either null hypothesis or if we can reject both.

When the dependent variable is interpersonal knowledge, there seems to be no statistically significant different between the two methods. The Wald statistics for the null hypotheses that D=0 and B=0 equal 2.40, which has a p-value of 0.966 against the X^2_8 distribution, and 17.43, which has a p-value of 0.625 against the X^2_{20} distribution, respectively. While the interpretation of the asymmetric measures makes more intuitive sense and performs slightly better in predicting which matches a respondent will already know, those differences are not statistically significant. As we see in the next section, however, the asymmetric social distance measures significantly outperform the more standard, symmetric ones when we study the information links people choose to make.

b) Asking for information

Table 5 summarizes the probit estimates of the decision of approach one's match in order to obtain information in response to the four questions above.

[TABLE 6]

Wald tests of the joint statistical significance of particular sets of variables show that one's attributes are not, as a rule, important to explain information links, as can be seen by the values of the associated Wald statistic. This is especially true for those problems that as less ordinary (fertilizer application and marketing) and is less true for information on planting techniques (where the respondents' attributes are jointly significant at the 15% significance level). The lone exception relates to queries regarding unhealthy crops, where own attributes (just) matter at the 5% significance level.

By contrast, each set of variables that measure the costs and benefits of a link in terms of differences between the respondent and his or her match are jointly statistically significant, usually at levels of significance below 1%. These two results – that the respondent's own attributes do not matter significantly to establishing a link, but that differences between the prospectively linked individuals matter a great deal – suggest that relative social position matters most.

Social distance between the respondent and match provides the strongest explanation for the choice to establish an information link, moreso than difference in agricultural experience, in non-agricultural activities or in occupations. Co-residence in the same household sharply increases the likelihood that one would ask the match. Differences in migrant status matter and reasonably symmetrically, with migrants less likely to ask nonmigrants and vice versa, with point estimates that are not statistically significantly different from one another. Conditional on knowing each other – which, recall from Table 4, is heavily impacted by wealth – people are more likely to ask questions of those who are less wealthy than themselves than of those of equal or greater wealth.

Gender differences matter a lot with respect to agricultural production matters, but not with respect to marketing. Perhaps more interesting, there appear to be strongly asymmetric effects of gender differences on individuals' incentives to establish information network links. Men are very reluctant to ask agricultural production questions of women, and women are less likely to ask such questions of each other, albeit statistically significantly so only with respect to plant health. But a woman is more likely to ask questions of a man than is another man.

Differences in experience with pineapple production seem to significantly affect respondents' search for information. Farmers appear significantly more likely to ask questions of matches with more experience than they have and less likely to ask questions of those with less experience. The differences are much larger in both magnitude and in statistical significance with respect to experience in pineapple, a new crop in the region, than in maize, a long-established staple crop. This sort of pattern is likewise intuitive and consistent with basic models of learning.

Differences in non-agricultural activities and in occupation likewise matter to farmers' propensity to ask questions of their random matches. Those who work more off-farm are considerably more likely to ask questions of those matches they know who spend less time off-farm and more time in agriculture than they do. Conversely, people are less likely to ask agricultural production questions of teachers or traders but far more likely to ask agricultural marketing questions of teachers. This suggests that those who are regarded as more detached from ordinary farming problems can nonetheless be seen as “bridges” towards solutions to nontraditional problems. This interpretation, which hints at the strength of weak ties hypothesis suggested by Granovetter (1973), will be tested more formally, using information on the frequency of contact, in section 5.

As the preceding discussion implies, the asymmetric effects of social distance appear significant in understanding the endogenous formation of information networks in rural Ghana. We establish this more formally following the statistical approach enumerated earlier, based on estimating equations (7’) and (8’) and then testing the exclusionary restrictions on the unique components of each regression.[14] Table 6 displays the results of these tests. The asymmetric social distance measure dramatically outperforms the symmetric measure based on simple Euclidean distance in each of the three models related to agricultural production information. Although the asymmetric treatment performs modestly better with respect to agricultural marketing information as well, there is no statistically significant difference between the two social distance measures with respect to that dependent variable.

[TABLE 6]

The differences in determinants of network linkages between information related to agricultural production and to agricultural marketing (Tables 3, 6 and 7) raise the question of the degree to which information networks interlink. Are people more likely to ask a question of a randomly matched individual the more likely they are to ask a different question of that same individual? Or do people target questions differentially at distinct individuals, building different networks with relatively limited interconnectedness? We turn now to explore this question.

[14] To conserve space, we omit the estimation results of the models that use the Euclidean norm measure of distance. They are available upon request.

4. Interlinked networks

As mentioned above, the similarities between the three different agricultural production-related information networks are quite striking. We test for their interlinkage by re-estimating the above models, now also including as explanatory variables the fitted values of the probability of establishing a link in the remaining models. That is, we estimate the system of equations

$$l_{ij\text{-}k}^{*} = X'\,\beta + u_0 \qquad u_0 \sim N(0,1) \qquad (9)$$

$$l_{ijk}^{*} = l_{ij\text{-}k\sim}\mu + X'\,\beta^{*} + u_1 \qquad u_1 \sim N(0,1) \qquad (10)$$

where $l_{ij\text{-}k}^{*}$ is i's propensity to address j in order to get information to solve any of the problems other than the k^{th} problem, while $l_{ij\text{-}k\sim}$ is defined as

$$l_{ij\text{-}k\sim} = \text{prob}\,(l_{ij\text{-}k}^{*}>0) \qquad (11)$$

The estimated coefficients on the fitted $l_{ij\text{-}k}$ regressor are presented in Table 7, together with the value of the Wald statistic of the test of the joint null hypothesis that all other variables (those enumerated in Table 5) equal zero.

[TABLE 7]

We can define these networks as interlinked when our estimates from equation (10) permit us to reject the hypothesis that the μ parameter relating l_{ijk}^{*} to $l_{ij\text{-}k}^{*}$ equals zero while at the same time failing to reject the null hypothesis that $\beta^{*} = 0$, i.e., that other network linkages fully explain an individual's likelihood of establishing an information link to a randomly matched person.

