The pick of the crop: agricultural practices and clustered networks in village economies

A. Groeger (UAB) Y. Zylberberg (Bristol)

II Spanish Workshop in Development Economics, 2024

Agricultural productivity is low in developing economies [Gollin et al., 2014a,b], and there is disparity in (measured) yield and practices [Restuccia and Santaeulalia-Llopis, 2017; Gollin and Udry, 2021].

In Vietnam, there is a large crop productivity gap observed across farms with different portfolios of agricultural commodities.

Research question: what is the role of local networks in explaining crop adoption?

An interesting context: the numerous agricultural commodities of rural Vietnam, and unique data on,

- agricultural production within delineated land parcels,
- the entire networks of family, neighbors, co-workers, friends, etc.,

within four repopulated villages of rural Vietnam.

This paper identifies how the structure of social networks affects the dynamic adoption of high-return agricultural practices.

1. We use the nature of village formation through staggered population resettlement to isolate exogenous variation in network linkages.

2. We find a large network multiplier in crop adoption.

3. A large multiplier may coexist with a relatively low adoption rate when:

   - there is large homophily between connected households,
   - as the endogenous outcome of a dynamic propagation of agricultural practices through clustered networks.

Policies targeting the “in-betweeners”—villagers connecting the different clusters—would be efficient.

Our main contribution is to relate the structure of (full) social networks to the limited adoption of highly-productive agricultural practices.

Agricultural productivity (gaps) in developing countries [Udry, 1996; Restuccia and Rogerson, 2008; Gollin et al., 2014a,b; Chen, 2017; Restuccia and Santaeulalia-Llopis, 2017; Gollin and Udry, 2021; Adamopoulos et al., 2022; Blarel et al., 1992; Shaw-Taylor, 2001; Chen, 2017; Burchardi et al., 2019; Perego, 2019; Adamopoulos and Restuccia, 2020; Le, 2020; Chen et al., 2021; Laskievic, 2021; Lagakos et al., 2018].

Technology adoption and learning through networks: Griliches [1957]; Foster and Rosenzweig [1995]; Bandiera and Rasul [2006]; Duflo et al. [2008]; Conley and Udry [2010]; Dercon and Christiaensen [2011]; Duflo et al. [2011]; Suri [2011]; Emerick et al. [2016]; Kala [2017]; Beaman and Dillon [2018]; BenYishay and Mobarak [2019]; Fabregas et al. [2019]; Comola et al. [2021]; de Janvry et al. [2022].

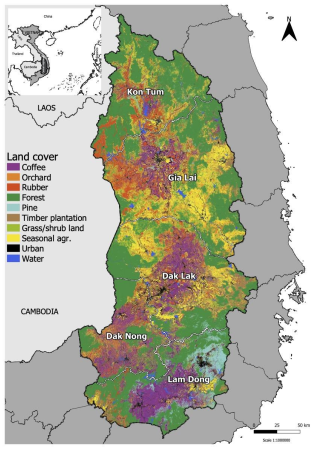
Network structure and targeting: clustering and homophily [Calvó-Armengol et al., 2009; Acemoglu et al., 2011; Golub and Jackson, 2012; Halberstam and Knight, 2016; Ferrali et al., 2020; Jackson et al., 2023], seed targeting [Akbarpour et al., 2023; Sadler, 2023], or influence maximization [Banerjee et al., 2013; Kim et al., 2015; Cai et al., 2015; Banerjee et al., 2019; Beaman et al., 2021].

Observable networks: Banerjee et al. [2013]; Cai et al. [2015]; BenYishay and Mobarak [2019]; BenYishay et al. [2020]; Beaman et al. [2021]; Chakraborty [2022]; Bandiera et al. [2023]; networks “without network data” [Banerjee et al., 2019; Breza et al., 2020; Burlig and Stevens, 2023].

Identification: peer effects [Granovetter, 1973; Manski, 1993; Bramoullé et al., 2009; De Giorgi et al., 2010; Bramoullé et al., 2020; Galeotti et al., 2020; Jochmans, 2023; Lewbel et al., 2024; Chandrasekhar et al., 2024], endogenous network formation [Graham, 2017; De Paula, 2020].

Agricultural production in Vietnam

Crop diversity in (the Central Highlands of) Vietnam:

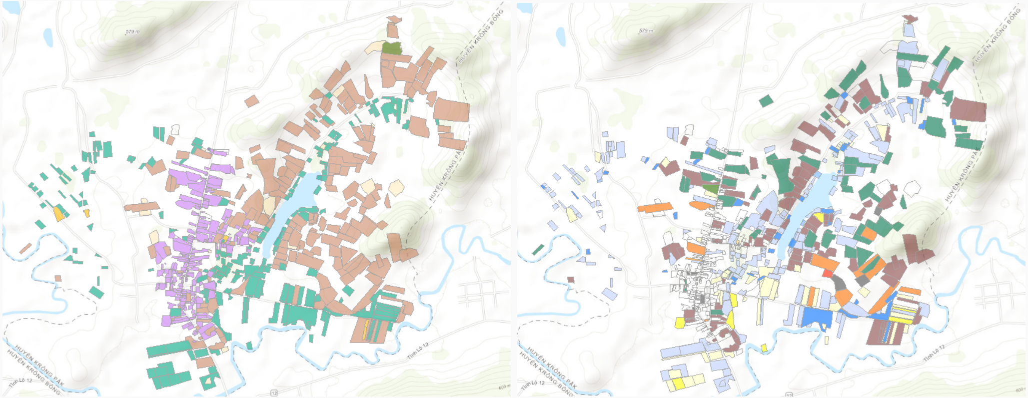


(a) Crop diversity

(b) Buon Ma Thuot Coffee festival

Notes: Panel (a) shows agricultural diversity in the Central Highlands of Vietnam, as inferred from satellite imagery (source: Coffee Vision Project, HEIG-VD/HES-SO). Our villages are located in Dak Lak where the production of coffee (in purple), rubber (in dark orange), and rice/wheat/cassava (in yellow) is widespread. Panel (b) is a photograph of the Buon Ma Thuot Coffee festival organized in 2013 (in Dak Lak); this illustrates the efforts from local/central governments to promote coffee to international investors and to potential local producers.

