www.elsevier.com/locate/worlddev

World Development Vol. 94, pp. 375–392, 2017
0305-750X/© 2017 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.worlddev.2017.02.002

Characterizing social networks and their effects on income diversification in rural Kerala, India

JUDIT JOHNY, BRUNO WICHMANN and BRENT M. SWALLOW*
University of Alberta, Edmonton, Canada

Summary. — Income diversification continues to be a key strategy for poor rural households, including those that are progressively developing and those operating under increasing distress. The ability of a household to diversify has been shown to depend upon its demographic and economic characteristics and its physical and social context. This paper considers the effects of intra-village social networks on household income diversification in one of the poorest and most ethnically diverse areas of the Indian state of Kerala. Using techniques adapted from spatial econometrics, we find that social connections within a village magnify the impacts of household characteristics such as education and number of adults by a factor of 3.6 times. Models with alternative measures of network centrality (degree and eigenvector) indicate that the number of network connections that a household has is more important than the centrality of those connections. Finally, we use social contact information to calculate assortative mixing based on caste. The results suggest social stratification in these villages, with higher levels of stratification associated with lower levels of income diversification.
© 2017 Elsevier Ltd. All rights reserved.

Key words — social networks, income diversification, rural India, network effects, network centrality, social stratification

1. INTRODUCTION

Income diversification by individuals and households is ubiquitous in the rural economies of developing countries (Barrett, Reardon, & Webb, 2001; Ellis, 2000; Martin & Lorenzen, 2016). Why? If markets were complete and costless, income diversification can be interpreted as the opposite of specialization: failure to deepen investment in specific enterprises and thus benefit from economies of scale (Ellis, 2000). With restricted and costly exchange, however, income diversification can be interpreted as the net result of trading off advantages of specialization in specific enterprises with complementarities and economies of scope between enterprises. For example, with thinly-traded labor markets, diversification can be a rational response to seasonality of labor requirements in agricultural enterprises that entail periods of labor surplus and shortage (Ellis, 2000). Or with incomplete credit or insurance markets, diversification can be an appropriate ex ante or ex post approach for managing weather or price risks (Barrett *et al.*, 2001; Ellis, 2000). Income diversification can involve different forms of production agriculture (e.g. cereals, perennials, livestock, horticulture), participation in both production and value addition (e.g. sales, processing), or engagement in both on-farm and off-farm employment (e.g. casual labor, formal sector employment).

Income diversification is thus one of the few strategies available to farm families living in situations of constrained markets, a situation that is prevalent in developing country economies (Stiglitz, 1989). The proximate determinants of diversification vary from case to case, although the literature shows empirical regularities (Barrett *et al.*, 2001; De Janvry & Sadoulet, 2001; Himanshu, Lanjouw, Murgai, & Stern, 2013). For a household on an upward trajectory from poverty to increased prosperity, households diversify income sources in order to absorb seasonal labor shortages, exploit economies of scope between enterprises, or leverage limited financial capital. For a household on a downward trajectory toward worsening poverty, diversification can be a constrained response to expected future income shocks. The motives associated with these upward and downward trends are alternatively described as "opportunity-led or survival-led" (Alobo Loison, 2015), "necessity or choice" (Ellis, 2000), "pull or push" (Barrett *et al.*, 2001), "progressive success or distress" (Martin & Lorenzen, 2016), or "asset-based or insurance-based" (Anderson & Deshingkar, 2005). In many circumstances, households may respond to both types of motivations, especially where a single event, such as a prolonged illness of a breadwinner, can make the difference between moving into or out of poverty (Krishna, 2010).

Regardless of the motives in specific contexts, however, the evidence indicates that most poor rural households benefit from opportunities to diversify farm and non-farm income sources. From a review of evidence from 11 Latin American countries, Reardon, Berdegue, and Escobar (2001) found diversification of rural incomes to be highest in countries with lowest average income, but, controlling for country and region, highest among households with highest average incomes. Barrett *et al.* (2001) review results from Ethiopia, Tanzania, Côte d'Ivoire, and Kenya that show income diversification to be positively associated with household welfare measures. Using data from a nationally representative survey of India, Birthal, Roy, and Negi (2015) find that smallholder households that diversify toward high-value crops have higher per capita household expenditures than household that diver-

* This project was funded by the International Development Research Centre (Canada) and the Canadian Department of Foreign Affairs, Trade and Development (now known as Global Affairs Canada) through the Canadian International Food Security Research Fund (CIFSRF). It was conducted within a project "Alleviating Poverty and Malnutrition in Agrobiodiversity Hotspots" implemented by the M.S. Swaminathan Research Foundation in Chennai, India (MSSRF) and the University of Alberta in Edmonton, Canada. The authors acknowledge the contributions of Arun Raj, Lisha Reni, LathaAravind, K.M. Johnson, and Nisha for their support of the social network survey and the whole MSSRF/U of A team involved in larger household survey. Final revision accepted: February 8, 2017.

sify less. Of course, it can be difficult to determine causality in such studies: do households diversity more because they have higher income or do they have higher income because they diversify more?

Several household characteristics have been commonly found to influence the extent of income diversification. This includes: (1) positive or negative association with the age of the household head (Agyeman, Asuming-Brempong, & Onumah, 2014; Khatun & Roy, 2012); (2) positive association with amount or value of assets (Agyeman *et al.*, 2014; Ellis, 2000; Khatun & Roy, 2012); (3) positive association with diversity of assets (Martin & Lorenzen, 2016), (4) positive association with availability of household labor (Agyeman *et al.*, 2014; Ellis, 2000; Liu & Lan, 2015), (5) positive association with level of education (Agyeman *et al.*, 2014; Barrett *et al.*, 2001; Khatun & Roy, 2012; Liu & Lan, 2015), and (6) positive association with the prevalence of market or production risks (Alobo Loison, 2015).

Physical context can affect opportunities for diversification. For example, households located near urban centers, mines, or plantations tend to have greater opportunities to earn income in those sectors, while households located near towns, highways, or market centers may have greater opportunities to market raw or processed food (e.g., Agyeman *et al.*, 2014; Joshi, Gulati, Birthal, & Tewari, 2004). Households located in remote forested areas are more likely to rely on the consumption and sale of products gathered from the forest than households located further away (e.g., Belcher, Achdiawan, & Dewi, 2015).

Social context may be an equally important determinant of diversification. In India, for example, caste and ethnicity provide the basis for multi-functional social networks. Rural families use caste-based networks to find marriage partners for girls in other rural areas, for men to identify opportunities for temporary migration, and for families to reduce income risk through gifts and loans (Munshi & Rosenweig, 2016). These gifts and loans function as substitutes for formal insurance and state-sponsored safety nets. Munshi and Rosenweig (2016) propose that rural insurance networks underlie the persistence of large wage gaps between urban and rural India.

In this paper we evaluate the role of intra-village social networks in enabling household income diversification in rural India. Understanding intra-village social networks and their effects can help with the design of rural service programs and the assessment of such interventions. Social network analysis shows who will be included or excluded if advisory services are supplied through focal point, interest group, or affirmative action approaches that target social marginalized groups (Glendenning, Babu, & Asenso-Okyere, 2010). The magnitude of the network effect on diversification also has implications for resource allocation—a larger effect of network on diversification implies that a larger share of resources should be devoted to strengthening those networks. Network approaches can also be used to measure ripple effects as the outcome of a policy intervention spreads through socially connected individuals (a social multiplier effect). While previous studies have examined the role of social networks on economic outcomes such as agricultural technology adoption (Maertens & Barrett, 2013; Matuschke & Qaim, 2009), risk sharing (Fafchamps & Lund, 2003; Munshi & Rosenweig, 2016), labor markets (Calvo-Armengol & Jackson, 2004), and diffusion of micro-finance (Banerjee, Chandrasekhar, Duflo, & Jackson, 2013), only a few previous studies have explored connections between social networks and diversification. No other study has applied the same network analysis and econometric methods to the study of the effects of social networks on diversification.

Cinner and Bodin (2010) use a network analysis approach to map occupations and the connections between those occupations in 27 coastal communities in 5 western Indian Ocean countries (Kenya, Tanzania, Madagascar, Seychelles, and Mauritius). They relate the position of each occupation (e.g. measures of centrality) in the "livelihood landscape" to indicators of socio-economic development and network statistics (e.g. density) to community-level development and population density. Their findings suggest a positive association between specialization and development at the household level, but no particular association at the community level. Baird and Gray (2014) consider income diversification and social networks of exchange as alternative mechanisms that pastoral households use to manage risk and uncertainty in Northern Tanzania, finding that income diversification and inter-household exchange serve as substitutes.

The context for the current study is a small contiguous region within the Western Ghats region of the Indian state of Kerala. The population of the area is comprised of a mixture of ethnic groups and castes, and households engage in a range of livelihood activities. We represent intra-village household-to-household networks through an adjacency matrix derived from household interview data we collected on several dimensions of social contact. These data also allow us to construct standard measures of network centrality of households, as well as village-level measures of social stratification based on social contacts within and between castes and tribes. The data also allow us to use methods adapted from spatial econometrics to estimate network multipliers, i.e. the multiplicative effects of social networks on the determinants of income diversification. To avoid mischaracterization of network statistics, we conducted interviews with all households in each of nine villages, which allowed us to examine complete networks. This is important as sample-based statistics may misrepresent their population counterparts (Costenbader & Valente, 2003; Lee, Kim, & Jeong, 2006).

While Baird and Gray (2014) focus on exchange networks per se, we took an inclusive approach to social networks very similar to the approach that Banerjee *et al.* (2013) took in their study of the effects of social networks on the diffusion of micro-finance in Karnataka, India. This approach is most consistent with a concept of social network as a vehicle for exchanging resources and information. We conclude that intra-village social networks play very important roles in enabling diversification in our study context. Our results indicate the existence of: (i) a social multiplier effect with respect to income diversification, i.e., network diversification is positively associated with own diversification; (ii) a social position effect on diversification, i.e., household centrality is positively associated with income diversification; and (iii) social stratification, i.e., social connections reveal assortative mixing (connections within caste are more prominent than between caste).

This paper proceeds as follows. Section 2 provides a brief description of the study site and sampling, while Section 3 provides a description of the data and the methods used to model social networks, income diversification, and their interrelationship. Section 4 provides results and Section 5 offers a discussion and conclusion.

2. STUDY SITE AND DATA

(a) *Study location*

We conducted this study in nine villages of Meenangadi Panchayat (decentralized territorial unit) in Wayanad District of Kerala, southern India. The primary sources of income in the region are plantation crops including coffee, tea, cocoa, pepper, and rubber. Rice, banana, tubers, and fruits are the most common food crops. In recent years, there has been an increase in non-farm activities and during fieldwork we observed that people were involved with a variety of non-farm activities including non-agricultural labor, government service, shop keeping, small-scale retail ventures, transportation services (operation of rickshaws, trucks, and tractors), as well as employment offered by the Mahamta Ghandi National Rural Employment Guarantee Scheme (Ravi & Engler, 2015).

The area has remarkably high ethnic heterogeneity. As noted by Munshi and Rosenweig (2016), Indian society is highly stratified by the caste system. Mohindra, Haddad, and Narayana (2006, p. 1010) describe caste as a "hereditary, endogamous, usually localized group, having a traditional association with an occupation and a particular position in the hierarchy of castes". The castes are categorized into four main groups namely Scheduled Tribes (ST), Scheduled Castes (SC), Other Backward Classes (OBC), and General caste (General). The Scheduled Tribes are indigenous people of India, also known as Adivasis, who continue to face systematic discrimination in Indian society. One of the most distinctive features of Wayanad District is the magnitude and diversity of the tribal population. There are about 35 tribal communities in Kerala and Wayanad District has the highest concentration of tribal populations in Kerala (Census of India, 2011). The five main Scheduled Tribes in the study area are Kurichya, Kuruma, Paniya, Adiya, and Kattunaika. On the basis of systematic differences in health, education, and employment status, Rajasenan, Abraham, and Rajeev (2013) distinguish the Kurichya and Kuruma as "forward tribes" (relatively high levels of socio-economic development) and the Paniya, Adiya, and Kattunaika as "backward tribes" (relatively low levels of socio-economic development).