Given this definition, we can only conclude that the decision to ask for advice on fertilizer application is not interlinked with other decisions to establish an information network. As for the other prospective information topics, we either cannot reject the null hypothesis that the other network linkages do not add to our understanding of links regarding questions of unhealthy crops or planting technique or we can reject the null hypothesis that once one controls for the other network links, the remaining covariates have no statistically significant relation to the likelihood of asking about agricultural marketing. [15]

Nonetheless, since there is no reason to expect that differences in identity would impose costs of establishing a link that would differ with the problem about which a respondent might inquire of his or her match, this result strikes us as further evidence of the asymmetric benefits of establishing an information link in one's network, with the asymmetry in this case arising from relatively modest differences in the nature of the information sought.

5. Strength of ties

The evidence in Figures 1 and 2 and Table 5 raises the intriguing possibility that Granovetter's classic "strength of weak ties" hypothesis might find support in these data. In formulating our hypothesis, we follow the original intuition of the author (Granovetter, 1973): "A natural *a priori* idea is that those with whom one has strong ties are more motivated to help with job information. Opposed to this greater motivation are the structural arguments I have been making: those to whom we are weakly tied are more

[15] The nonlinearity of the first stage estimate of $l_{ij\text{-}k\sim}$ obviates the collinearity problem that would otherwise exist, permitting identification of d.

likely to move in circles different from our own and will thus have access to information different from that which we receive”.

Of course, it is not exactly clear how one ought to measure the strength of a tie. In the original exposition of this hypothesis, Granovetter (1973) writes that “most intuitive notions of the “strength” of an interpersonal tie should be satisfied by the following definition: the strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie.” In an early review of studies that tried to test this hypothesis, Granovetter (1982) identifies two major ways of operationalizing the concept of “strength of tie”: (i) frequency of contact, as we use here and as did Granovetter (1973), and (ii) the assumption that ties with different people (e.g., kin, friends, colleagues and acquaintances) had inherently different strength, something that is not directly measurable in these data.

Instead of assuming that the probability of establishing a link is a monotonically decreasing function of frequency of contact, we posit that such relation may be non–linear. As is plain from Table 1, people need to already know someone in order for them to feel comfortable approaching them for information or advice. So there is plainly a sharp increase in the likelihood of establishing a link as one makes initial contact. The salient question is how much additional contact increases people’s propensity to establish informational links. Up to some point, more frequent contact may permit better evaluation of the costs and benefits of the link (e.g., the likely accuracy of the information obtained and the motivation level of the match to respond to a request for assistance), after which point high frequency interaction may signal that little new or relevant information can be transmitted by the potential contact.

In order to test the "strength of weak ties" hypothesis we therefore estimate the following system of equations:

$$z_{ij} = X'\ \delta + v_{ij} \qquad v_{ij} \sim N(0, \sigma^2) \qquad (14)$$

$$l_{ijk}^{*} = \alpha\, z + \lambda\, z^2 + X'\, \beta + u_{ij} \qquad u_{ij} \sim N(0, 1) \qquad (15)$$

where z_{ij} is frequency of contact between i and j, X is the same vector of explanatory variables as before, and z and z^2 are, respectively, the predicted value of frequency of talk and its square, obtained by estimating equation (14).[16]

In order to obtain the fitted values of frequency of talk, we regressed this variable on the same variables as in the previous models, but using only a subsample of the original data. In particular we excluded those observations with a abnormally high number of "average number of talks in a normal month".[17] The results are presented in Table 8.

[TABLE 8]

[16] Notice that the parameter α is identified by the statistical significance of the variable "teacher" in Table 8 (where we explain the number of talks in an average month) and its absence in Table 6 (where we explain the decisions of establishing a link for any of the four problems under analysis). Although this is a rather ad-hoc process, motivated by the lack of instruments that can, reasonably, separate the frequency of contact from the propensity to ask for advice, it is supported by our general result that own attributes seem not to matter in the models presented in Table 6. The re-estimation of these last models after the exclusion of other attributes (such as "cocoa" or "cassava"), in order to increase the precision of the estimates, produced the same results.

[17] We excluded those observations where the frequency of contact in an average month was in the range 60-300 times; this represents less than 2% of the total number of observations. As above, we performed a Kolmogorov-Smirnov test of equality of the distributions of all dependent and independent variables and failed to reject the null hypothesis that these distributions were identical.

As one would expect, the frequency with which the respondent and his or her random match speak is sharply increasing if they reside in the same household. Men speak to each other more frequently than do women with either men or women. Teachers and traders speak with people more frequently than do people in other occupations. And, in general, the frequency of contact decreases with social distance along any of several dimensions: wealth, income, occupation, etc.

Table 9 summarizes the coefficient estimates associated with the variables z and z^2. These estimates indicate precisely the hypothesized strong concave relation between frequency of contact and likelihood of establishing a link. The last row in Table 9 presents the estimated number of talks per month that maximizes the probability of contacting one's match. The estimated marginal effect is shown in Figure 3 for the case of asking about fertilizer application. The marginal effects for the other three questions look very similar.

[TABLE 9]

[FIGURE 3]

In the case of traditional agricultural production questions (concerning planting techniques and plant health), about which farmers have less to learn, the frequency of contact that maximizes the likelihood that someone asks a question of a match is almost 25% larger than in those cases where new or more extraordinary information is sought d (regarding fertilizer application and finding a buyer). This is consistent with the sociological observation that heterophilous relations – contacts with people less like oneself, which occur less frequently – are especially valuable in addressing novel situations. Farmers trade off social proximity for access to new information as most appropriate to the question at hand. In all cases, however, the greatest marginal effect on the likelihood of making the information network link occurs at very low frequency of contact. Weak ties indeed seem to have strong effects.

6. Conclusions

The structure of one's social networks has been increasingly recognized as a crucial determinant of access to information, credit and, more generally, influence. Although social networks' role in shaping decisions and outcomes seems important, quantification of this effect has proved relatively elusive, not least of which because analysts have typically been unable to control adequately for the obvious endogeneity of networks.