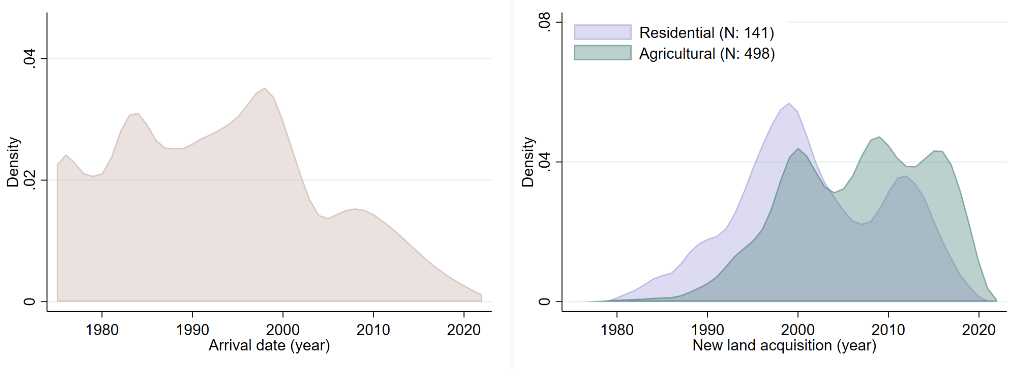
Crop diversity within villages:



(a) Land use (b) Crops

Notes: This map shows the dispersion of agricultural land parcels within “Village 3” in the Central Highlands of Vietnam. The left panel reports the main land usage: residential (purple), perennial (light brown), annual (green), other (yellow). The right panel reports crop types: rice (shades of blue), coffee (brown), cashew nuts (green), pepper (gray), rubber (orange), maize (yellow), others (blank).

Our villages were formed through successive, individual migration spells of Northern families.



(a) Settlement date (b) Land acquisition

Notes: Panel (a) displays the distribution of settlement date across our 950 households; panel (b) displays the distribution of land acquisition date for all land parcels acquired through a market transaction (i.e., not claimed, inherited, or allocated through a government program).

The nature of land acquisition within our villages is a mix between formal resettlement programs, informal settlements, and later land transactions.

Data

a. A novel dataset

A novel dataset

A high-quality panel census of 950 households across 4 villages in the Central Highlands of Vietnam with a focus on agricultural production:

- land geolocation module,
- subjective land evaluation (own, from others),
- objective land evaluation (soil samples),
- expenditures on each production input for each crop across different parcels and activities (e.g., sowing, harvesting, threshing).,

and social networks: structure, usage.

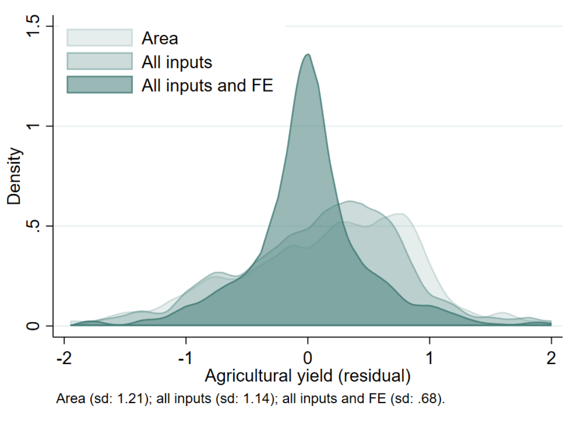
Other modules: standard (household, education, health, activities, income, assets, transfers, savings), beliefs about risk and shocks (including modules on floods and climate change), future strategies.

In summary: around 1,000 questions per household in total.

b. The productivity gap across agricultural commodities

An agricultural productivity gap:

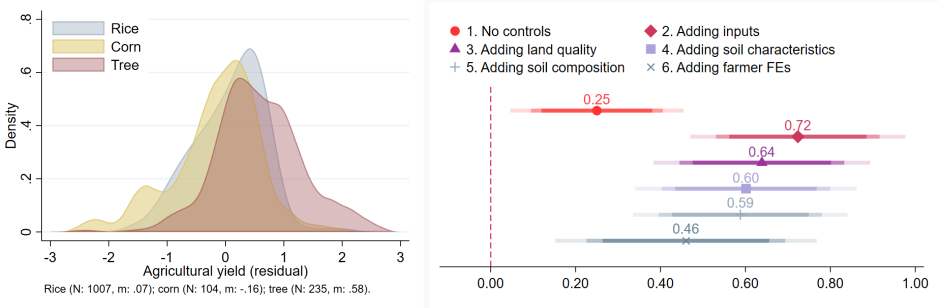
$$\ln y = \ln z + \ln f(\mathbf{x})$$



Notes: This Figure shows the distribution of agricultural TFP, $\ln z_{ic}$, when controlling for: area as the only input; all inputs (area, labor, intermediary, capital); and all inputs and farmer fixed-effects.

A crop productivity gap:

$$\ln z = \mu_c + \varepsilon$$



(a) Distribution across crops

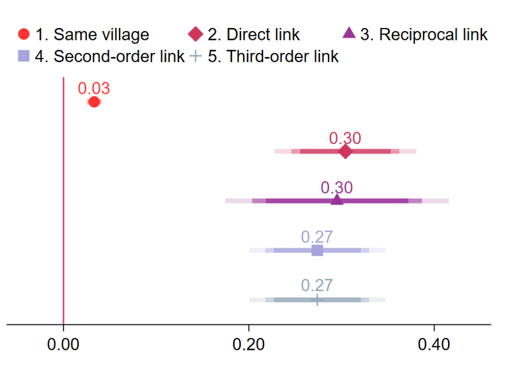
(b) The “tree” premium

Notes: Panel (a) shows the crop-specific distribution of agricultural Total Factor Productivity when controlling for all inputs. Panel (b) shows the tree premium in (log) agricultural Total Factor Productivity without controls and adding sequentially controls for: inputs, land quality, soil characteristics, soil composition, and farmer fixed-effects.