In the past, tribal communities did not have permanent houses and tended to live in remote forests. However, economic change and targeted programs by federal and state governments have enabled most tribal communities to build permanent houses and own small parcels of land. In Wayanad, most tribal communities settle in groups or clusters with houses located close together. These clusters are commonly referred to as tribal settlements or colonies. A typical village has one or more tribal settlement. Extended families tend to live in the same settlement.

(b) *Sampling and data collection*

Data required for this study was collected from a census of all 301 households residing in nine villages in Meenangadi Panchayat. The unit of analysis is a household. Respondents were selected from a census of approximately 1000 households living in 31 villages that were previously included in a related study, "Alleviating Poverty and Malnutrition in Agrobiodiversity Hotspots." The 31 villages are located within a contiguous area of Meenangadi Panchayat selected to test the efficacy of a number of agricultural technologies and enterprises for improving household nutrition and income (Ragu *et al.*, 2013). The villages are located relatively near to each other, in an area of rolling hills and forest remnants in the Western Ghats region of southwest India. Out of the 31 villages, all small villages (<20 households) and large villages (>100 households) were excluded from consideration. Nine intermediate-sized villages (between 20 and 100 households) were thus selected. This focus on intermediate-sized villages reduced survey costs, helped ensure a reasonable sample of villages, and allowed us to depict the complete picture of social networks within each village. A sampling design that surveys all households in the village is important because network measures (e.g., centrality) based on a portion of the network may distort complex relationships of the true (or complete) network structure (Costenbader & Valente, 2003; Lee *et al.*, 2006). As a result, to avoid measurement error bias, all households in each of the nine villages were included in the household survey.

Some of our data were obtained from a cross-sectional survey implemented in 2013 by the M.S. Swaminathan Research Foundation (India) and the University of Alberta (Canada) as part of a larger research project (Ragu *et al.*, 2013). In order to capture income diversification, household incomes were categorized into different sources, identified through a pilot survey, namely: sales of surplus staple crops, sales of fruits and vegetables, sales of crop by-products, sales of livestock products, oxen rental, agricultural wages, income from off-farm activities, migration income, and salaried employment. Annual income in rupees earned from each of these sources was obtained from recall questions posed to the household respondents. That survey also collected information on household demographics such as age, household size, education of the household head, religion, and ethnic background. The survey also collected information on household assets and infrastructure as well as incidences of household shocks like loss of crops and assets.

To measure the social networks of each household, a questionnaire was drafted based on the social network survey developed by Banerjee *et al.* (2013) for their study of diffusion of microfinance in villages of rural Karnataka, India. The site of the Banerjee *et al.* (2013) study is less than 200 km from our Meenangadi field site. We adapted the Banerjee *et al.* (2013) survey design to capture within-village inter-household social interactions. Our social network survey included questions on thirteen possible dimensions of interactions through which households within a village could be connected. Local field technicians who implemented the earlier baseline survey were also involved in implementing the social network survey between November 2013 and January 2014. Specifically, the survey collected from each household answers to the following questions:

1. Which members of your family do you maintain good relations with?
2. Who comes to your house to watch television?
3. Who would you ask for help if you had a medical emergency?
4. If you needed to borrow Kerosene, rice, wheat, sugar or some other necessary good, whom would you go to?
5. With whom would you go to visit places of worship?
6. If you suddenly needed to borrow Rs. 100 for a day, whom would you ask for money?
7. If you had to make a difficult personal decision, whom would you ask for advice?
8. Who do you go to for advice?
9. Who are the people that you consider to be influential and respected by everyone?
10. With whom do you share foods produced in your home garden?
11. To whom do you sell agricultural food products?
12. To whom do you sell other products?
13. Whom do you ask for information about government programs/policies?

The survey was administered in the main local language (Malayalam) to a member of the household older than 18 years of age. Enumerators translated the questionnaire into other local languages for respondents not comfortable speaking Malayalam (eg. Paniya, Adiya, Kattunayakan, Kurichiya, Kuruma). In most cases (91% of the observations) the head of the household answered the questions and in some cases the spouse of the head answered, or both answered jointly. For each question, respondents were asked to list residents of multiple households if appropriate. The responses were used to construct household-level networks. For example, if the respondent of household 1 declared that in case of a medical emergency he/she would ask for help from residents of households 2 and 4; then we interpret that household 1 is connected to both 2 and 4 in a network for medical emergency support.

We investigate the possibility of gender differences in our network elicitation instrument as men and women may have very different responses to the questions above. We statistically test for differences in mean network centralities (an important network characteristic, see 3(d)) between male- and female-headed households. The t-test results indicate that we cannot reject the null hypothesis of equality of mean centrality between male- and female-headed households. This suggests that gender is not influencing network structure in our sample.

3. METHODS

(a) *Measuring income diversification*

Most empirical studies in the literature measure income diversification as an index such as the Simpson, Herfindahl, or Shannon. These indices account for both the number of income sources and the balance among them (Ersado, 2006; Joshi *et al.*, 2004; Minot, Epprecht, Tran, Tram, & Le Quang, 2006). In this paper we use the Simpson Index of diversity which is defined as

$$Simpson_i = 1 - \sum_{s=1}^{k} P_{i,s}^2,$$

where i is a particular household, s is a particular income source, and $P_{i,s}$ is the proportion of household i's income from source s out of k possible sources. The value of the Simpson index falls between zero and $1 - 1/k$ (Minot *et al.*, 2006), which equals 0 if there is just one source of income ($k = 1$, $P_{i,k} = 1$), indicating no diversification. As the number of income sources increases, the shares P_s decline and so does the sum of the squared shares. As a result, *Simpson* approaches an upper limit of $1 - 1/k$. Thus a higher Simpson index implies greater diversification.

(b) *Determinants of social networks*

As discussed in the introduction, we draw a distinction between household characteristics and contextual factors as determinants of diversification. We thus collected data on variables that have been found to be important determinants of both push and pull diversification in previous studies, specifically different types of assets, exposure to risks, and social structure.

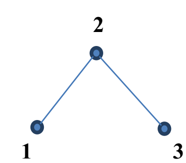
(c) *Representing social networks*

In general a social network is a set of nodes joined by some relation (e.g., households connected for medical emergency support). Nodes can be persons, groups, organizations, or entities such as texts, artifacts, and even concepts (Butts, 2008). This study represents each of the 9 villages as a network of households linked through social ties. The medium-sized villages (20–100 households) included in this study are small enough for all households to know each other but large enough to capture many forms of social interactions. Thus a household's social network consists of all households within their home village that it interacts with one or more of the 13 dimensions.

In empirical contexts, network data are often represented by an adjacency square matrix whose ijth cell is equal to 1 if node i has a link to node j, and 0 otherwise (Butts, 2008; Jackson, 2008). By convention, the diagonal of the matrix is 0 indicating that a node is not connected to itself. Below is an example of an adjacent matrix $\boldsymbol{W}$ representing a network with 3 nodes.

$$\boldsymbol{W} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

In the example, node 1 is connected with 2, node 2 is connected with both 1 and 3, and node 3 is connected with 2. Graphically, these relationships can be illustrated by the following network graph.



We used the answers to the social network questions to construct an adjacent matrix representing the social network for each of the nine villages in our study. We constructed undirected networks, i.e. an i–j connection implies a j–i connection such that the adjacent matrix is symmetric. Therefore, a link between households i and j exists if: (i) household i mentions a member in household j as a contact in response to one or more of the network questions; and/or (ii) a member of household j mentions a member in household i as a contact. For example, household i may report borrowing money or material goods from j although j did not borrow from i. We assume that the one-way borrowing link is sufficient to create a link between the two households through which information, services, or skills could flow in either direction.

The network data collected enabled us to construct 13 matrices that take account of the connections between households along each of the 13 dimensions included in the questionnaire. For this paper, we constructed a representative social network matrix where two households are considered to be linked if they have a relationship in any of the thirteen dimensions. The representative social network is thus the union of the 13 matrices. This approach offers several advantages to our study of income diversification. First, as discussed by Banerjee *et al.* (2013), this is an appropriate network measure since the emphasis here is on the links between two households and any of the dimensions included in the survey creates an opportunity for interaction and information exchange. Second, collapsing the 13 different dimensions of social interaction in a single unweighted network avoids giving (possibly ad hoc) weights to different connections as it is difficult (if not impossible) to predict the relevance of each dimension for income diversification. Third, the union of all networks avoids the occurrence of isolated households, which would lead to difficulties in the identification of the econometric model of average network effect, and in the calculation of eigenvector centrality of isolated nodes.

In this paper we use this social network data to estimate an econometric model that captures average network effects on diversification. In other words, we are interested in estimating how the diversification of a household's social network affects the household's own diversification. A discussion of this approach is presented in Section 3(e) and (i).

We also use the social network data to examine caste-based social stratification. Specifically, our data allow us to measure network homophily, i.e. the propensity of social contact between similar people to occur at a higher rate than among dissimilar people. The *assortativity coefficient* developed by Newman (2003) is typically used to quantify the level of homophily in a network. Formally, Newman's assortativity coefficient is defined as

$$r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$$

where e_{ij} is the share of links between nodes of type i and j, $a_i = \sum_j e_{ij}$, and $b_j = \sum_i e_{ij}$. We construct the coefficient to consider assortative mixing in networks according to the caste or ethnicity of each household. In general, $-1 < r < 1$. Positive values of r indicate assortative mixing (or homophily) while negative values indicate disassortative mixing. A value of zero reflects neutrally mixed networks.

(d) *Measuring social network centrality*

This paper uses indices of social network centrality as a second method to characterize social interactions between households within a village. Centrality indices measure the influence or position of nodes within in a network. We use centrality measures to examine whether social position of a household influences their diversification profile.

Two commonly used centrality measures are degree centrality and eigenvector centrality. Degree centrality indicates how well a node is connected in terms of direct connections with other nodes (Butts, 2008). In the case of an undirected network, the *degree* of a node is the number of direct connections the node has with other nodes (Jackson, 2008). *Degree centrality* is the proportion of connections of a node, i.e. the degree divided by (*n-1*), where *n* is the number of nodes in the network (hence, *n-1* is the maximum number of connections a node may have). Formally, degree centrality is measured as:

$$C_d(i, A) = N(i)/(n - 1)$$

where, $C_d(i, \boldsymbol{W})$ represents the degree centrality C_d of node i in network $\boldsymbol{W}$, N is the number of connections attached to node i, and n is the number of nodes in the network. As degree centrality is the proportion of connections of a particular node, it ranges from 0 to 1 and is one of the simplest measures of the social position (or influence) of that node. For example, in a network of size 10 (10 nodes), a node with a degree of 9 is directly connected to all other nodes and thus is very central to the network (Butts, 2008). A node with high degree centrality is expected to have larger influence on those around it and possibly the entire network.

While degree captures centrality by examining the number of connections of a node (a quantity measure), eigenvector centrality is based on the idea that a node's importance is determined by the importance of the nodes it connects with. This measure qualifies each connection by taking into account a node's proximity to other important nodes (Jackson, 2008). Let $C_e(i, \boldsymbol{W})$ denote the eigenvector centrality of node i in network W. The centrality of node i is proportional to the sum of the centrality of the nodes that it connects with (i.e. its neighbors):

$$\lambda C_e(i, W) = \sum_{j=1}^{n} w_{ij} C_e(j, \boldsymbol{W}),$$

where w_{ij} is the *ij*th element of the adjacent network matrix W, and λ is a scalar.