This paper argues for the need to treat social networks as endogenous and for the importance of recognizing the asymmetries that underlie the concept of social distance as it relates to the establishment of information network linkages. We propose and test a new measure of social distance that accommodates possible asymmetries in social distance and find that, for most of the problems analyzed in this paper, this new measure outperforms the Euclidean norm most commonly used in the extant literature.

Using this improved measure of social distance, we stress the need to consider explicitly the cost-benefit calculus in which agents engage when deciding whether or not to link to another individual. Networks are not exogenously determined on the basis of inherited identity; they are a direct product of individual choice, albeit choice that is conditioned by social distance that is predetermined (e.g., with respect to occupation) or exogenous (e.g., with respect to gender). Our results show that social distance plays a major role indeed in shaping network structure, but that other factors related to the inherent costs and benefits of linkage matter significantly as well.

Reinforcing this point, we find that in spite of the evident similarities in the structure of the four networks we study and although social distance is exactly the same between each respondent and his or her random matches in each prospective network, there are statistically and economically significant differences between these networks. The costs of establishing a link with respect to social distance should be the same whatever the problem, yet the benefits will vary across the question at hand. Thus, for example, people more actively seek out particular types of individuals – teachers, traders, those with whom they have less frequent contact – when faced with nontraditional questions than with problems with which they have much prior experience themselves. We interpret these differences as evidence that cross-sectional variation in the benefits of links affect agents' behavior in constructing and maintaining information networks. This point is further reinforced by our empirical corroboration of the sociological "strength of weak ties" hypothesis, that there exists inverted-U effect of frequency of contact upon the decision to form a network link, with the greatest marginal effects occurring at the lowest frequencies of interpersonal contact.

References

Akerlof, George (1997), Social identity and social decisions, Econometrica, 65, 5, 1005-1027

Akerlof, George and Rachel Kranton (2000), Economics and identiy, Quarterly Journal of Economics, 115, 3, 715-753.

Arrow, Kenneth J. (2000), Observations on social capital, pp. 3-5, in Partha Dasgupta and Ismail Serageldin (eds), *Social Capital. A multifaceted perspective* (Washington:World Bank).

Barrett, Christopher B., editor (forthcoming), *The Social Economics of Poverty: Identities, Groups, Communities and Networks* (London: Routledge).

Barrett, Christopher B., Shane M. Sherlund and Akinwumi A. Adesina (2004), "Macroeconomic shocks, human capital and productive efficiency: Evidence from West African farmers ," Cornell University working paper.

Behrman, Jere R., Hans-Peter Kohler and Susan Cotts Watkins (2001), How can we measure the causal effects of social networks using observational data? Evidence from the diffusion of family planning and AIDS worries in South Nyanza District, Kenya, MPIDR WP 2001-022

Berry, Sara (1993), No condition is permanent. *The social dynamics of agrarian change in Sub-Saharan Africa*, (Madison: The University of Wisconsin Press).

Breiger, Ronald (1990), Social control and social networks: a model from Georg Simmel in Craig Calhoun, Marshall W. Meyer and W. Richard Scott (eds), *Structures of power and constraint. Papers in honor of Peter M. Blau*, (Cambridge: Cambridge University Press), pp. 453-476.

Conley, Timothy and Christopher R. Udry (2002), Learning about a new technology: pineapple in Ghana, working paper, Yale University

Dercon, Stefan and Joachin deWeerdt (2002), Risk-sharing networks and insurance against illness, The Centre for the Study of African Economies Working Paper Series, WP 175.

deWeerdt, Joachin (forthcoming), Risk-sharing and endogenous networks, in Dercon, Stefan (ed) *Insurance against poverty*, (Oxford: Oxford University Press).

Durlauf, Steven N. (1999), The membership theory of inequality: Ideas and implications, pp. 161-177 in Elise S. Brezis and Peter Temin (eds), *Elites, Minorities and Economic Growth*, (Amsterdam: Elsevier Science).

Ensminger, Jean (1997), Transaction costs and Islam: explaining conversion in Africa, Journal of Institutional and Theoretical Economics, 153, 1, pp. 4-29.

Fafchamps, Marcel and Susan Lund (2003), Risk-sharing networks in rural Philippines, Journal of Development Economics, 71, 2, pp. 261-287.

Goldstein, Markus and Christopher Udry (1999), Agricultural innovation and resource management in Ghana, Final Report to IFPRI

Granovetter, Mark (1973), The strength of weak ties, American Journal of Sociology, 78, 1360-1380

Granovetter, Mark (1982), The strength of weak ties: a network theory revisited, pp. 105-130 in Peter Marsden and Nan Lin (eds), *Social structure and network analysis* (Thousand Oaks: Sage Publications).

Granovetter, Mark (1985), Economic action and social structure: the problem of embeddedness, American Journal of Sociology, 91, pp. 481-510

Haaland, Gunnar (1969), Economic determinants in ethnic processes, pp. 58-73 in Fredrik Barth (ed) *Ethnic groups and boundaries*, (Boston:Little, Brown and Company).

Jackson, Matthew O. (forthcoming), A survey of models of network formation. Stability and efficiency, Gabrielle Demange and Myrna Wooders (ed), *Group formation in economics: networks, clubs and coalitions* (Cambridge: Cambridge University Press).

Maddala, G. S. (1983), *Limited-dependent and qualitative variables in econometrics*, (Cambridge: Cambridge University Press).

Manski, Charles (1993), Identification of endogenous social effects: the reflection problem, Review of Economic Studies, 60, 204, pp. 531-42

Manski, Charles (2000), Economic analysis of social interactions, The Journal of Economic Perspectives, 14, 3, pp. 115-136

Sobania, Neil (1991), Feats, famines and friends: nineteenth century exchange and ethnicity in the Eastern Lake Turkana region, in John G. Galaty and Pierre Bonte (eds) *Herders, Warriors and Traders. Pastoralism in Africa*, (Boulder, Co: Westview Press), pp. 118-142

Wellman, Barry and S. D. Berkowitz (1988a), Preface in Barry Wellman and S. D. Berkowitz (eds), *Social Structures. A network approach*, (Greenwich: JAI Press), pp. xiii-xviii

Wellman, Barry and S. D. Berkowitz (1988b), Introduction: studying social structures in Barry Wellman and S. D. Berkowitz (eds), *Social Structures. A network approach* (Greenwich: JAI Press), pp. 1-14.