High-return crops (HRC) (**coffee**, cashew, rubber and pepper) are **60–100% more productive** than staple crops and this difference is not explained by inputs or by the general skills of farm managers.

c. The structure of networks

(c) Duration



(d) Homophily

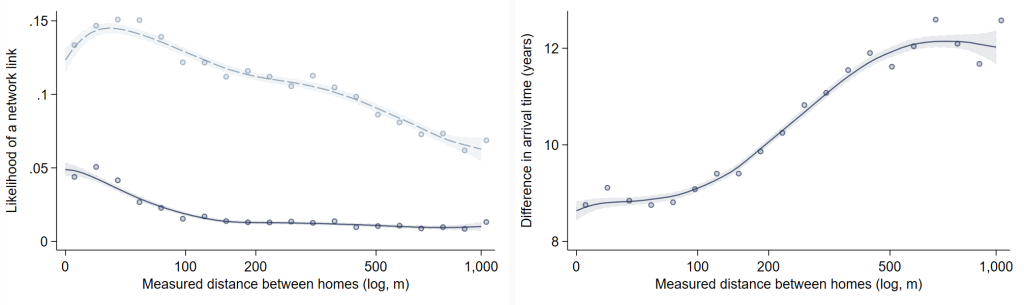
Notes: Panel (a) shows the distribution of a link "duration". Panel (b) shows the correlation between two nodes of a link in terms of growing a prime tree crop.

An illustration, some descriptives, the motivation, and the degree of homophily.

Empirical strategy

Predicting network links

Home proximity, network linkages, and arrival time:



(a) Network linkages (b) Arrival time

Notes: Panel (a) shows the correlation between the existence of a network linkage (first-order in darker blue, second-order in lighter blue and dashed line) and distance between homes across all pairs of unrelated households (using a logarithmic scale). Panel (b) shows the correlation between proximity in arrival times in years and distance between homes across all pairs of households. Note that arrival times are obtained through a retrospective question to households and not from administrative data.

Other correlations.

A treatment through the network

Consider a household i,

- φ_i its portfolio of land parcels including the residential place, and $p \in \varphi_i$ the index of land parcels;
- treatment T_{pi} is defined at the level of a parcel p and is equal to 1 if household i grows a high-return perennial crop;
- the set of non-family-related, yet directly-linked households $\mathcal{L}_i^1$, etc.

The exposure to the treatment through first-order links is defined as follows,

$$\vartheta_i^1 = \frac{\sum_{j \in \mathcal{L}_i^1} \max_{p \in \varphi_j} T_{pj}}{\sum_{j \in \mathcal{L}_i^1} 1} = \sum_{j \in \mathcal{L}_i^1} \frac{1}{\sum_{j \in \mathcal{L}_i^1} 1} \max_{p \in \varphi_j} T_{pj}.$$

We construct similar exposure, but predicted by residential proximity,

$$\theta_i^h = \frac{\sum_{j \in \mathcal{D}_i^h} \max_{p \in \varphi_j} T_{pj}}{\sum_{j \in \mathcal{D}_i^h} 1} = \sum_{j \in \mathcal{D}_i^h} \frac{1}{\sum_{j \in \mathcal{D}_i^h} 1} \max_{p \in \varphi_j} T_{pj}.$$

Results

Predicting exposure:

Exposure (ϑ_i^1)	(1)	(2)	(3)
Predicted exposure (θ_i^h)	0.206	0.210	0.216
	(0.055)	(0.052)	(0.053)
Controls (instrument)	Yes	Yes	Yes
Controls (soil)	No	Yes	Yes
Controls (network)	No	No	Yes
Observations	2,203	2,203	2,203

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include sub-network fixed effects. The dependent variable is the standardized exposure to the treatment; the explaining variable is the standardized, predicted exposure to the treatment—as predicted by proximity between homes. In both cases, the exposures are computed using the allocation of treatment in 2019. The set of (instrument) controls include: the previous status of the parcel in 2019 (treated or not), the number of households in immediate proximity, the average (absolute) altitude differential with other homes in the village, the density of parcels with high-return perennial crops around the various parcels owned by the household, and the density of parcels around the various parcels owned by the household. The set of (soil) controls include: parcel characteristics (area, bulk density, organic carbon content, elevation, slope, distance to the homestead), the latitude, longitude and altitude of the home location. The set of (network) controls include the number of first-order linkages, of second-order linkages, and indicators of network centrality (betweenness, closeness, eigenvector centrality, clustering), and sub-network fixed effects. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

The network multiplier between 2019–2022—a linear specification:

Adoption	(1)	(2)	(3)
Exposure (ϑ_i^1)	0.087	0.090	0.109
	(0.049)	(0.049)	(0.052)
Controls (instrument)	Yes	Yes	Yes
Controls (soil)	No	Yes	Yes
Controls (network)	No	No	Yes
Observations	2,203	2,203	2,203
F-stat	14.08	14.34	16.31

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include sub-network fixed effects. The explaining variable is the standardized exposure to the treatment; the instrument is the standardized, predicted exposure to the treatment—as predicted by proximity between homes. In both cases, the exposures are computed using the allocation of treatment in 2019. The set of (instrument) controls include: the previous status of the parcel in 2019 (treated or not), the number of households in immediate proximity, the average (absolute) altitude differential with other homes in the village, the density of parcels with high-return perennial crops around the various parcels owned by the household, and the density of parcels around the various parcels owned by the household. The set of (soil) controls include: parcel characteristics (area, bulk density, organic carbon content, elevation, slope, distance to the homestead), the latitude, longitude and altitude of the home location. The set of (network) controls include the number of first-order linkages, of second-order linkages, and indicators of network centrality (betweenness, closeness, eigenvector centrality, clustering), and sub-network fixed effects. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

Additional specifications:

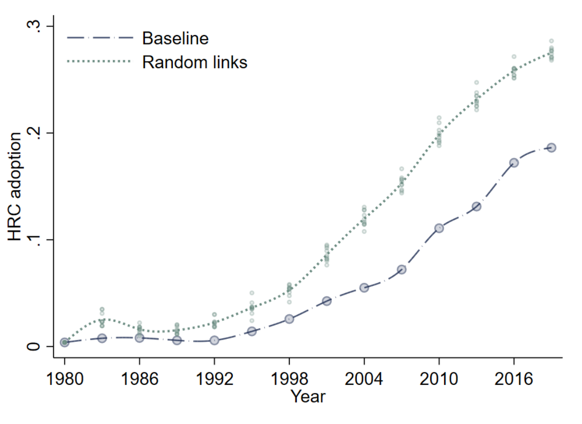
- a probit model;
- a pseudo-panel approach;
- possible mechanisms.