In matrix notation, eigenvector centrality can be represented as $\lambda C_e = \boldsymbol{W} C_e$, where C_e is a vector that collects the eigenvector centrality of all nodes in $\boldsymbol{W}$. Note that C_e is an eigenvector of the matrix $\boldsymbol{W}$, and λ is its corresponding eigenvalue (Jackson, 2008).[1] Our eigenvector measure also ranges from 0 to 1. As with degree centrality, a higher value of eigenvector centrality indicates a higher importance or centrality. In summary, the two centrality measures capture different dimensions of node importance (Borgatti, 2005).

(e) *Empirical models*

In this paper we use two econometric models to examine the influence of social networks on income diversification. The first model estimates the effect of household i's social network on i's own income diversification. The second model estimates the effect of household i's network centrality on their income diversification. Other possible determinants of diversification are included as control variables.

(i) *Average network effect model*

To build an econometric model we first construct a matrix A that is a row normalization of the adjacent matrix $\boldsymbol{W}$ that represents the network. Formally, the elements of A are $a_{ij} = w_{ij} / \sum_j w_{ij}$, where w_{ij} are elements of $\boldsymbol{W}$. Therefore, a_{ij} can be interpreted as the fraction of network weight that household i attributes to j. In spatial econometrics, A is often referred to as the spatial or weighing matrix.

It is useful to present our first empirical model in matrix notation:

$$\mathbf{Y} = \beta_0 + \beta_1 \mathbf{AY} + \beta_2 \mathbf{X} + \varepsilon \tag{1}$$

where $\mathbf{Y}$ is a vector of income diversification (measured by the Simpson index), $\mathbf{A}$ is the row normalized social network matrix, $\mathbf{X}$ is a matrix of other determinants of diversification (e.g. demographics, shocks and assets), ε is an independently and identically distributed (iid) mean zero error term ($E(\varepsilon|\boldsymbol{X}) = 0$), and βs are parameters to be estimated.

Notice that the vector $\mathbf{AY}$ captures the average diversification of a household's social network. The ith row of $\mathbf{AY}$ is equal to the average of the Simpson indices of all households connected to household i. Including $\mathbf{AY}$ in an econometric model of income diversification allows us to examine how a household's diversification is influenced by the average diversification of its network. Therefore, the estimated coefficient β_1 captures the *average network effect* on diversification.

The matrix $\mathbf{X}$ contains data on other determinants of income diversification that have been widely studied in the literature, including variables representing household characteristics, access to infrastructure, assets, risk exposure, and social structure as described in the introduction. Household characteristics include age of the household head, gender of the household head, education level the household head, caste or social category of the household, and number of adult members in household. Access to infrastructure is represented by use of electricity and use of cooking gas. Household assets include land, cattle, and hoes owned by the household. Household risk exposure is captured through dummy variables that indicate whether or not the household experienced a loss of assets or a crop failure within one year prior to the interview.

Our first empirical model is related to the Spatial Autoregressive model, with the spatial matrix replaced by a network matrix. Conceptually, our model is consistent with Manski's theory of social interactions which suggests that the outcome of each individual depends on her own characteristics and the mean outcome of her reference group (Manski, 1993). Bramoulle, Djebbari, and Fortin (2009) applies this theory to develop an econometric model for identifying peer effects through social networks. The model used in this study closely follows Bramoulle *et al.* (2009).

Our goal is to test whether or not a household's diversification behavior is influenced by the average diversification of its network (i.e. $\beta_1 \neq 0$) and how (i.e. is β_1 large or small, positive or negative?). The presence of a negative social network effect (i.e. $\beta_1 < 0$) suggests a competitive environment in which the strategies adopted by household *i*'s network to increase diversification lead to a decrease in *i*'s own diversification. The presence of a positive network effect (i.e. $\beta_1 > 0$) suggests a social multiplier where aggregate effects amplify individual effects (Glaeser, Sacerdote, & Scheinkman, 2003). This amplification is known as a social multiplier effect. Glaeser *et al.* (2003) has explained this using an example of criminal behavior. Crime deterrence is expected to be affected by changes in policing or punishment. However, if one person's inclination toward crime influences his neighbor's criminal behavior, then a change in policing will have both a direct effect on crime and an indirect effect through social influence. This presence of positive spillovers or strategic complementarities creates a "social multiplier" where aggregate coefficients will be greater than individual coefficients (Glaeser *et al.*, 2003).

Mathematically, this multiplier effect can be demonstrated through the reduced form representation of the structural model (1):

$$\mathbf{Y} = (\mathbf{I} - \beta_1 \mathbf{A})^{-1}(\beta_0 + \beta_2 \mathbf{X} + \varepsilon) \tag{2}$$

where $\mathbf{I}$ is the identity matrix. Notice that β_2 captures the direct effect of X on Y. The term $(\mathbf{I} - \beta_1 \mathbf{A})^{-1}$ represents the social multiplier as it multiplies the effect of β_2. The average social multiplier η is normally approximated as $\eta \approx 1/(1-\beta_1)$ and is interpreted as how much on average network interactions intensify the effect of variation in $\mathbf{X}$ on the outcome $\mathbf{Y}$ (Wichmann, 2014). In our application, η captures the ripple effect of variation in demographics, assets, and shocks on income diversification. For example, providing agricultural training opportunities and initial support may lead to a lower incidence of crop shocks and would ultimately affect income diversification. Such a training program could have a ripple effect through the village network (due, for instance, to information transmission) so that the behavior of a single household may modify the information available to the rest of the agents in its network (Maurin & Moschion, 2009).

Estimation of the parameters of model (3) can be challenging because of endogeneity due to reverse causality. This is evident when the model is written in matrix notation as it highlights the fact that Y is on both sides of the structural equation. Manski (1993, 2000) refers to this issue as the reflection problem. In our paper, it arises from the fact that household *i*'s diversification affects its network diversification which, in turn, affects *i*'s own diversification. Hence, the variable that captures the average diversification of the social network ($\mathbf{AY}$) is an endogenous variable. As a result, ordinary least squares (OLS) estimation of equation (2) provides biased and inconsistent estimators of βs.

We follow the spatial econometrics literature to implement an instrumental variable strategy and consistently estimate the parameters of model (3). This literature shows how instrumental variables can be constructed using network transformations of the variables in $\mathbf{X}$ (Bramoulle *et al.*, 2009; Kelejian & Prucha, 1998; Lee, 2003). This literature demonstrates that pre-multiplying $\mathbf{X}$ with $\mathbf{A}$, $\mathbf{A}^2$, and $\mathbf{A}^3$ leads to variables that have intuitive interpretations and are valid instruments for $\mathbf{AY}$. The variable $\mathbf{AX}$ represents the average diversification determinants of a household's network. The variable $\mathbf{A^2X}^2$ captures the average determinants of the network of a household's network. The corresponding third-degree relationship is therefore captured by $\mathbf{A^3X}$. This instrumental variable strategy delivers consistent parameter estimates under the assumption of strict exogeneity of the regressors $\mathbf{X}$, $(E(\varepsilon|X) = 0)$, and can be easily implemented using the generalized method of moments or GMM/IV approach (Bramoulle *et al.*, 2009; Kelejian & Prucha, 1998; Lee, 2003). Moreover, as Bramoulle *et al.* (2009) discuss, this approach is robust when networks have a fixed or exogenous structure. As a result, an underlying assumption of our estimator is that income diversification does not play a significant role in forming our matrix A. Note that our approach to construct the social network A captures 13 different dimensions of social interaction, each very different from the others. For example, the construction of A considers kinship ties, medical emergency contacts, sources of general advice, and networks of recreational social contacts like watching television. This social heterogeneity allows us to apply the GMM/IV strategy with a reasonable degree of confidence.

(ii) *Network centrality model*

We estimate a second empirical model to test whether or not network centrality influences income diversification. This approach modifies model (1) by replacing the average behavior of the network $\mathbf{AY}$ with a centrality measure $\mathbf{C}$. Equation 4 below represents this model using matrix notation:

$$\mathbf{Y} = \alpha_0 + \alpha_1 \mathbf{C} + \alpha_2 \mathbf{X} + \varepsilon \tag{3}$$

where $\mathbf{C}$ is a vector of centrality measures and α s are the parameters to be estimated. We estimate two specifications, one in which $\mathbf{C}$ captures degree centrality and another in which $\mathbf{C}$ captures eigenvector centrality.

The parameter of interest is α_1, the effect of network centrality on income diversification. The centrality indices allow us to identify households in positions of prominence, e.g. households whose positions enable actions such as information dissemination. Further, examining the effect of centrality of households on income diversification can guide the design and implementation of initiatives that are developed to facilitate diversification in rural areas. For example, if household centrality positively affects diversification ($\alpha_1 > 0$), then policy instruments (such as providing information regarding new diversification opportunities, offering support and training that might assist people in starting and maintaining new income-generating activities) should target central actors within a network. This approach can enhance the effectiveness of new initiatives. For example, Banerjee *et al.* (2013) show that participation in microfinance is significantly higher when the injection points, i.e. households who first adopted the microfinance program, have higher eigenvector centrality.

4. RESULTS

(a) *Income diversification*

Table 1 provides a description of the various income sources, amount of income obtained from each source, and share of total income from each source. Households in our sample have an average income of 156,512 Rupees per year. [2] The results show significant variation within and between income sources. Average income per source ranges from 33 Rupees for income from plowing services to 71,842 Rupees for income from off-farm activities. For all income sources, we find that the standard deviation is greater than the average, indicating significant within source heterogeneity. Table 1 also shows, for each income source, the average and standard deviation of shares across households. [3] On average, 53% of household income came from off-farm activities.

We further explore heterogeneity in income sources by distinguishing households into two categories: those that do not diversify income sources (i.e., all household income comes from one source) and those that do diversify. Summary statistics based on this distinction are presented in Table 2. 20% of households do not diversify income sources. On average, households that diversify earn 34,749 Rupees (27%) more income than those that don't diversify. This difference is statistically significant at the 5% level.

Table 3 shows summary statistics for the Simpson index. The average across all households is 0.274. In our study, there are 11 possible income sources and thus a household with equal amounts of income from each source would have a Simpson index of $1 - 1/11 = 0.909$. As discussed above, 50 households do not diversify and as a result have a Simpson index of zero. The maximum value of the Simpson Index in our sample was 0.719, 21% less than the theoretical maximum.

(b) *Social networks*

Figures 2–10 (in the Appendix A) provide visual depictions of the social networks in each of the 9 villages. The figures represent the social network adjacent matrix **W** discussed in 3(b). The figures are constructed using the library *igraph* in R with the Fruchterman and Reingold (1991) layout. [4] The network figures are color coded to represent different attributes of the households which is either the caste or tribal affiliation. For each network **W** a row normalized matrix **A** was constructed to allow for estimation of model (1).

In general, the figures reveal network homophily, i.e. a tendency for nodes (households) with similar characteristics (in our case caste or tribe) to be connected to one another. Table 4 shows Newman's assortative coefficient for the social networks of each village in our sample. In all villages, except village 6, the coefficient is positive indicating a tendency of households of the same caste (or ethnicity) to be more likely to socially connect to one another. We also find a negative correlation between village-level social stratification and income diversification (correlation coefficient of −0.32). This result suggests that social stratification can deter diversification. Unfortunately the small number of villages in this study prevents a deeper village-level analysis of homophilia.