Udry, Christopher R. and Timothy Conley (forthcoming), Social networks in Ghana, in Christopher B. Barrett, editor, *The Social Economics of Poverty: Identities, Groups, Communities and Networks* (London: Routledge).

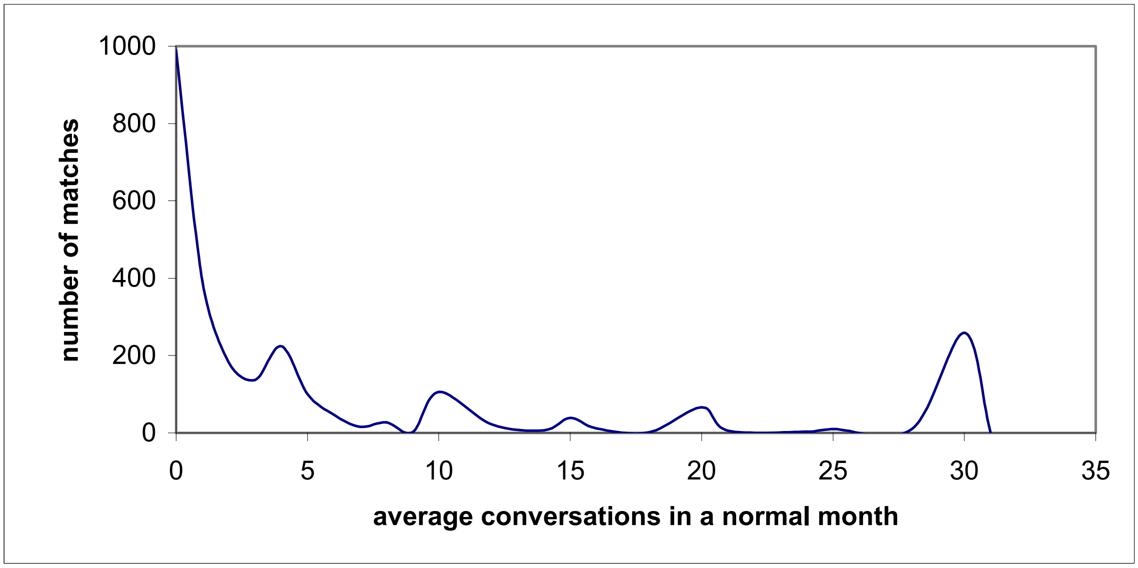

Figure 1 – frequency of the average number of conversations in a normal month

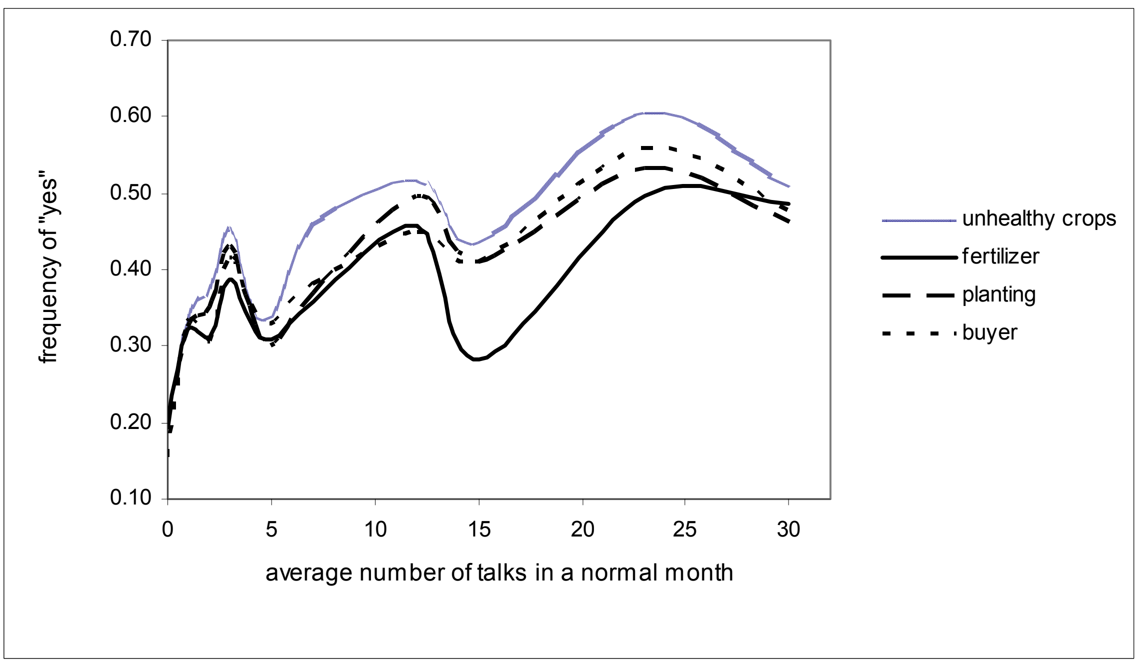

Figure 2 – frequency of “yes” as a function of average number of talks in a normal month