Robustness checks:

- the OLS specification;
- additional controls (longitude/latitude, land quality, soil characteristics, characteristics of friends);
- a placebo;
- alternative instruments/treatments;
- periods of interest and tenure.

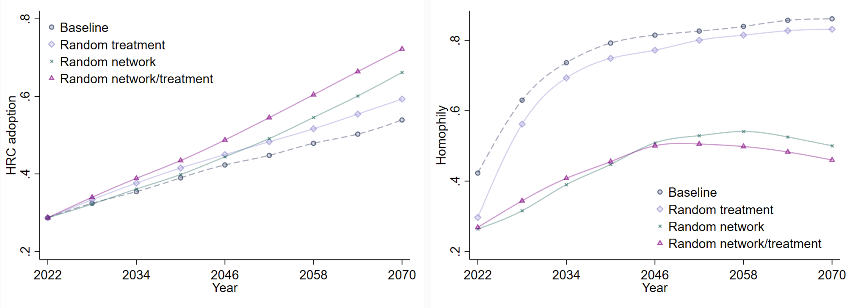
The role of network structure

Network structure and the propagation of treatment—randomizing exposure:



Notes: This Figure compares the actual share of households with a treated parcel (i.e., growing coffee, rubber, cashew nuts, or pepper) to counterfactual shares between 1980–2019. The counterfactuals are based on: (i) the wave-specific predicted likelihood to adopt and treatment exposure at that time, as inferred from estimating wave-specific variations of Equation (3); and 10 re-sampled random networks. In practice, we proceed in a recursive fashion to re-sample networks: 1. we consider the actual settlement date for all households and populate the village accordingly across time; 2. in every year, we re-sample the links that were declared as formed in this exact year among the yet unlinked households; 3. we take the formed links as perennial and proceed to the next wave. The dashed blue line represents the actual share of households; the green dots represent the counterfactual share in each of the 10 experiments; and the green line is the average share across the 10 counterfactuals.

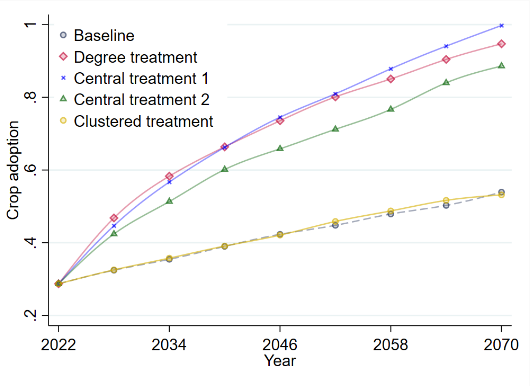
Network structure and the propagation of treatment—homophily and clustering:



(a) Adoption (b) Homophily

Notes: Panel (a) compares the actual share of households with a treated parcel (i.e., growing coffee, rubber, cashew nuts, or pepper) to counterfactual shares between 2022–2070. All simulations assume that crop adoption follows a variation of Equation (3), i.e., $P(y_{in+1} = 1|y_{in} = 0) = a + b\vartheta_i^1$ where n is a wave and i is a land parcel. We consider four scenarios: (i) a baseline projection with the actual network structure and distribution of agricultural practices in 2022 [dashed curve, circles], (ii) randomized agricultural practices in 2022 [light blue, diamonds], (iii) randomized network linkages [green, crosses], and (iv) randomized agricultural practices in 2022 and network linkages [purple, triangles]. Panel (b) displays the evolution of homophily within the network where homophily is the correlation in treatment calculated across undirected links, as in panel (d) of Figure 3. Note that we randomize agricultural practices in 2022 such as to keep the same exact incidence for each village as in the baseline.

Network structure and the propagation of treatment—targeted policies:



Notes: This Figure compares the actual share of households with a treated parcel (i.e., growing coffee, rubber, cashew nuts, or pepper) to counterfactual shares between 2022–2070. All simulations assume that crop adoption follows a variation of Equation (3), i.e., $P(y_{in+1} = 1 | y_{in} = 0) = a + b\vartheta_i^1$ where n is a wave and i is a land parcel. We consider five scenarios: a baseline projection with the actual network structure and distribution of agricultural practices in 2022 [dashed curve, circles], (T1) reshuffled high-return crops to households with the highest number of undirected links [red, diamonds], (T2) reshuffled high-return crops to households with the highest betweenness centrality measure [blue, crosses], (T3) reshuffled high-return crops to households with the highest closeness centrality measure [green, triangles], (T4) reshuffled high-return crops to households with the highest clustering coefficient [gold, circles]. Note that we reshuffle agricultural practices in 2022 such as to keep the same exact incidence for each village as in the baseline.

Thanks!

Appendix

References

Acemoglu, D., Ozdaglar, A. and Yildiz, E.: 2011, Diffusion of innovations in social networks, *2011 50th IEEE Conference on Decision and Control and European Control Conference*, pp. 2329–2334.

Adamopoulos, T., Brandt, L., Leight, J. and Restuccia, D.: 2022, Misallocation, selection, and productivity: A quantitative analysis with panel data from china, *Econometrica* **90**(3), 1261–1282.

Adamopoulos, T. and Restuccia, D.: 2020, Land reform and productivity: A quantitative analysis with micro data, *American Economic Journal: Macroeconomics* **12**(3), 1–39.

Akbarpour, M., Malladi, S. and Saberi, A.: 2023, Just a Few Seeds More: Value of Network Information for Diffusion, *Research papers*, Stanford University, Graduate School of Business.

Bandiera, O., Burgess, R., Deserranno, E., Morel, R., Sulaiman, M. and Rasul, I.: 2023, Social incentives, delivery agents, and the effectiveness of development interventions, *Journal of Political Economy Microeconomics* **1**(1).