Table 5 presents summary statistics for the two network centrality measures. Mean degree centrality is 0.25 indicating that, on average, households in our sample have network links with 25% of the other households in their respective villages. [5] Eigenvector centrality is a measure that accounts for the network importance of one's network links. Figure 1 is a scatter plot of degree and eigenvector centrality measures calculated for the households in our sample. Although the two measures are positively correlated, the plot shows that some households have high-degree centrality (i.e. several connections) but low eigenvector centrality (i.e. these connections are not to central nodes).

(c) *Demographics, Assets, and shocks*

Table 6 presents summary statistics for the socioeconomic characteristics of households included in our survey. These are determinants of income diversification that have been previously explored in the literature (e.g., Belcher *et al.*, 2015; Ellis, 1998; Martin & Lorenzen, 2016). The average household has 3.1 adult members. 81% of households are headed by males and 40% are scheduled caste or scheduled tribe. The average age of household heads is 53 years and average education is between the primary and upper primary levels.

Eighty-six percent of the sample households use electricity while 70% use gas. On average, these households own 0.17 goats and 1.67 hoes. Average area of owned land is approxi-

Table 1. *Annual income from the different earning sources in Meenangadi, Wayanad*

Sources of income	Income (in Rupees)		Income shares	
	Average	Std. Dev.	Average	Std. Dev.
Sale of market surplus staple crops	18,721.2	47,156.0	0.109	0.212
Sale of surplus fruits and vegetables	81.9	751.6	0.001	0.007
Sale of crop by-products	311.0	1,051.4	0.003	0.019
Sale of livestock, birds	1,452.3	8,169.7	0.009	0.044
Sale of livestock products	10,480.9	45,791.2	0.053	0.149
Income from services by livestock	33.3	407.6	0.000	0.003
Income from agricultural wages	6,134.2	19,972.3	0.047	0.153
Income from off-farm activities (business, trade, non-ag wages)	71,841.8	75,215.9	0.530	0.404
Income from migration	12,389.4	40,947.9	0.070	0.209
Salaried employment	31,063.4	86,872.4	0.139	0.294
Income from other activities	3,885.9	14,807.8	0.039	0.130
Total income	156,511.7	119,914.4	1	—

Source: Authors' analysis of data from the APM Monitoring and Evaluation Survey. See Ragu *et al.* (2013). Exchange Rate: 1 USD = Rs. 59. *Note:* For each income source, Table 1 reports the average share across households. For source s, this average is calculated as $Avg\ Share(s) = \frac{1}{n}\sum_{i=1}^{n} P_{i,s}$. This procedure allow us to calculate standard deviations for each share: $Std.\ Dev.(s) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}\left(P_{i,s} - Avg\ Share(s)\right)^2}$. The average share across sources can be obtained by dividing the average income of source s by the total income of 156,511.7.

Table 2. *Income differences by households who have diversified vs non-diversified in Meenangadi, Wayanad*

	Diversified	Non-diversified
Average income (Rs.)	162,303.2	127,554.2
Standard deviation (Rs.)	8,059.5	9,161.2
Number of households	250	50
Ho: Mean (non-diversified) = Mean (Diversified)		
Ha: Mean (non-diversified) < Mean (Diversified)		
P-value = 0.031		

Source: Authors' analysis of data from the APM Monitoring and Evaluation Survey, 2013. See Ragu *et al.* (2013) and survey conducted by the authors for this paper. Exchange rate: 1 USD = Rs. 59.

Table 3. *Summary statistics – Simpson index*

Variable	Obs	Mean	Std. Dev.	Min	Max
Simpson	300	0.274	0.224	0.000	0.719

Source: Authors' analysis of data from the APM Monitoring and Evaluation Survey, 2013. See Ragu *et al.* (2013).

Table 4. *Assortativity and Simpson indices*

Village	Assortativity coefficient	Simpson index*
1	0.3472	0.3013
2	0.5205	0.2487
3	0.1427	0.1739
4	0.0649	0.3102
5	0.2156	0.3514
6	−0.1626	0.3222
7	0.2818	0.2594
8	0.3414	0.2689
9	0.3548	0.2386

Source: Authors' analysis of data collected by authors.
*denotes village average.

Table 5. *Summary statistics – network centrality*

Variable	Obs	Mean	Std. Dev.	Min	Max
Degree Centrality	300	0.246	0.158	0.028	0.947
Eigenvector Centrality	300	0.436	0.253	0.003	1.000

Source: Authors' analysis of data collected by authors.

mately one acre. Ten percent of the households experienced a crop failure within the 12-month period preceding the interview and 1% experienced a loss of assets in that same time period.

(d) *Estimates of the empirical models*

Table 7 reports parameter estimates for model (1). Statistical inference is based on bootstrapped standard errors because the dependent variable is an index. In addition, we cluster the standard errors to allow for correlations within village, which is relevant for our village-level network model. The table reports parameter estimates based on the GMM/IV strategy which are consistent under the assumption of strict exogeneity of **X** (see Section 3(e) and (i)).

We estimate a statistically significant average network effect of 0.721. That is, a 1% increase in the diversification of a household's social network increases its own diversification by 0.721%. This result suggests a social multiplier of 3.6. In other words, the social network amplifies the effect of other determinants of livelihood diversification (such as education) by 3.6 times. [6]

We find statistical evidence that two demographic characteristics influence income diversification among our sample. First, higher levels of education are associated with higher income diversification. Specifically, a one-step increase in education leads to a direct increase of 1.2% in the Simpson index. However, this effect is amplified by social network feedback loops (i.e., household i affects j that affects i and so on…). With the social multiplier of 3.6, the total effect of a one-step increase in education is a 4.3% increase in the Simpson index. Second, the number of adults in the household is also positively associated with diversification. An additional adult in the household leads to a direct increase of 3% in the household's Simpson index, and almost 11% when the network multiplier is considered.

Although our GMM/IV estimates of the effect of assets on diversification are generally positive and consistent with the previous diversification literature (e.g., Anderson & Deshingkar, 2005; Martin & Lorenzen, 2016), these coeffi-

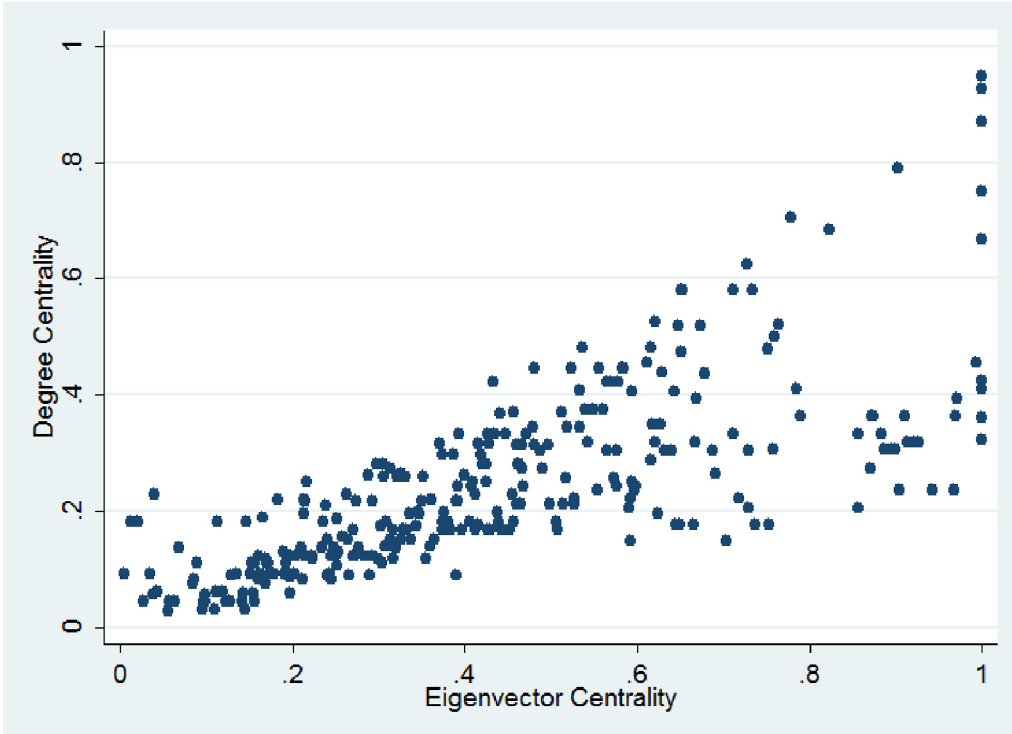


Figure 1. *Scatter plot of centrality measures. Source: Authors' analysis of data collected by authors.*

cients are imprecisely estimated and not statistically significant. We also investigate how households respond to negative shocks. We find that households that suffered a recent crop failure are more diversified. Specifically, having a recent crop failure increases the Simpson index by 12%.

Next we discuss results for the network centrality model. With this model we investigate how social networks affect income diversification by estimating the impact of network centrality on the Simpson index. Contrary to the network model, the centrality model is not affected by the reflection problem discussed by Manski (1993, 2000). These models are thus estimated using OLS.

Next, we discuss two alternative approaches to estimate the average network effect model that serve as robustness checks for our finding that social networks influence income diversification. First, note that Table 2 shows that the Simpson index has a probability mass at 0 and this can be thought of as a corner solution outcome (i.e., some household do not diversify). As Wooldridge (2002, Ch. 16) explains, these corner solution models can be estimated using a censored regression model, or Tobit model. In our case, given the endogeneity of network diversification, we use a maximum likelihood approach to estimate a Tobit model with endogenous explanatory variables (Wooldridge 2002, p. 530). The results corroborate the existence of strong network effects on income diversification. The Tobit estimate of the average network effect is 0.882 ($p < 0.01$).[7] Its 95% confidence interval [0.45,1.31] overlaps with that of the GMM/IV estimate [0.39,1.05], which indicates that these two parameters are not statistically different. Note, however, that consistency of the Tobit estimate depends on homoscedasticity and normality assumptions, which are not needed in the linear GMM/IV approach. In addition, as we cluster standard errors at the village level, statistical inference of the GMM/IV estimate is based on a heteroscedasticity–auto correlation consistent (HAC) variance estimator.

Our second analysis checks the robustness of the social network effect to the measurement of income diversification. The Simpson index takes into consideration all 11 possible sources of income described in Table 1 and, as such, is able to capture fine differences in diversification strategies. For example, two households that fully rely on agricultural activities and have similar total income can have different Simpson measures if one focuses on 3 agricultural activities (e.g., sales of stable crops, by-products, and vegetables), and the other focuses 2 activities (e.g., livestock sales and services). A simpler way to measure income diversification is to compute the share of income that comes from non-agricultural activities. We estimate Eqn. (2)—the average network effect model—by replacing the Simpson index with the share of non-farm income of each household. Results again indicate statistically significant network effects. Specifically, we estimate an average network effect of 0.425 ($p < 0.05$), i.e., a 1% increase in the share of nonfarm income of a household's social network increases its own nonfarm share by 0.425%.[8]

Table 8 reports regression estimates for two representations of network centrality, one in which centrality is measured by the degree index and another where centrality is measured by the eigenvector index. We find that network centrality has a positive and statistically significant effect on diversification. In addition, our estimates suggest that degree centrality has a stronger influence on income diversification than eigenvector centrality. Specifically, a 1% increase in degree centrality increases the Simpson index by 0.19%, while the corresponding increase in eigenvector centrality is 0.12%. These results suggests that the number of social connections of a household has an effect on diversification and that having more connections may be more important than having connections that are centrally positioned in the social network of the village.