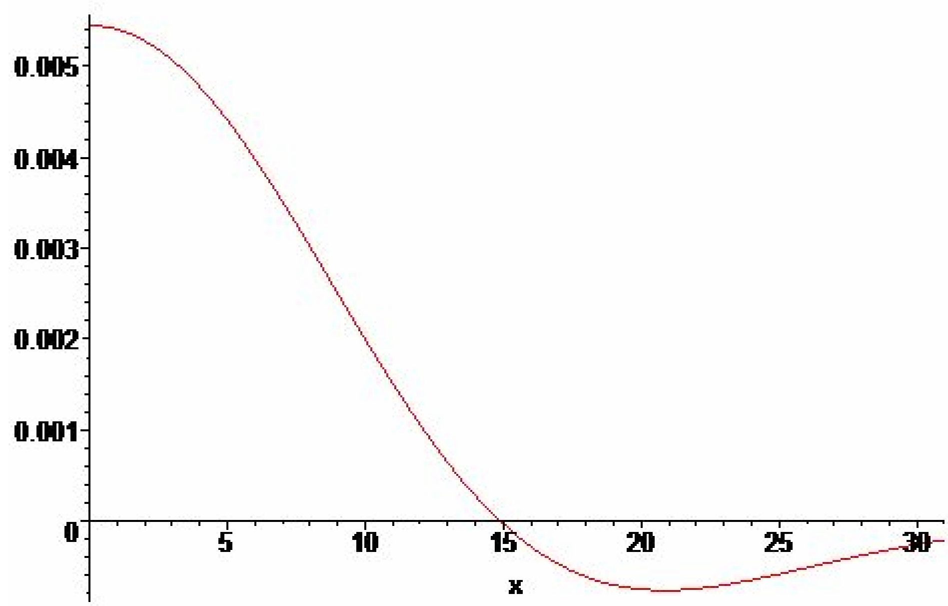

Figure 3: marginal effect of the number of talks per month on establishing a link in order to get information on fertilizer application.

Table 1: Prospective information links conditional on prior knowledge of match

	Can you go to ___							
	If you had a problem with unhealthy crops?		for advice about when to apply a new kind of fertilizer?		if you wanted to discuss changing your method of planting?		if you wanted to find a buyer for any of your crops?	
Know	no	yes	no	yes	no	yes	no	yes
no	312	1	312	1	312	1	313	0
yes	1755	959	1819	895	1834	880	1852	862

Table 2: Explanatory variables

Variable	Definition (i = respondent, j = randomly matched individual)	Mean (Std Deviation)
Same_hh	0-1 dummy variable that equals 1 if both i and j belong to the same household	.009 (.098)
Same_clan	0-1 dummy variable that equals 1 if both i and j belong to the same clan	.270 (444)
Both_migrant	0-1 dummy variable that equals 1 if both i and j are migrants	.001 (.033)
Migrant_nmigrant	0-1 dummy variable that equals 1 if only i is migrant	.044 (.204)
Nmigrant_nmigrant	0-1 dummy variable that equals 1 if both i and j are not migrant	.914 (.280)
Ff	0-1 dummy variable that equals 1 if both i and j are female	.260 (.438)
Fm	0-1 dummy variable that equals 1 if i is female and j is male	.292 (.455)
Mf	0-1 dummy variable that equals 1 if i is male and j is female	.207 (.405)
Older	Age difference between i and j if i is older than j, 0 otherwise	6.95 (10.68)
Younger	Absolute value of age difference between i and j if i is younger than j, 0 otherwise	7.23 (10.75)
Literacy	0-1 dummy variable that equals 1 if i is illiterate and j is literate	.167 (.373)
Wealthier	Difference in wealth between i and j (in 10^5 cedis) if i is wealthier than j, 0 otherwise	2.56 (11.76)
Poorer	Absolute value of the difference in wealth between i and j (in 10^5 cedis) if i is poorer than j, 0 otherwise	2.83 (9.03)
Maize_more	Difference in experience with maize (in years) between i and j if i has more experience than j, 0 otherwise	6.57 (10.38)
Maize_less	Absolute value of the difference in experience with maize (in years) between i and j if i has less experience than j, 0 otherwise	7.56 (10.81)
Cassava_more	The same as Maize_more, for cassava	6.60 (10.43)
Cassava_less	The same as Maize_less for cassava	7.61 (10.94)
Pineapple_more	The same as Maize_more, for pineapple	1.31 (2.95)
Pineapple_less	The same as Maize_less for pineapple	2.47 (4.37)
Cocoa_more	The same as Maize_more, for cocoa	2.93 (8.34)
Cocoa_less	The same as Maize_less for cocoa	2.68 (7.27)
Yam_more	The same as Maize_more, for yam	5.19 (10.26)
Yam_less	The same as Maize_less for yam	6.20 (12.07)
Wage_more	Difference in wage received between i and j if i has received more wage than j, 0 otherwise	28631.32 (115113.4)
Wage_less	Absolute value of the difference in wage received between i and j if i has received less wage than j, 0 otherwise	23208.75 (95119.840
Time_job_more	Difference in time spent on non-farm job between i and j if i spent more time than j, 0 otherwise	.088 (.242)
Time_job_less	Absolute value of the difference in time spent on non-farm job between iand j if i spent less time than j, 0 otherwise	.075 (.221)

Net_rev_more	Difference in net revenue from own business between i and j if i received a bigger net revenue than j, 0 otherwise	43600.24 (144527.4)
Net_rev_less	Absolute value of the difference in net revenue from own business between i and j if i received a smaller net revenue than j, 0 otherwise	55643.44 (154279.7)
Time_bus_more	Difference in time spent on own business between i and j if i spent more time than j, 0 otherwise	.186 (.417)
Time_bus_less	Absolute value of the difference in time spent on own business between i and j if i spent less time than j, 0 otherwise	.177 (.405)
Both_farmer	0-1 variable that equals 1 if both i and j identify themselves as farmers	.730 (.444)
Farm_nfarm	0-1 variable that equals 1 if only i identifies himself as farmer	.110 (.313)
Nfarm_farm	0-1 variable that equals 1 if only j identifies himself as farmer	.122 (.327)
Both_teacher	0-1 variable that equals 1 if both i and j identify themselves as teachers	.001 (.034)
Teacher_nteacher	0-1 variable that equals 1 if only i identifies himself as teacher	.021 (.145)
Nteacher_teacher	0-1 variable that equals 1 if only j identifies himself as teacher	.016 (.126)
Both_trader	0-1 variable that equals 1 if both i and j identify themselves as traders	.041 (198)
Trad_ntrad	0-1 variable that equals 1 if only i identifies himself as trader	.129 (.335)
Ntrad_trad	0-1 variable that equals 1 if only j identifies himself as trader	.133 (.339)
Male	0-1 variable that equals 1 if i is male	.447 (.497)
Age	Age of i	40.02 (13.44)
Wealth	Value of non-land assets owned by i (in 10^5 cedis)	3.29 (12.16)
Maize	Years of experience with maize of individual i	24.21 (14.04)
Cassava	Years of experience with cassava of individual i	24.14 (14.11)
Pineapple	Years of experience with pineapple of individual i	2.06 (3.53)
Cocoa	Years of experience with cocoa of individual i	3.37 (8.76)
Yam	Years of experience with yam of individual i	9.22 (12.95)
Wage	Value of wage received by i	30728 (117224)
Time_job	Time spent on job by i	.096 (.254)
Net_rev	Value of net revenue from own business received by i	35481 (153573)
Time_bus	Time spent on own business by i	.229 (.456)
Farmer	0-1 variable that equals 1 if i identifies himself as farmer	.839 (.367)
Teacher	0-1 variable that equals 1 if i identifies himself as teacher	.022 (.148)
Trader	0-1 variable that equals 1 if i identifies himself as trader	.167 (.373)