Bandiera, O. and Rasul, I.: 2006, Social networks and technology adoption in northern mozambique, *The Economic Journal* **116**(514), 869–902. **URL:** *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1468-0297.2006.01115.x*

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O.: 2013, The diffusion of microfinance, *Science* **341**(6144), 1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. and Jackson, M. O.: 2019, Using gossips to spread information: Theory and evidence from two randomized controlled trials, *The Review of Economic Studies* **86**(6), 2453–2490.

Beaman, L., BenYishay, A., Magruder, J. and Mobarak, A. M.: 2021, Can network theory-based targeting increase technology adoption?, *American Economic Review* **111**(6), 1918–43.

Beaman, L. and Dillon, A.: 2018, Diffusion of agricultural information within social networks: Evidence on gender inequalities from mali, *Journal of Development Economics* **133**, 147–161.

BenYishay, A., Jones, M., Kondylis, F. and Mobarak, A. M.: 2020, Gender gaps in technology diffusion, *Journal of Development Economics* **143**, 102380.
URL: *https://www.sciencedirect.com/science/article/pii/S030438781930197X*

BenYishay, A. and Mobarak, A. M.: 2019, Social learning and incentives for experimentation and communication, *The Review of Economic Studies* **86**(3), 976–1009.

Blarel, B., Hazell, P., Place, F. and Quiggin, J.: 1992, The economics of farm fragmentation: evidence from ghana and rwanda, *The World Bank Economic Review* **6**(2), 233–254.

Bramoullé, Y., Djebbari, H. and Fortin, B.: 2009, Identification of peer effects through social networks, *Journal of Econometrics* **150**(1), 41–55.

Bramoullé, Y., Djebbari, H. and Fortin, B.: 2020, Peer effects in networks: A survey, *Annual Review of Economics* **12**, 603–629.

Breza, E., Chandrasekhar, A. G., McCormick, T. H. and Pan, M.: 2020, Using aggregated relational data to feasibly identify network structure without network data, *American Economic Review* **110**(8), 2454–2484.

Burchardi, K. B., Gulesci, S., Lerva, B. and Sulaiman, M.: 2019, Moral hazard: Experimental evidence from tenancy contracts, *The Quarterly Journal of Economics* **134**(1), 281–347.

Burlig, F. and Stevens, A. W.: 2023, Social networks and technology adoption: Evidence from church mergers in the us midwest, *American Journal of Agricultural Economics* .

Cai, J., De Janvry, A. and Sadoulet, E.: 2015, Social networks and the decision to insure, *American Economic Journal: Applied Economics* **7**(2), 81–108.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y.: 2009, Peer effects and social networks in education, *The review of economic studies* **76**(4), 1239–1267.

Chakraborty, A.: 2022, Network-based targeting with heterogeneous agents for improving technology adoption, *Technical report*.

Chandrasekhar, A. G., Goldsmith-Pinkham, P., McCormick, T., Thau, S. A. and Wei, J.: 2024, Non-robustness of diffusion estimates on networks with measurement error, *Technical report*, National Bureau of Economic Research.

Chen, C.: 2017, Untitled land, occupational choice, and agricultural productivity, *American Economic Journal: Macroeconomics* **9**(4), 91–121.

Chen, C., Restuccia, D. and Santaeulàlia-Llopis, R.: 2021, The effects of land markets on resource allocation and agricultural productivity, *Review of Economic Dynamics* .

Comola, M., Inguaggiato, C. and Mendola, M.: 2021, Learning about Farming: Innovation and Social Networks in a Resettled Community in Brazil, *IZA Discussion Papers 14092*, Institute of Labor Economics (IZA).
URL: *https://ideas.repec.org/p/iza/izadps/dp14092.html*

Conley, T. G. and Udry, C. R.: 2010, Learning about a new technology: Pineapple in ghana, *American Economic Review* **100**(1), 35–69.

De Giorgi, G., Pellizzari, M. and Redaelli, S.: 2010, Identification of social interactions through partially overlapping peer groups, *American Economic Journal: Applied Economics* **2**(2), 241–275.

de Janvry, A., Rao, M. and Sadoulet, E.: 2022, Seeding the seeds: Role of social structure in agricultural technology diffusion, *Technical report*.

De Paula, Á.: 2020, Econometric models of network formation, *Annual Review of Economics* **12**, 775–799.

Dercon, S. and Christiaensen, L.: 2011, Consumption risk, technology adoption and poverty traps: Evidence from ethiopia, *Journal of Development Economics* **96**(2), 159–173.

Duflo, E., Kremer, M. and Robinson, J.: 2008, How high are rates of return to fertilizer? evidence from field experiments in kenya, *American Economic Review* **98**(2), 482–88.

Duflo, E., Kremer, M. and Robinson, J.: 2011, Nudging farmers to use fertilizer: Theory and experimental evidence from kenya, *American Economic Review* **101**(6), 2350–90.

Emerick, K., De Janvry, A., Sadoulet, E. and Dar, M. H.: 2016, Technological innovations, downside risk, and the modernization of agriculture, *American Economic Review* **106**(6), 1537–61.

Fabregas, R., Kremer, M. and Schilbach, F.: 2019, Realizing the potential of digital development: The case of agricultural advice, *Science* **366**(6471), eaay3038.

Ferrali, R., Grossman, G., Platas, M. R. and Rodden, J.: 2020, It takes a village: Peer effects and externalities in technology adoption, *American Journal of Political Science* **64**(3), 536–553.

Foster, A. D. and Rosenzweig, M. R.: 1995, Learning by doing and learning from others: Human capital and technical change in agriculture, *Journal of Political Economy* **103**(6), 1176–1209.

Galeotti, A., Golub, B. and Goyal, S.: 2020, Targeting interventions in networks, *Econometrica* **88**(6), 2445–2471.

Gollin, D., Lagakos, D. and Waugh, M. E.: 2014a, Agricultural productivity differences across countries, *American Economic Review* **104**(5), 165–70.

Gollin, D., Lagakos, D. and Waugh, M. E.: 2014b, The agricultural productivity gap, *The Quarterly Journal of Economics* **129**(2), 939–993.

Gollin, D. and Udry, C.: 2021, Heterogeneity, measurement error, and misallocation: Evidence from african agriculture, *Journal of Political Economy* **129**(1), 1–80.