Table 6. *Summary statistics – income, demographics, assets, and shocks*

Variable	Description	Mean	Std. Dev.	Min	Max	Skewness
Income	Total income (in thousands of Rupees)	156.5	119.9	4.3	1,079	3.00
Demographics						
Age	Age of household head	52.75	13.04	28	97	0.32
Gender	Gender of household head (male = 1)	0.81	0.40	0	1	−1.55
Education	Education level of household head: 1 = Illiterate, 2 = Read & Write (not formal education), 3 = Primary, 4 = Upper Primary, 5 = High School, 6 = Higher Secondary, 7 = Diploma/ITI, 8 = Graduate & Above	3.50	1.67	1	8	0.18
Number of adults	Count of household members who are above 18 years of age	3.07	1.09	1	8	0.79
Social category	=1 if the household belongs to the scheduled caste or scheduled tribe	0.40	0.49	0	1	0.41
Assets						
Electricity	=1 if the household uses electricity	0.86	0.35	0	1	-2.04
Gas	=1 if the household uses gas	0.70	0.46	0	1	-0.89
Goats	Number of goats owned by the household	0.17	0.67	0	4	4.24
Hoes	Number of hoes owned by the household	1.67	1.71	0	15	2.82
Land	Total land owned in acres	0.97	1.72	0	13	3.74
Shocks						
Crop failure	=1 if the household has faced any crop failure in the past year	0.10	0.30	0	1	2.61
Loss of assets	=1 if the household has faced any loss of assets in the past year	0.01	0.08	0	1	12.02

Source: Authors' analysis of data from the APM Monitoring and Evaluation Survey, 2013. See Ragu *et al.* (2013) and survey conducted by the authors for this paper.

Table 7. *Estimates of the average network effect model*

Dependent variable: *Simpson*	GMM/IV estimates
Average network effect	0.721***
	(0.182)
Demographics	
Age	0.001
	(0.001)
Gender	−0.005
	(0.029)
Education	0.012**
	(0.006)
Number of adults	0.030**
	(0.011)
Social category	0.034
	(0.025)
Assets	
Electricity	0.021
	(0.038)
Gas	0.011
	(0.025)
Goats	0.031
	(0.024)
Hoes	0.009
	(0.009)
Land	−0.002
	(0.009)
Shocks	
Crop failure	0.118***
	(0.034)
Loss of assets	0.189
	(0.184)
Constant	−0.172**
	(0.067)
Observations	272
Wald chi2(13)	122.63
Prob > chi2	0.000

Notes: Bootstrapped standard errors clustered at the village level in parentheses (1000 replications).
Source: Authors' analysis of data from the APM Monitoring and Evaluation Survey, 2013. See Ragu *et al.* (2013) and survey conducted by the authors for this paper.

** $p < 0.05$.
*** $p < 0.01$.

Estimates of the marginal effects of demographics, assets, and income shocks on diversification are robust across the two models and are also similar to the estimates from the network effect model. This suggests that the network effect (1) and network centrality (3) models represent different but consistent approaches to investigate the influence of social networks on diversification. There are, however, a few differences between the models. In particular, the centrality model results show a statistically significant effect of three variables that were not significant in the GMM/IV estimate of model (1): *social category*, *goats*, and *hoes*. This result is, in part, a result of the efficiency loss associated with the instrumental variable approach.

5. DISCUSSION AND CONCLUSIONS

In developing countries, it is rare for households to specialize in a single income-earning activity (Reardon, 1997).

Table 8. *Estimates of the network centrality model*

Dependent variable: *Simpson*	Degree centrality	Eigenvector centrality
Network centrality	0.185**	0.123*
	(0.086)	(0.071)
Demographics		
Age	0.001	0.001
	(0.001)	(0.001)
Gender	−0.008	−0.008
	(0.018)	(0.019)
Education	0.016**	0.014**
	(0.007)	(0.006)
Number of adults	0.032**	0.030**
	(0.015)	(0.013)
Social category	0.070**	0.064**
	(0.023)	(0.024)
Assets		
Electricity	0.031	0.027
	(0.048)	(0.049)
Gas	−0.006	0.001
	(0.019)	(0.019)
Goats	0.035*	0.037*
	(0.021)	(0.019)
Hoes	0.021**	0.021**
	(0.009)	(0.010)
Land	0.001	0.002
	(0.008)	(0.008)
Shocks		
Crop failure	0.112**	0.116**
	(0.043)	(0.045)
Loss of assets	0.193*	0.166
	(0.104)	(0.129)
Constant	−0.054	−0.045
	(0.069)	(0.069)
Observations	272	272
Wald chi2(13)	131.73	129.77
Prob > chi2	0.000	0.000

Notes: Bootstrapped standard errors clustered at the village level in parentheses (1000 replications).
Source: Authors' analysis of data from the APM Monitoring and Evaluation Survey, 2013. See Ragu *et al.* (2013) and survey conducted by the authors for this paper.
* $p < 0.10$.
** $p < 0.05$.

Results in this paper support the importance of income diversification in this poor and socially diverse region of rural Kerala, India. Only 20% of households specialize in a single income-generation activity, the remaining 80% earn income from two or more agricultural or non-agricultural activities. The average Simpson index for the 301 households is 0.274. Using similar income categories and survey methods, Khatun and Roy (2012) found average Simpson indices of 0.206 and 0.562 in two districts of West Bengal, while Saha and Bahal (2014) found an average Simpson index of 0.46 in two other districts of West Bengal. In a study of Western Ghana, Agyeman *et al.* (2014) found an average Simpson Index of 0.338, and in another study of four Chinese provinces, Wei, Chao, and Yali (2016) found Simpson Indices ranging from 0.50 to 0.68. Overall, this suggests a low-to-moderate level of diversification in Wayanad district compared to other developing regions.

The link between social networks and diversification is a relatively unexplored area of research. The main focus of this study was to examine whether social network effects play an important role on income diversification among rural households in Wayanad. We represent villages as networks of households connected through social linkages. In order to examine the effects of social networks, we use a network econometric model, based on the Spatial Autoregressive approach, by replacing the spatial matrix with a network matrix. We also compute network centrality measures to examine whether social position is associated with income diversification. Finally, we measure social stratification using Newman's assortative coefficient and correlate this index with the diversification index.

Average diversification of a household's social network has a positive effect on its diversification. This suggests the importance of ripple effects created through the network interactions. Given the relevance of diversification among rural households, it is important to recognize that social network effects can be critical in promoting diversification activities. We estimate a network multiplier for income diversification of 3.6.[9]

Our multiplier estimate indicates that social networks are an important component of income diversification. This finding adds to previous evidence about the importance of social networks in India. Munshi and Rosenweig (2016) find that gifts and loans flowing through connections between caste members work as substitutes for formal insurance and state-sponsored safety nets. They argue that intra-caste connections are strong and constrain rural–urban migration.

We also use our multiple social dimensions data to measure households' social positions in the village network, i.e. degree and eigenvector centralities. Regression analyses indicate that households with central positions in their village social network have higher income diversification. This finding corroborates previous results in the literature suggesting that unfavorable social position can be an important constraint to economic development. Examples include the influence of informal social support networks on economic vulnerabilities in Georgia (Dershem & Gzirishvili, 1998), and the effect of social capital on household expenditure in South Africa (Maluccio, Haddad, & May, 2000). In rural India, Rao (2001) finds that households that actively participate in festivals have higher social status and receive more invitations to meals from other families. Also in rural India, Banerjee *et al.* (2013) explore a natural experiment when a microfinance institution invited "leaders" of villages in rural India to an informational meeting and requested them to spread information about microfinance programs. The authors find that the centrality of these first-informed villagers is an important predictor of microfinance participation.

Diversification appears to be desirable and highly affected by social networks. What are the implications for program and policy design? This suggests that households in positions of prominence or central households may have positive influence on diversification. Central households tend to have better access to resources due to their position in the network structure. On the other hand this information can also be used to target households (with high centrality) within networks to disseminate important information regarding training opportunities, government initiatives, and availability of skills & services that can facilitate and promote diversification in rural areas.

In contrast with studies such as Baird and Gray (2014) that focus on exchange networks, our interviews capture a range of social and economic connections between households. The descriptive results on the structure of the social networks show that all 9 study villages contain a mixture of castes and/or ethnic groups. Our data allow us to measure social stratification in these rural villages. We use the assortativity coefficient of Newman (2003) to capture village-level homophily. In conjunction with caste information, this coefficient captures the tendency for nodes of the same caste (or ethnic group) to be connected to one another. Therefore, this coefficient can be thought of as a measure of social stratification. Results indicate that caste-based homophily is present in 8 out of the 9 villages in our sample. This result is consistent with the claim of Munshi and Rosenweig (2016) that strong intra-caste connections constrain rural–urban migration in India. Finally, we examine the relationship between the social stratification and income diversification indices at the village level. We conclude that social stratification is negatively correlated with diversification of income sources. This result suggests that while social networks may foster positive influences by serving as conduits for economic development, the benefits of social interactions could be even greater if rural villages in India were less socially stratified.

The study found relatively few significant determinants of diversification. Two of the four significant determinants—education, number of adult members—are consistent with the opportunity-led model of diversification. One of the significant determinants—crop failure—supports the survival-led model, while the final significant determinant—social group—could be consistent with either model. These results are consistent with other similar studies. Assets appear to be less important in this study than other similar studies.

Our work has several limitations. First, our sampling strategy focuses on intermediate-sized villages so that all households in each village are surveyed. While this is important to avoid possible measurement error biases associated with sampled networks, it raises a question about what information is lost by cutting off the tails of the village size distribution. While our sampling design does not offer significant variation to explore the relationship between centrality and village size, the social network data of rural villages in India collected by Banerjee *et al.* (2013) does. [10] Their study offers the opportunity for this examination as they sampled 75 Indian villages (not so far from our study sites) with significant variation in size. Using these data, we regress average degree (i.e. number of connections) on village size and find that the slope coefficient is close to zero and not statistically significant. This result suggests that village size does not systematically affect centrality. However, further diversification research using larger sample sizes is needed to formally test whether or not the effect of network centrality on income diversification varies with village size.

Second, social network analysis requires the specification of network boundaries. Our network models are based on within-village social contacts. It is possible, however, for individuals to engage in between-village interactions that are not captured in our analysis. Nevertheless, we do not think these interactions, if they exist, have large impacts. Our rationale for setting the village as the network boundary of social interaction is twofold: (i) it is well-documented in the literature that the strength of social interactions between individuals decreases with social, or geographic, distance (Akerlof, 1997; Bramoulle *et al.*, 2009). This is also known as the distance-decay effect (Matous, Todo, & Mojo, 2013) and is the foundation of the widespread use of distance decay mechanisms to model spatial dependence in spatial econometric models (Anselin, 1986); (ii) each village in our sample represents a distinct geographical area where clearly identifiable clusters of households form a village. [11] A possible mechanism for social interactions between villages in our sites is the existence of dairy and other agricultural and marketing cooperatives that historically represent an important component of development. However, we have qualitative evidence from focus groups that these cooperatives are not currently perceived to be an important tool for aggregation of social capital; hence, we do not anticipate significant social interaction effects between households of different villages. [12] Despite the limitation of having to specify a "finite" social network, the above arguments make us confident that within-village networks capture the bulk of social influences in this region of India.

Third, our finding of a negative correlation between social stratification and the diversification of income sources is based on a small number of villages. This data restriction prevents us from using social network analysis to further explore caste-based stratification. Future research should concentrate on gathering sufficient village-level information to allow for more robust statistical analyses.

The social network dataset collected by our research team is very rich and its potential and applicability goes beyond the study of income diversification. Other economic dimensions may be influenced by network interactions and should be the focus of future research. Subsequent projects may also focus on the disaggregation of the different dimensions of social contacts to conduct an in-depth analysis of social structure in rural India.