Village 1	0-1 variable that equals 1 if i lives in village 1	.27 (.44)
Village 2	0-1 variable that equals 1 if i lives in village 2	.20 (40)
Village 3	0-1 variable that equals 1 if i lives in village 3	.30 (.46)
Abs_male	0-1 variable that equals 1 if I and j have the same gender	.49 (.50)
Abs_age	Absolute value of the difference in age between i and j	14.13 (11.33)
Abs_wealth	Absolute value of the difference in wealth between i and j (in 10^5 cedis)	5.63 (15.39)
Abs_maize	Absolute value of the difference in years of experience with maize between i and j	14.78 (11.96)
Abs_pineapple	Absolute value of the difference in years of experience with pineapple between i and j	3.78 (4.58)
Abs_cocoa	Absolute value of the difference in years of experience with cocoa between i and j	5.43 (9.99)
Abs_yam	Absolute value of the difference in years of experience with yam between i and j	11.40 (13.78)
Abs_wage	Absolute value of the difference in earnings from non-farm job between i and j (in 10^5 cedis)	0.59 (1.60)
Abs_time_job	Absolute value of the difference in time spent on non-farm job between i and j	.17 (31)
Abs_net_rev	Absolute value of the difference in earnings from own business between i and j (in 10^5 cedis)	.35 (.52)
Abs_net_rev	Absolute value of the difference in time spent in own business between i and j	0.93 (1.94)

Table 3: Network interlinkages

	Number	Percentage of "yes" responses, per problem and per combination of answers			
		unhealthy crops	fertilizer	planting	buyer
don't ask for advice on any problem	1575				
ask for advice only for unhealthy crops	48	5	0	0	0
ask for advice only for fertilizer application	33	0	3.7	0	0
ask for advice only for changes in planting technique	1	0	0	0.1	0
ask for advice only for getting a buyer	111	0	0	0	12.9
ask for advice on all but getting a buyer	139	14.5	15.5	15.8	0
ask for advice on all problems	658	68.5	73.4	74.7	76.3
other combinations	149	12.0	7.4	9.4	10.8
total	2714	100	100	100	100

Table 4: Probit Estimation Results of Likelihood of Knowing the Random Match

A	Coef.	Robust Std. Err.	P>|z|	B	Coef.	Robust Std. Err.	P>|z|
same_clan	**0.244**	0.114	0.032	same_clan	**0.219**	0.108	0.043
both_migrant	-1.186	1.036	0.252	both_migrant	-1.039	0.876	0.235
migrant_nmigrant	**-0.570**	0.218	0.009	migrant_nmigrant	**-0.608**	0.202	0.003
nmigrant_migrant	-0.095	0.235	0.686	nmigrant_migrant	-0.108	0.228	0.637
ff	-0.135	0.159	0.396				
fm	**-0.438**	0.158	0.006	abs_male	**-0.285**	0.088	0.001
mf	-0.101	0.164	0.538				
older	**-0.020**	0.009	0.030	abs_age	-0.007	0.005	0.138
younger	0.001	0.007	0.832				
literacy	0.090	0.135	0.505	literacy	0.061	0.135	0.648
wealthier	**-0.130**	0.058	0.025	abs_wealth	**0.037**	0.012	0.002
poorer	**0.052**	0.021	0.014				
maize_more	-0.009	0.009	0.319	abs_maize	-0.006	0.005	0.203
maize_less	-0.001	0.007	0.930				
pineapple_more	-0.076	0.056	0.172	abs_pineapple	**0.033**	0.014	0.018
pineapple_less	**0.031**	0.017	0.062				
cocoa_more	-0.008	0.020	0.681	abs_cocoa	0.011	0.007	0.107
cocoa_less	0.011	0.008	0.193				
yam_more	0.007	0.010	0.457	abs_yam	**0.008**	0.005	0.073
yam_less	0.003	0.006	0.591				
wage_more	-1.060	0.754	0.159	abs_wage	0.084	0.056	0.131
wage_less	0.079	0.063	0.208				
time_job_more	1.291	1.557	0.407	abs_time_job	**-0.490**	0.211	0.020
time_job_less	-0.305	0.233	0.189				
time_bus_more	**-0.647**	0.392	0.099	abs_time_bus	0.046	0.117	0.691
time_bus_less	-0.039	0.123	0.753				
net_rev_more	**-0.145**	0.069	0.035	abs_net_rev	**0.077**	0.034	0.022
net_rev_less	**0.151**	0.042	0.000				
both_farmer	-0.081	0.256	0.753	both_farmer	-0.009	0.255	0.972
farm_nfarm	-0.394	0.262	0.133	farm_nfarm	-0.330	0.258	0.202
nfarm_farm	-0.259	0.253	0.304	nfarm_farm	-0.161	0.252	0.523
teacher_nteacher	0.227	0.252	0.368	teacher_nteacher	0.292	0.249	0.241
nteacher_teacher	**1.272**	0.328	0.000	nteacher_teacher	**1.114**	0.324	0.001
both_trader	-0.252	0.221	0.255	trad_ntrad	0.237	0.209	0.257
ntrad_trad	-0.203	0.145	0.164	ntrad_trad	**-0.324**	0.136	0.017
male				male	**0.294**	0.120	0.015
age	-0.006	0.009	0.454	age	**-0.017**	0.006	0.003
wealth	**0.124**	0.058	0.031	wealth	**-0.039**	0.012	0.001
maize	**0.061**	0.028	0.027	maize	**0.054**	0.026	0.038
cassava	-0.040	0.027	0.136	cassava	-0.037	0.026	0.144
pineapple	0.081	0.051	0.111	pineapple	-0.007	0.016	0.674
cocoa	0.004	0.020	0.860	cocoa	-0.015	0.009	0.112
yam	-0.013	0.010	0.185	yam	**-0.010**	0.006	0.075
wage	1.080	0.755	0.151	wage	-0.081	0.094	0.390
time_job	-1.422	1.555	0.360	time_job	0.420	0.373	0.260