Golub, B. and Jackson, M. O.: 2012, How homophily affects the speed of learning and best-response dynamics, *The Quarterly Journal of Economics* **127**(3), 1287–1338.

Graham, B. S.: 2017, An econometric model of network formation with degree heterogeneity, *Econometrica* **85**(4), 1033–1063.

Granovetter, M. S.: 1973, The strength of weak ties, *American Journal of Sociology* **78**(6), 1360–1380.

Griliches, Z.: 1957, Hybrid corn: An exploration in the economics of technological change, *Econometrica* **25**(4), 501–522.

Halberstam, Y. and Knight, B.: 2016, Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter, *Journal of Public Economics* **143**, 73–88.

Hengl, T.: 2018, Soil bulk density (fine earth) 10 x kg / m-cubic at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution.
URL: *https://doi.org/10.5281/zenodo.1492157*

Hengl, T. and Wheeler, I.: 2018, Soil organic carbon content in x 5 g / kg at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution.
URL: *https://doi.org/10.5281/zenodo.1475458*

Jackson, M. O., Nei, S. M., Snowberg, E. and Yariv, L.: 2023, The dynamics of networks and homophily, *Technical report*, National Bureau of Economic Research.

Jochmans, K.: 2023, Peer effects and endogenous social interactions, *Journal of Econometrics* **235**(2), 1203–1214.
URL: *https://www.sciencedirect.com/science/article/pii/S0304407622001853*

Kala, N.: 2017, Learning, adaptation, and climate uncertainty: Evidence from indian agriculture, *Technical report*.

Kim, D. A., Hwong, A. R., Stafford, D., Hughes, D. A., O'Malley, A. J., Fowler, J. H. and Christakis, N. A.: 2015, Social network targeting to maximise population behaviour change: a cluster randomised controlled trial, *The Lancet* **386**(9989), 145–153.

Lagakos, D., Mobarak, A. M. and Waugh, M. E.: 2018, The welfare effects of encouraging rural-urban migration, *Technical report*, National Bureau of Economic Research.

Laskievic, A. A.: 2021, Commodity booms and structural transformation: The role of input use and land inequality, *Technical report*.

Le, K.: 2020, Land use restrictions, misallocation in agriculture, and aggregate productivity in vietnam, *Journal of Development Economics* **145**, 102465.

Lewbel, A., Qu, X. and Tang, X.: 2024, Ignoring measurement errors in social networks, *The Econometrics Journal* **27**(2), 171–187.

Manski, C. F.: 1993, Identification of endogenous social effects: The reflection problem, *The Review of Economic Studies* **60**(3), 531–542.

Perego, V. M.: 2019, Crop prices and the demand for titled land: Evidence from uganda, *Journal of Development Economics* **137**, 93–109.

Restuccia, D. and Rogerson, R.: 2008, Policy distortions and aggregate productivity with heterogeneous establishments, *Review of Economic Dynamics* **11**(4), 707–720.

Restuccia, D. and Santaeulalia-Llopis, R.: 2017, Land misallocation and productivity, *Technical report*, National Bureau of Economic Research.

Sadler, E.: 2023, Seeding a simple contagion, *Technical report*.

Shaw-Taylor, L.: 2001, Parliamentary enclosure and the emergence of an english agricultural proletariat, *The Journal of Economic History* **61**(3), 640–662.

Suri, T.: 2011, Selection and comparative advantage in technology adoption, *Econometrica* **79**(1), 159–209.

Udry, C.: 1996, Gender, agricultural production, and the theory of the household, *Journal of Political Economy* **104**(5), 1010–1046.

TVSEP

Data source: a high-quality panel survey of 2,200 households across 220 villages (TVSEP).

Survey sampling and attrition:

- random sampling within province (36), sub-district (2) and village (10),
- minimal attrition of about 10% over 10 years (2007, 2008, 2010, 2013, 2016, and 2017).

Household characteristics: demographics, labor status, health, education, migration, income and expenditures, assets, borrowing/saving, shocks/risks.

Agricultural production: expenditures on each production input for each crop across different parcels and activities (e.g., sowing, harvesting, threshing).

Back

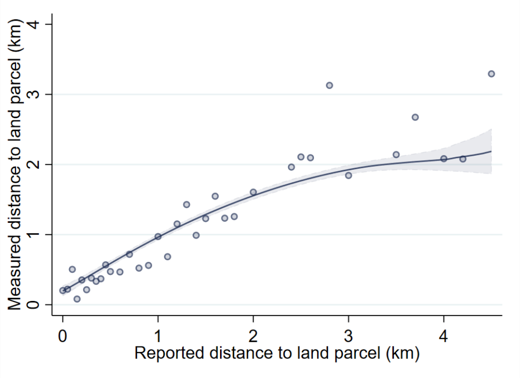
A geolocation module

The geo-localization of parcels proceeds as follows:

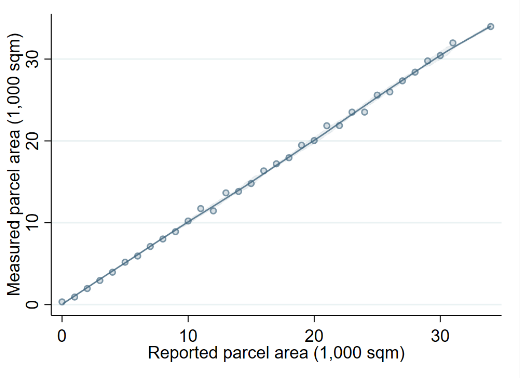
- An offline satellite map is prepared and augmented by the addition of points of interest (e.g., gas stations, supermarkets, schools etc.)—the map covers a radius of 8 kilometers around the village centroid.
- The software automatically centers the map around the current location (the “House”); the interviewer then helps the respondent navigate by showing her/him the main points of interest, the main roads, the waterways. In practice, the most efficient way of finding a land parcel is to ask the respondent to follow the usual route on the map, starting from the house to the land parcel.
- Once the location of the land parcel is identified by the respondent on the map, the interviewer draws a polygon under the instructions of the respondent.
- Additional questions help capture possible issues with the geo-localization, e.g., how sure the respondent may be, how much help was needed etc.