NOTES

1. As described by Jackson (2008), eigenvectors are vectors that, when acted upon by the network matrix, give back some rescaling of themselves. This is useful to capture the idea that an individual is central when it is connected to central individuals. This concept has a self-referential problem: centrality is a function of centrality. The eigenvector of the network matrix associated with the eigenvalue of 1 captures this concept and returns centrality without rescaling.

2. The average exchange rate in 2014 was 63.5 rupees for 1 US dollar.

3. Clarification note: for each income source, table 1 reports the average share across households. For source s, this average is calculated as $Avg\ Share(s) = \frac{1}{n}\sum_{i=1}^{n} P_{i,s}$. This procedure allow us to calculate standard deviations for each share: $Std.\ Dev.(s) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(P_{i,s} - Avg\ Share(s))^2}$. The average share across sources can be obtained by dividing the average income of source s by the total income of 156,511.7.

4. Fruchterman-Reingold algorithm draws undirected graphs according to five generally accepted aesthetic criteria: i) distribute the vertices evenly in the frame; ii) minimize edge crossings; iii) make edge lengths uniform; and iv) reflect inherent symmetry; and v) conform to the frame. Refer to Fruchterman and Reingold (1991) for details.

5. Average degree centrality is often referred in the network literature as network density.

6. The social multiplier is approximated by $\eta \approx 1/(1 - \beta_1)$ (see section 3.4.1). We also estimate the average network effect model using a naive OLS approach to illustrate the importance of controlling for the endogeneity of the network effects. The OLS estimate of β_1 is 0.328, implying a social multiplier of 1.5, a value much smaller than the 3.6 estimate produced by the consistent GMM/IV model. This highlights the downward bias of the OLS model and the importance of controlling for reverse causality in the estimation of the network effect model.

7. Full regression results are available upon request.

8. Full regression results are available upon request.

9. We are the first to estimate multipliers for income diversification; nevertheless, to put this number in perspective, our multiplier is greater than those estimated by Glaeser *et al.* (2003) ranging from 1.67 to 2.17 for wage spillovers in the U.S. They find important social multipliers in different social contexts. They report social multipliers ranging from 1.4 to 2.2 for group membership among college roommates, and from 1.7 to 8.2 for crime rates.

10. The data is available online at https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/21538.

11. In our experience, this contrasts with several study sites in Africa where the boundary between villages are not clearly geographically identified and different villages often constitute a continuous housing landscape.

12. The focus group methodology is described in Breitkreuz *et al.* (in press).

REFERENCES

Agyeman, B. A. S., Asuming-Brempong, S., & Onumah, E. E. (2014). Determinants of income diversification of farm households in the western region of Ghana. *Quarterly Journal of International Agriculture, 53*(1), 55–72.
Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica: Journal of the Econometric Society*, 1005–1027.
Alobo Loison, S. (2015). Rural livelihood diversification in Sub-Saharan Africa: A literature review. *The Journal of Development Studies, 51*(9), 1125–1138.
Anderson, E., & Deshingkar, P. (2005). Livelihood diversification in rural Andra Pradesh, India. In F. Ellis, & HA. Freeman (Eds.). *Rural livelihoods and poverty reduction policies* (pp. 408). Oxon: Routledge.
Anselin, L. (1986). Non-nested tests on the weight structure in spatial autoregressive models: Some Monte Carlo results. *Journal of Regional Science, 26*(2), 267–284.
Baird, T. D., & Gray, C. L. (2014). Livelihood diversification and shifting social networks of exchange: A social network transition?. *World Development, 60*, 14–30.
Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The diffusion of microfinance. *Science, 341*(6144), 1236498.
Barrett, C. B., Reardon, T., & Webb, P. (2001). Nonfarm income diversification and household livelihood strategies in rural Africa: Concepts, dynamics and policy implications. *Food Policy, 26*, 315–331.
Belcher, B., Achdiawan, R., & Dewi, S. (2015). Forest-based livelihoods strategies conditioned by market remoteness and forest proximity in Jharkhand, India. *World Development, 66*, 269–279.
Birthal, P. S., Roy, D., & Negi, D. S. (2015). Assessing the impact of crop diversification on farm poverty in India. *World Development, 72*, 70–92.
Borgatti, S. P. (2005). Centrality and network flow. *Social Networks, 27*(1), 55–71.
Bramoulle, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics, 150*(1), 41–55.
Breitkreuz, R., Stanton, C. J., Brady, N., Pattison, J., King, E. D. I. O., Mishra, C.S., et al. (in press). The Mahatma Gandhi national rural employment guarantee scheme: A policy solution to rural poverty in india? *Development Policy Review*.
Butts, C. T. (2008). Social network analysis with sna. *Journal of Statistical Software, 24*(6), 1–51.
Calvo-Armengol, A., & Jackson, M. O. (2004). The effects of social networks on employment and inequality. *American Economic Review*, 426–454.
Cinner, J. E., & Bodin, Ö. (2010). Livelihood diversification in tropical coastal communities: A network-based approach to analyzing 'livelihood landscapes'. *PLoS One, 5*(8), e11999.
Census of India. (2011). Retrieved from <http://www.censusindia.gov.in/%28S%283zzub5izyowner45cpxr2nao%29%29/2011census/population_enumeration.aspx>.
Costenbader, E., & Valente, T. (2003). The stability of centrality measures when networks are sampled. *Social Networks, 25*(4), 283–307.
De Janvry, A., & Sadoulet, E. (2001). Income strategies among rural households in Mexico: The role of off-farm activities. *World Development, 29*(3), 467–480.
Dershem, L., & Gzirishvili, D. (1998). Informal social support networks and household vulnerability: Empirical findings from Georgia. *World Development, 26*(10), 1827–1838.
Ellis, F. (1998). Household strategies and rural livelihood diversification. *The Journal of Development Studies, 35*(1), 1–38.
Ellis, F. (2000). The determinants of rural livelihood diversification in developing countries. *Journal of Agricultural Economics, 51*(2), 289–302.
Ersado, L. (2006). *Income diversification in Zimbabwe: Welfare implications from urban and rural areas* (Vol. 3964). World Bank Publications.
Fafchamps, M., & Lund, S. (2003). Risk-sharing networks in rural Philippines. *Journal of Development Economics, 71*, 261–287.
Fruchterman, T. M., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience, 21*(11), 1129–1164.
Glaeser, E. L., Sacerdote, B. I., & Scheinkman, J. A. (2003). The social multiplier. *Journal of the European Economic Association, 1*(2–3), 345–353.
Glendenning, C. J., Babu, S., & Asenso-Okyere, K. (2010). Review of agricultural extension in India. IFPRI Discussion Paper 01048. Washington, DC, USA.
Himanshu Lanjouw, P., Murgai, R., & Stern, N. (2013). Nonfarm diversification, poverty, economic mobility, and income inequality: A case study in village India. *Agricultural Economics, 44*(4–5), 461–473.
Jackson, M. O. (2008). *Social and economic networks*. Princeton University Press.
Joshi, P. K., Gulati, A., Birthal, P. S., & Tewari, L. (2004). Agriculture diversification in South Asia: Patterns, determinants and policy implications. *Economic and Political Weekly*, 2457–2467.
Kelejian, H., & Prucha, I. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics, 17*(1), 99–121.
Khatun, D., & Roy, B. C. (2012). Rural livelihood diversification in west Bengal: determinants and constraints. *Agricultural Economics Research Review, 25*(1), 115–124.
Krishna, A. (2010). *One Illness Away: How People Escape Poverty and Why they Become Poor*. Oxford University Press.
Lee, L. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews, 22*(4), 307–335.
Lee, S. H., Kim, P.-J., & Jeong, H. (2006). Statistical properties of sampled networks. *Physics Review E, 73*, 016102.
Liu, Z., & Lan, J. (2015). The sloping land conversion program in China: Effect on the livelihood diversification of rural households. *World Development, 70*, 147–161.
Maertens, A., & Barrett, C. B. (2013). Measuring social networks' effects on agricultural technology adoption. *American Journal of Agricultural Economics, 95*(2), 353–359.
Maluccio, J., Haddad, L., & May, J. (2000). Social capital and household welfare in South Africa, 1993–98. *Journal of Development Studies, 36*(6), 54–81.
Martin, S. M., & Lorenzen, K. (2016). Livelihood diversification in Rural Laos. *World Development, 83*, 231–243.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies, 60*(3), 531–542.
Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives, 14*(3), 115–136.
Matous, P., Todo, Y., & Mojo, D. (2013). Boots are made for walking: Interactions across physical and social space in infrastructure-poor regions. *Journal of Transport Geography, 31*, 226–235.
Matuschke, I., & Qaim, M. (2009). The impact of social networks on hybrid seed adoption in India. *Agricultural Economics, 40*, 493–505.
Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E, 67*, 026126.
Maurin, E., & Moschion, J. (2009). The social multiplier and labour market participation of mothers. *American Economic Journal: Applied Economics, 1*(1), 251–272.
Minot, N., Epprecht, M., Tran, T., Tram, A., & Le Quang, T. (2006). *Income diversification and poverty in the Northern Uplands of Vietnam* Research report. International Food Policy Research Institute, Retrieved from <http://www.ifpri.org/>.
Mohindra, K. S., Haddad, S., & Narayana, D. (2006). Women's health in a rural community in Kerala, India: Do caste and socioeconomic position matter?. *Journal of Epidemiology and Community Health, 60* (12), 1020–1026.
Munshi, K., & Rosenweig, M. (2016). Networks and misallocation: Insurance, migration and the rural-urban wage gap. *American Economic Review, 106*(1), 46–98.
Ragu, P. T., Swallow, B. M., Manaloor, V., Kalaiselvan, N. N., Mahana, R., Arunraj, R., *et al.* (2013). *Alleviating poverty and malnutrition in agrobiodiversity hotspots: Baseline report*. M.S. Swaminathan Research Foundation and University of Alberta, Retrieved from <http://www.rees.ualberta.ca/en/Research/International%20Work.aspx>.
Rajasenan, D., Abraham, B. G., & Rajeev, B. (2013). Health, education and employment in a forward-backward dichotomy based on standard of living index for the tribes in Kerala. *Journal of Economics and Sustainable Development, 4*(7), 100–107.
Rao, V. (2001). Celebrations as social investments: Festival expenditures, unit price variation and social status in rural India. *Journal of Development Studies, 38*(1), 71–97.
Ravi, S., & Engler, M. (2015). Workfare as an effective way to fight poverty: The case of India's NREGS. *World Development, 67*, 57–71.
Reardon, T. (1997). Using evidence of household income diversification to inform study of the rural nonfarm labor market in Africa. *World Development, 25*(5), 735–747.
Reardon, T., Berdegue, J., & Escobar, G. (2001). Rural non-farm employment and incomes in Latin America: Overview and policy implications. *World Development, 29*(3), 395–409.
Saha, B., & Bahal, R. (2014). Livelihood diversification pattern among the farmers of West Bengal. *Economic Affairs, 59*(3), 321.
Stiglitz, J. E. (1989). Markets, market failures and development. *American Economic Review, 79*(2), 197–203.
Wei, D., Chao, H., & Yali, W. (2016). Role of income diversification in reducing forest reliance: Evidence from 1838 rural households in China. *Journal of Forest Economics, 22*, 68–79.
Wichmann, B. (2014). Agents of change and the approximation of network outcomes: A simulation study. *Networks and Spatial Economics, 15*(1), 17–41.
Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge/Massachusetts.

APPENDIX A.

See Figures 2–10.