net_rev	**0.209**	0.073	0.004	net_rev	0.049	0.065	0.447
time_bus	0.623	0.380	0.101	time_bus	-0.019	0.123	0.878
trader	-0.192	0.160	0.230	trader	**-0.434**	0.201	0.031
village1	**-0.428**	0.165	0.010	village1	**-0.468**	0.154	0.002
village2	-0.185	0.187	0.323	village2	**-0.296**	0.169	0.079
village3	**0.378**	0.212	0.074	village3	**0.397**	0.178	0.026
constant	**1.541**	0.439	0.000	constant	**1.926**	0.361	0.000
number of observations	2173				2173		
percent correctly predicted	91.1				90.9		
Log likelihood	-507.779				-535.582		
Pseudo R-squared	0.2288				0.1866		

In both specifications, the variables "Same_hh" and "Both_teacher" predict success perfectly when equal to 1; these variables were dropped and 25 observations were not used. Also, the variables "trad_ntrad", "farmer" and "teacher" were dropped due to collinearity. In the specification presented in column A, "male" is dropped due to collinearity.

Table 5: Probit Estimation Results of Likelihood of Establishing a Link

		unhealthy crops	fertilizer application	planting technique	getting a buyer
social distance	same_hh	**1.767**	**1.689**	**1.951**	**2.677**
		0.357	0.272	0.368	0.390
	same_clan	0.117	0.102	0.117	0.079
		0.077	0.078	0.078	0.083
	migrant_nmigrant	-0.575	**-0.710**	**-0.713**	**-0.956**
		0.391	0.402	0.425	0.485
	nmigrant_migrant	**-0.469**	**-0.471**	**-0.446**	-0.203
		0.180	0.197	0.193	0.182
	ff	**-0.315**	-0.199	-0.226	0.254
		0.159	0.157	0.163	0.162
	fm	0.132	**0.417**	**0.286**	0.142
		0.149	0.144	0.151	0.152
	mf	**-0.782**	**-0.936**	**-0.870**	-0.166
		0.136	0.187	0.163	0.128
	wealthier	**0.048**	**0.043**	**0.042**	-0.006
		0.022	0.022	0.022	0.020
	poorer	-0.004	-0.005	-0.004	-0.004
		0.003	0.003	0.003	0.003
differences in agricultural experience	maize_more	-0.008	0.000	-0.009	**-0.012**
		0.006	0.007	0.007	0.007
	maize_less	-0.005	-0.005	-0.004	-0.003
		0.006	0.006	0.006	0.006
	pineapple_more	**-0.049**	**-0.092**	**-0.054**	**-0.076**
		0.025	0.025	0.024	0.026
	pineapple_less	**0.014**	**0.030**	**0.015**	**0.020**
		0.008	0.009	0.008	0.008
differences in non-agricultural activities	wage_more	-0.549	-0.769	**-0.812**	-0.589
		0.482	0.483	0.489	0.597
	wage_less	-0.026	-0.066	-0.027	-0.026
		0.044	0.043	0.043	0.043
	time_job_more	**2.122**	**2.449**	**2.307**	**3.086**
		1.200	1.252	1.256	1.761
	time_job_less	-0.223	-0.212	-0.243	-0.156
		0.200	0.199	0.204	0.189
	time_bus_more	-0.220	-0.312	**-0.433**	0.019
		0.207	0.236	0.232	0.222
	time_bus_less	-0.059	-0.103	-0.099	-0.010
		0.095	0.108	0.100	0.093
	net_rev_more	0.060	0.089	0.050	0.029
		0.066	0.069	0.071	0.069
	net_rev_less	**0.068**	**0.059**	**0.054**	**0.067**
		0.025	0.026	0.025	0.025
differences in occupation	nteacher_teacher	**-0.875**	**-0.864**	**-0.838**	**0.921**
		0.345	0.347	0.351	0.265
	trad_ntrad	0.331	0.176	0.334	-0.339
		0.202	0.182	0.205	0.213

	ntrad_trad	**-0.219**	-0.191	-0.151	-0.095
		0.131	0.130	0.128	0.128
own attributes	age	-0.001	0.000	-0.004	-0.005
		0.009	0.009	0.009	0.009
	wealth	**-0.055**	**-0.049**	**-0.050**	-0.002
		0.022	0.022	0.022	0.020
	maize	0.015	0.010	**0.023**	**0.022**
		0.012	0.012	0.012	0.012
	yam	**-0.020**	-0.011	**-0.014**	-0.003
		0.008	0.007	0.007	0.008
	time_job	-1.709	**-2.122**	**-2.124**	-2.650
		1.220	1.254	1.287	1.720
	time_bus	0.312	0.309	**0.497**	0.140
		0.229	0.249	0.248	0.225
	trader	**-0.485**	-0.231	-0.410	0.156
		0.244	0.231	0.250	0.241
number of observations		1997	1997	1997	1997
percent correctly predicted		67.1	70.6	67.4	68.9
Log-likelihood value		-1095.8	-1002.3	-1049.8	-1042.7
Pseudo R-squared		0.17	0.23	0.19	0.19
Wald statistics					
Social distance:	W	86.79	115.58	92.09	68.13
	(n. df)	(10)	(10)	(10)	(10)
	[p-value]	[0.000]	[0.000]	[0.000]	[0.000]
Difference experience:	W	17.67	47.09	21.18	19.94
	(n. df)	(8)	(8)	(8)	(8)
	[p-value]	[0.024]	[0.000]	[0.067]	[0.011]
Diff. non-agric activities:	W	22.79	26.20	18.87	20.14
	(n. df)	(8)	(8)	(8)	(8)
	[p-value]	[0.004]	[0.001]	[0.016]	[0.010]
Difference occupation:	W	16.15	15.41	12.97	16.09
	(n. df)	(6)	(5)	(5)	(5)
	[p-value]	[0.013]	[0.009]	[0.024]	[0.007]
Own attributes:	W	20.18	8.87	16.77	9.67
	(n. df)	(11)	(12)	(12)	(12)
	[p-value]	[0.043]	[0.714]	[0.158]	[0.645]