Back

A geolocation module



(a) Distance to homestead

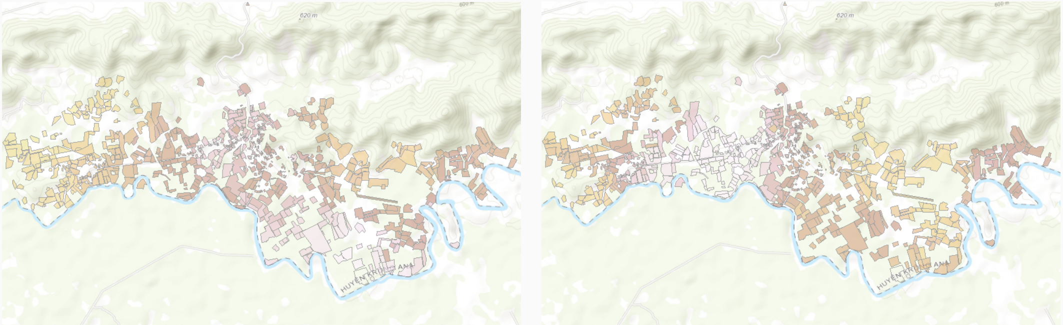


(b) Land area

Notes: These figures report a validation exercise for the geo-located land parcels. The left panel reports the relationship between the measured distance to the land parcel (x-axis) and the distance as *reported by the respondent* (y-axis). The right panel reports the relationship between the measured land area (x-axis) and the area as reported by the respondent (y-axis). In both instances, we create bins of observations along the x-axis variable and the dots represent the average of the y-axis variable within each bin. The lines are locally weighted regressions with the associated 95% confidence interval.

A soil module

We collect about 300 soil samples, and test for: *pH* (acidity), *OMts* (organic content), *Nts* (nitrogen), P_2O_5dt (phosphate), K_2Odt (potash).



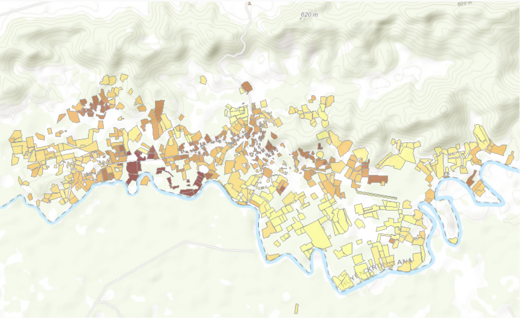
(a) *pH* (acidity) (b) P_2O_5dt (phosphate)

Notes: These figures report the inferred soil content from about 300 soil samples in the village of Buon Triet: *pH*, acidity; and P_2O_5dt, phosphate).

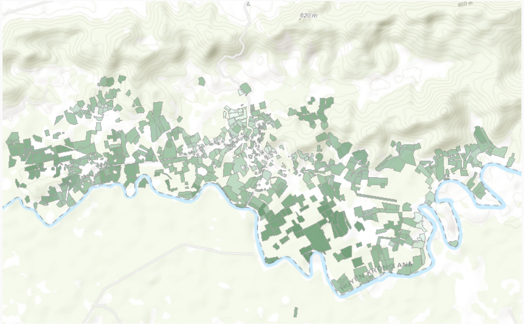
Back

A soil module

We extract high-quality data on topography—30m precision—and soil characteristics—100m precision [Hengl, 2018; Hengl and Wheeler, 2018].



(a) Soil bulk density at 0 cm depth



(b) Organic carbon content

Notes: These figures report inferred carbon content and soil bulk in the village of Buon Triet. We extract high-quality data on topography (30m precision), soil characteristics (100m precision) and temperature/precipitation (1km precision) from Google Earth Engine. More specifically, we construct the maximum, minimum and average elevation within each hexagon; the average slope; the soil bulk density at 0 cm depth as reconstructed from recent satellite imagery; the organic carbon content.

Agricultural production in the survey ($y_t, \mathbf{x}_t, c$)

Measures of agricultural production for a household i at time t growing a certain crop c:

- income (y_{ict});
- quantity (q_{ict});
- expenditures, e.g., on fertilizers (e_{ict});
- hours provided by family members or casual farm workers (l_{ict});
- expenditures on capital (k_{ict});
- intermediate inputs (m_{ict});
- and cultivated land area (t_{ict}).

Crops: areca nut, bamboo, cajuput tree, cashew nut, cassava, casuarina, coffee, cotton, eucalyptus, flower, fruits, gluey tree, grass, green bean, kapol, kenaf, lotus, maize, mulberry, nuts, palm oil, pepper, rubber, soybean, sugarcane, sweet potato, tea, tobacco, vegetables, rice.

Back

A network module

The identification of the household network relies on:

(i) a list of contacts with their name, age, gender, phone (last 6 digits),
(ii) a description of their relationship with the different household members,
(iii) and references to these contacts when relevant along the questionnaire.

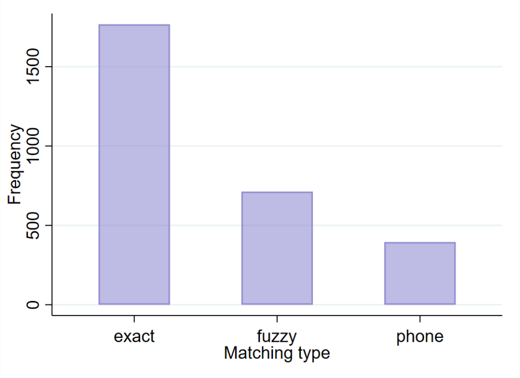
Enumerators would establish a preliminary list and update the list as the interview went along.

The matching algorithm proceeds in steps:

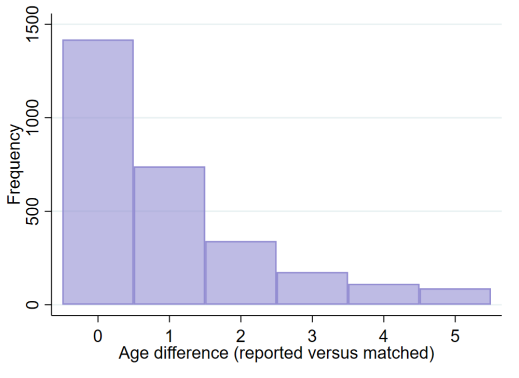
(i) matching is performed on gender, age (within a window of 5 years), and the last 6 digits of the phone number,
(ii) among unmatched entries, matching is then based on gender, age, and exact name matching,
(iii) and unmatched entries are finally matched through a fuzzy matching on names, accounting for specificities of the Vietnamese language.