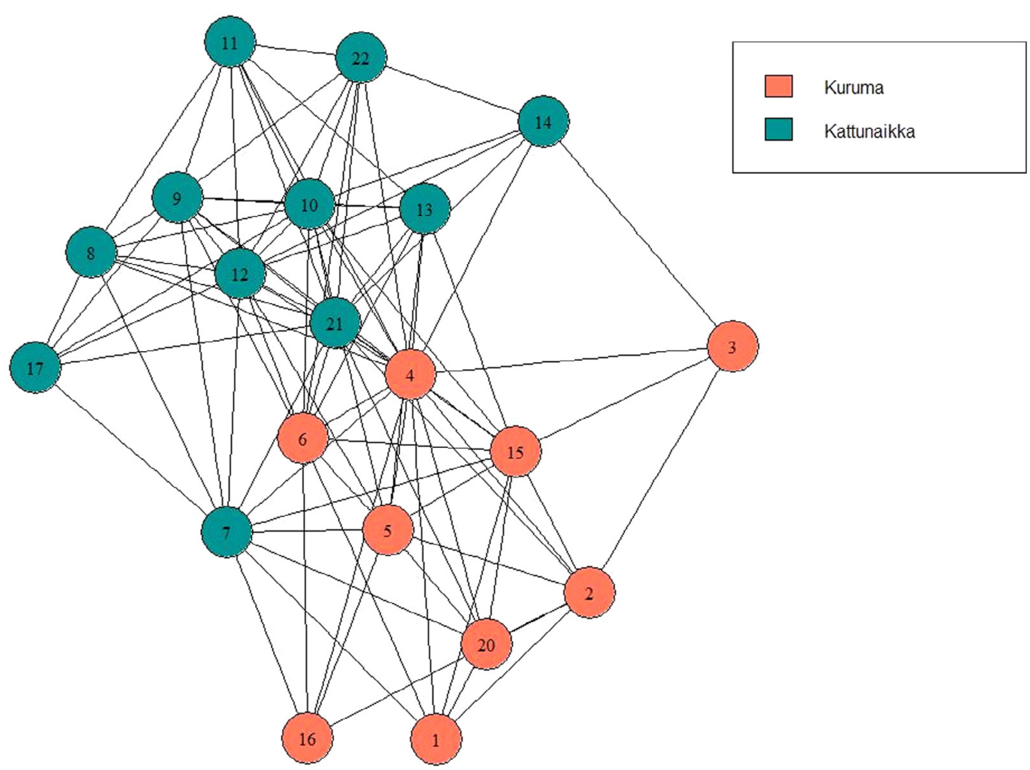


Figure 2. *Visualization of social network for village 1 in Meenangadi, Wayanad.*

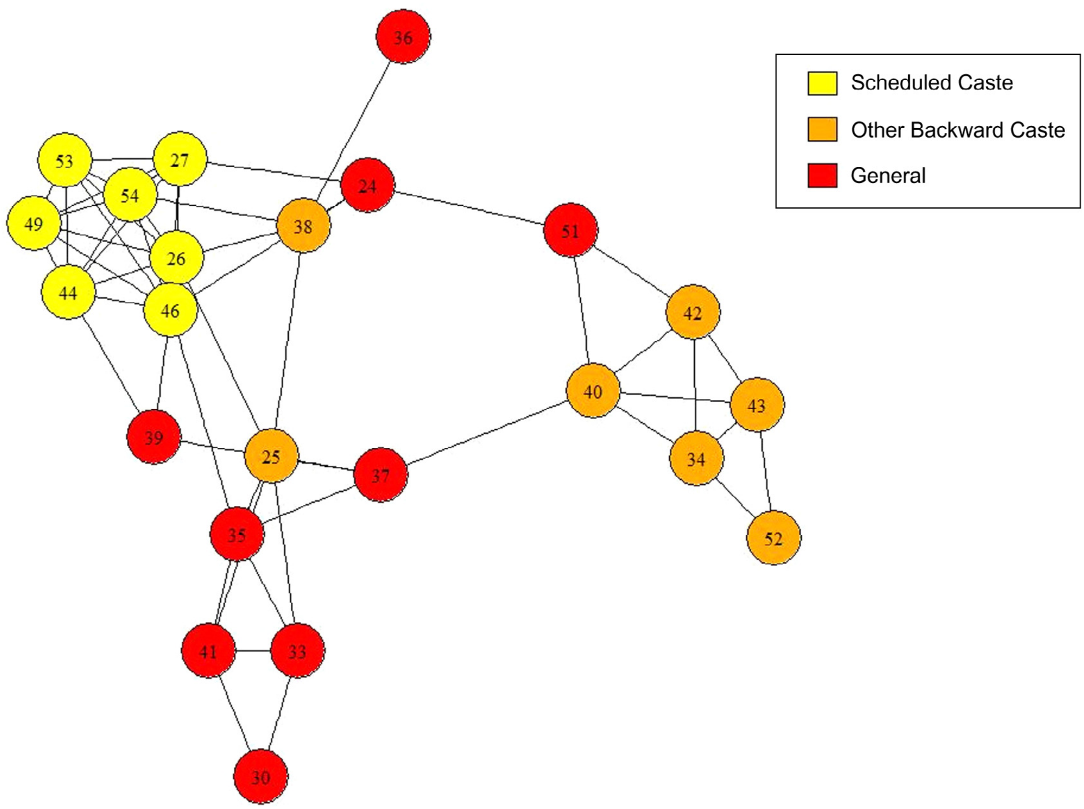


Figure 3. *Visualization of social network for village 2 in Meenangadi, Wayanad.*

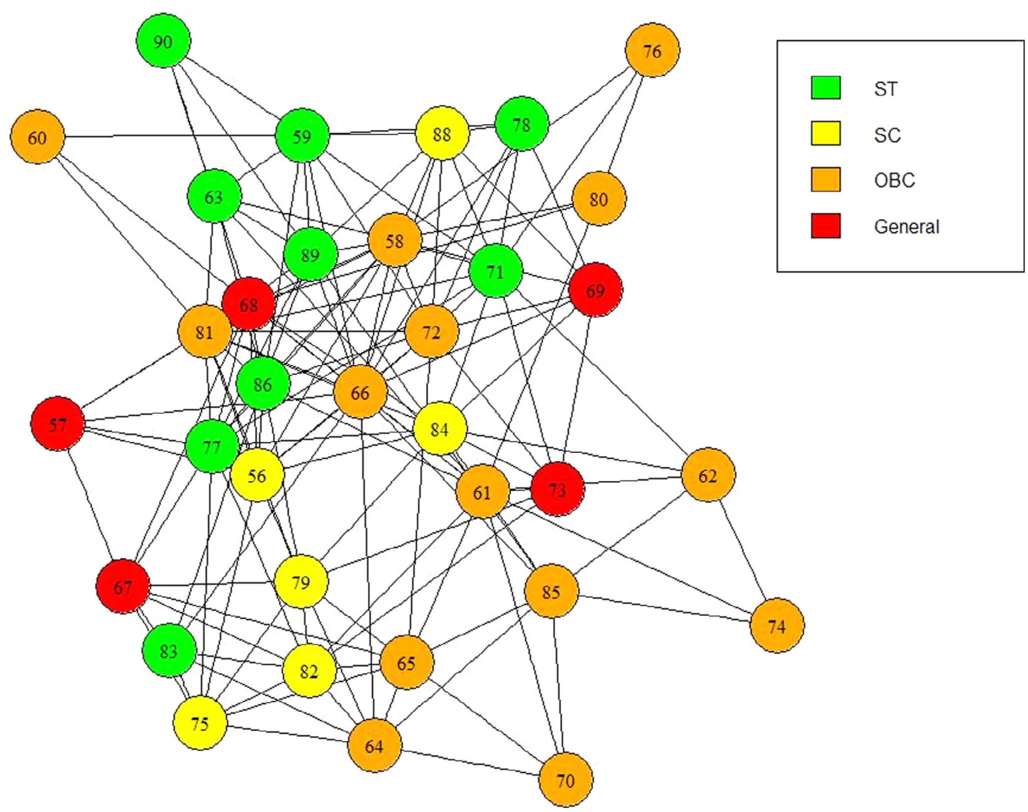


Figure 4. *Visualization of social network for village 3 in Meenangadi, Wayanad.*

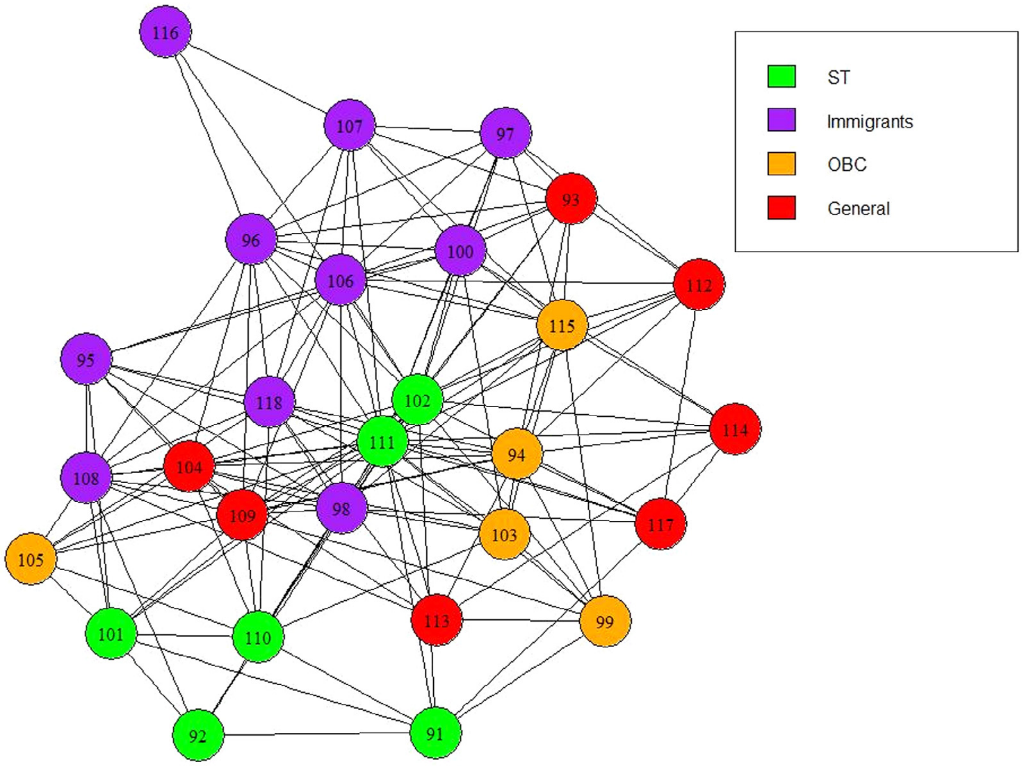


Figure 5. *Visualization of social network for village 4 in Meenangadi, Wayanad.*