Notes: The values in this table are the coefficients estimates of the probit model and (in smaller type) the respective standard deviations. Check table 2 for the definition of these variables.
The variables “both_migrant” and “both_teacher” predict failure perfectly, when equal to 1 - these variables were dropped and 5 observations were not used.
The variables “both_trader”, “teacher_nteacher” and “farmer” were dropped due to collinearity. For the same reason the variables that express differences in experience in cassava were dropped.
Variables included in the regression but not reported in this table are: constant, village dummies, older, younger, literacy, cocoa_more, cocoa_less, yam_more, yam_less, both_farmer, farm_nfarm, nfarm_farm, pineapple, cocoa, cassava, net_rev, wage and teacher.

Table 6: Comparing different measures of social distance

Difference variables expressed as:	Problem			
	Unhealthy crops	Fertilizer application	Planting technique	Getting a buyer
1) Euclidean norm	$W = 0.21 \sim X^2_8$ $prob > X^2 = 1.00$	$W = 0.81 \sim X^2_8$ $prob > X^2 = 0.99$	$W = 0.47 \sim X^2_8$ $prob > X^2 = 0.99$	$W = 0.22 \sim X^2_9$ $prob > X^2 = 1.00$
2) Double indicator	$W = 58.96 \sim X^2_{18}$ $prob > X^2 = 0.00$	$W = 75.36 \sim X^2_{17}$ $prob > X^2 = 0.00$	$W = 61.89 \sim X^2_{17}$ $prob > X^2 = 0.00$	$W = 6.05 \sim X^2_{18}$ $prob > X^2 = 0.99$
Conclusion	reject 1)	reject 1)	reject 1)	1) and 2) are equivalent

Table 7: testing for interlinkage between different networks

	Problem			
	Unhealthy crops	Fertilizer application	Planting technique	Getting a buyer
δ(eq. 10)	1.607 (1.554) [0.301]	1.683 (1.792) [0.348]	2.248 (1.482) [0.129]	1.934 (1.001) [0.053]
Wald statistic	$W = 41.61 \sim X^2_{36}$ $prob > X^2 = 0.24$	$W = 55.90 \sim X^2_{38}$ $prob > X^2 = 0.03$	$W = 35.85 \sim X^2_{37}$ $prob > X^2 = 0.52$	$W = 113.38 \sim X^2_{38}$ $prob > X^2 = 0.00$

Values in parenthesis are standard deviations; values in square backets are p-values.

Table 8 : Explaining “Number of talks in an average month”

	Coef.	Robust Std. Err.	P>|t|
same_hh	**18.261**	3.529	0.000
same_clan	0.222	0.533	0.677
both_migrant	-3.339	3.241	0.304
migrant_nmigrant	0.141	2.108	0.947
nmigrant_migrant	**-3.059**	0.820	0.000
ff	**-2.336**	0.916	0.011
fm	**-3.181**	0.918	0.001
mf	**-1.605**	0.704	0.023
older	-0.049	0.042	0.243
younger	-0.020	0.031	0.515
literacy	**-1.401**	0.744	0.061
wealthier	-0.122	0.144	0.399
poorer	**-0.046**	0.014	0.001
maize_more	0.001	0.041	0.982
maize_less	-0.004	0.037	0.919
pineapple_more	-0.247	0.159	0.122
pineapple_less	0.057	0.057	0.317
cocoa_more	**0.170**	0.063	0.007
cocoa_less	0.028	0.029	0.338
yam_more	-0.062	0.051	0.219
yam_less	0.030	0.027	0.261
time_job_more	11.077	8.936	0.216
time_job_less	2.186	1.409	0.122
time_bus_more	0.539	1.346	0.689
time_bus_less	**1.000**	0.593	0.093
both_farmer	0.072	1.149	0.950
farm_nfarm	-0.481	1.364	0.725
nfarm_farm	0.519	1.249	0.678
teacher_nteacher	**-20.583**	7.030	0.004
nteacher_teacher	**-3.304**	0.960	0.001
both_trader	0.049	1.215	0.968
ntrad_trad	-0.283	0.701	0.686
age	0.036	0.044	0.418
wealth	0.095	0.143	0.507
maize	**1.219**	0.080	0.000
pineapple	0.200	0.149	0.182
cassava	**-1.226**	-0.067	0.000
cocoa	**-0.176**	0.065	0.008
yam	0.060	0.050	0.231
teacher	**16.535**	6.734	0.015
trader	**1.542**	0.797	0.054

Number of obs = 1956 R-squared = 0.1305
F(52, 330) = 294.43 Prob > F = 0.0000
The variables “farmer”, “trad_ntrad” and “both_teacher” were dropped due to collinearity. Included in the regression but omitted from this table: “wage_more”, “wage_less”, “net_rev_more”, “net_rev_less”, “wage”, time_job”, “net_rev”,”time_bus”, village dummies and constant.

Table 9 : Probit estimates of the effect of strength of a tie on the Likelihood of establishing an information link

Parameter	Problem			
	Unhealthy crops	Fertilizer application	Planting technique	Getting a buyer
αa (eq. 15)	0.341 (0.114) [0.003]	0.385 (0.109) [0.000]	0.457 (0.132) [0.001]	0.581 (0.128) [0.000]
λ b (eq.15)	-0.0085 (0.0037) [0.025]	-0.0102 (0.0037) [0.005]	-0.0120 (0.0043) [0.006]	-0.0147 (0.0041) [0.000]
Number of talks maximizes the probability of success	20.93	14.96	21.47	14.86

Values in parenthesis are standard errors; values in square brackets are p-values