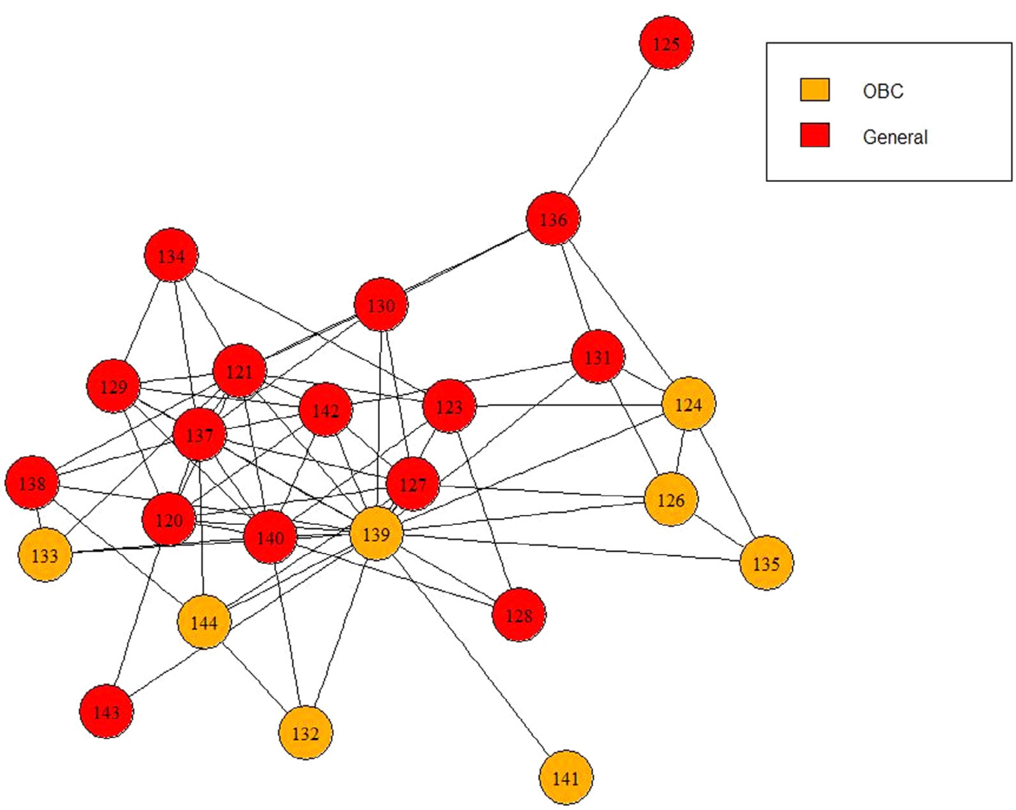


Figure 6. *Visualization of social network for village 5 in Meenangadi, Wayanad.*

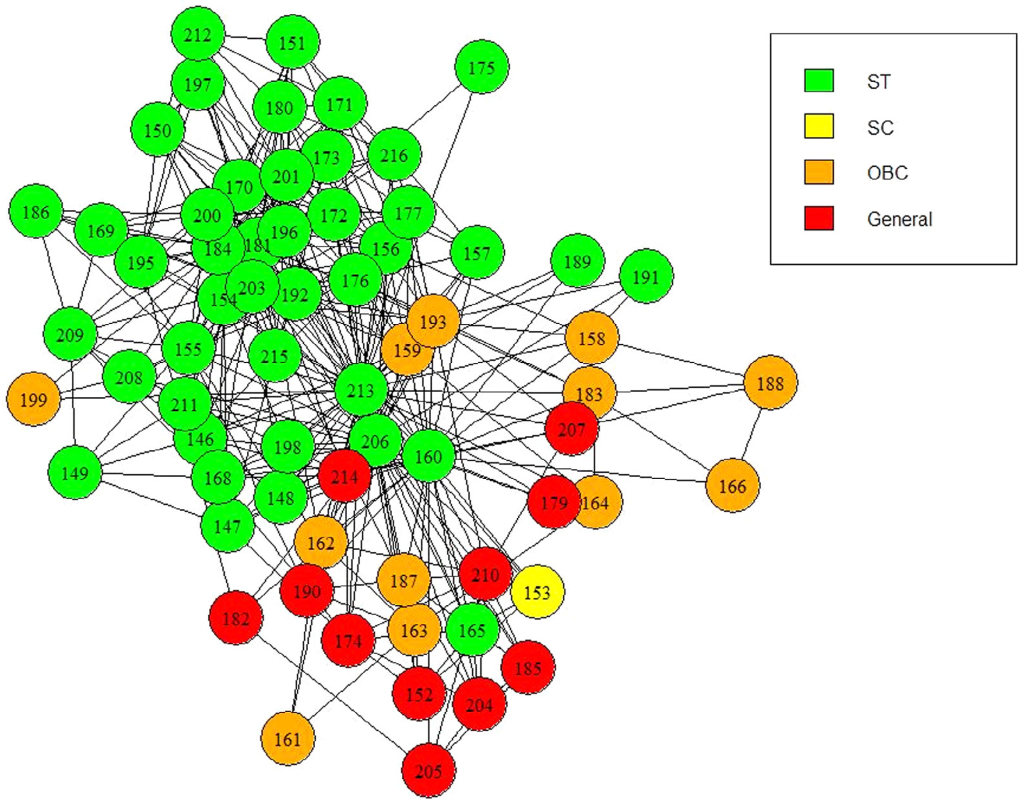


Figure 7. *Visualization of social network for network 6 in Meenangadi, Wayanad.*

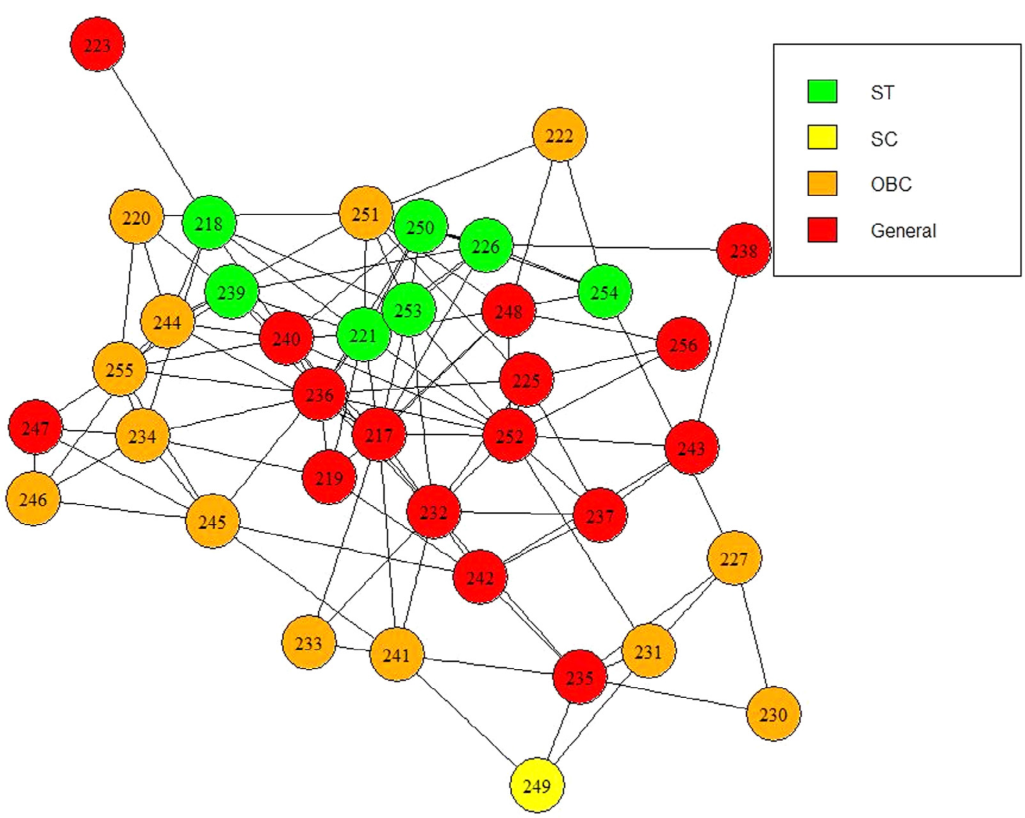


Figure 8. *Visualization of social network for village 7 in Meenangadi, Wayanad.*

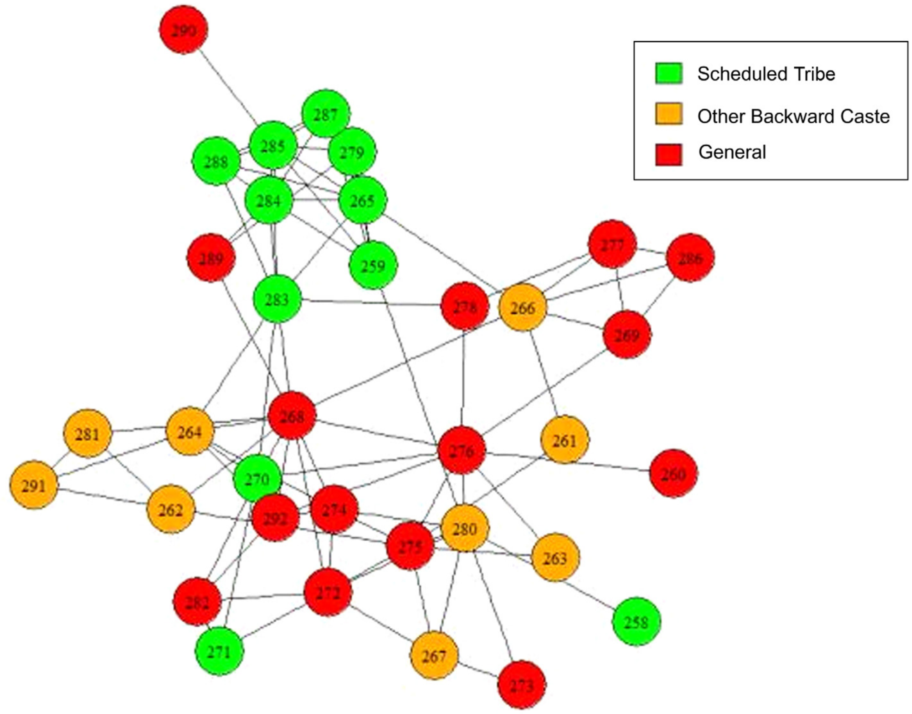


Figure 9. *Visualization of social network for village 8 in Meenangadi, Wayanad.*

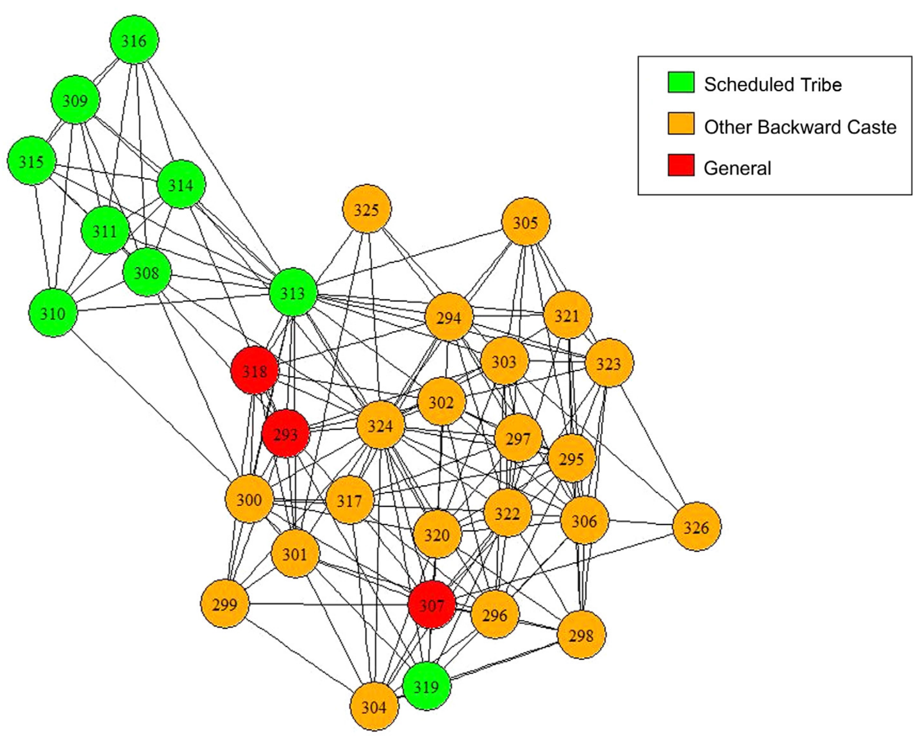


Figure 10. *Visualization of social network for village 9 in Meenangadi, Wayanad.*

Available online at www.sciencedirect.com
ScienceDirect