The outcome of this matching procedure is 2,900 linkages from 4,000 reported contacts—a match rate of about 71% because villagers may report names of contacts outside the scope of our household survey.

A network module



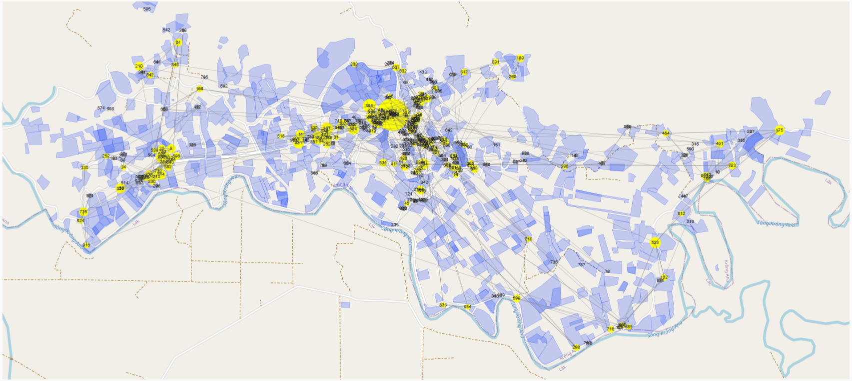
(a) Method



(b) Age difference (reported versus actual)

Notes: The left panel reports the number of matches per matching method (phone: match based on gender, phone number, and age; exact: match based on exact string matching between names; fuzzy: match based on fuzzy string matching between names). The right panel reports the distribution of age differences (reported versus actual) within the sample of matched contacts.

Back



Notes: The figure displays the location of houses (yellow circles) and land parcels (blue polygons) in Buon Triet. The size of circles indicate the number of times the members of a given household is mentioned as a contact by another respondent, and the arrows illustrate the spatial distributions of these links.

Social links form spatial clusters:

(i) households of a same extended family are more likely to live nearby,

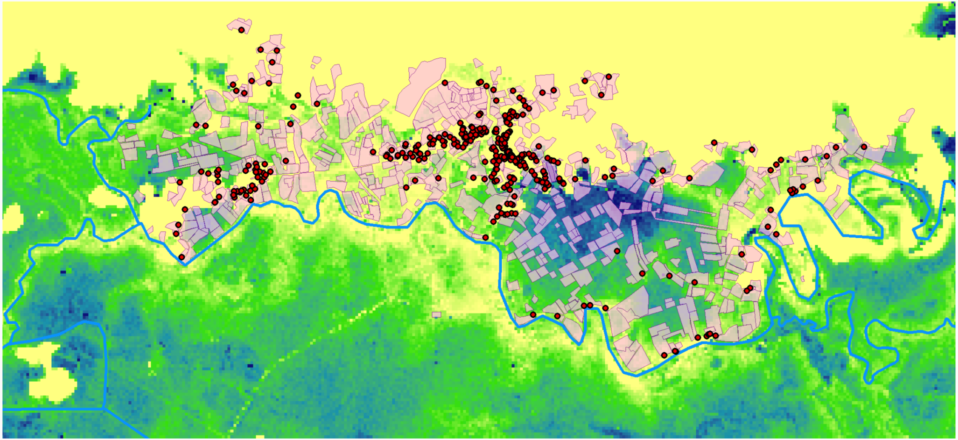
(ii) neighbors are more likely to form a labor exchange network

A flood module

We collect highly-precise elevation data from the FloodAdaptVN project (`https://floodadapt.eoc.dlr.de/`), at a resolution of 30m and a vertical accuracy below 1 meter.

A flood module

We integrate the highly-precise elevation data within the global Flood Hazard Model.



Notes: This map shows the dispersion of land parcels and deciles of flood risk in Buon Triet. Flood hazard is computed from simulations based on 1-in-100 years events. Land parcels located South-East of the village are very vulnerable to fluvial inundations, in contrast with low-lying land in the West and higher grounds around the village center.

A climate change module

We capture beliefs about climate change and mitigation strategies as follows:

(i) Have you experienced recent changes in [X] or do you expect [X] to increase/decrease/remain stable in the future?

(ii) If so, when/where etc.?

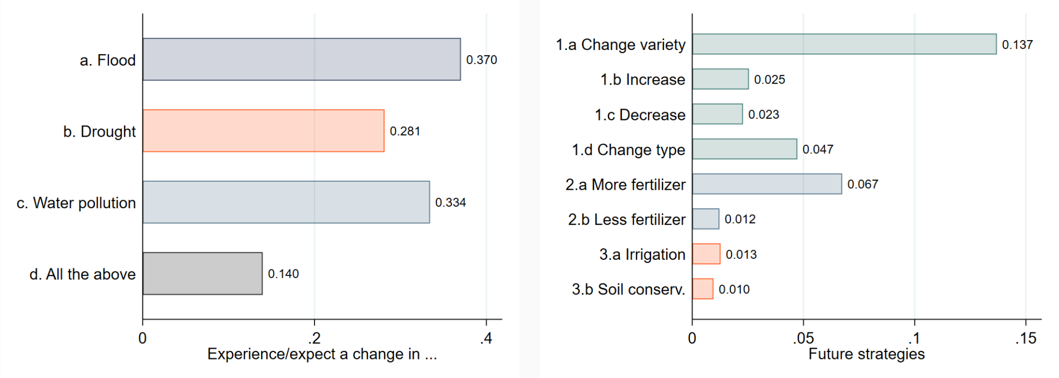
(iii) Which one of the following strategies do you expect to adopt in the future?

(iv) Is adopting [X] motivated by changes in agricultural conditions or environmental concerns?

We also capture plans about crops, irrigation etc. in the land module.

Back

Climate change and future strategies



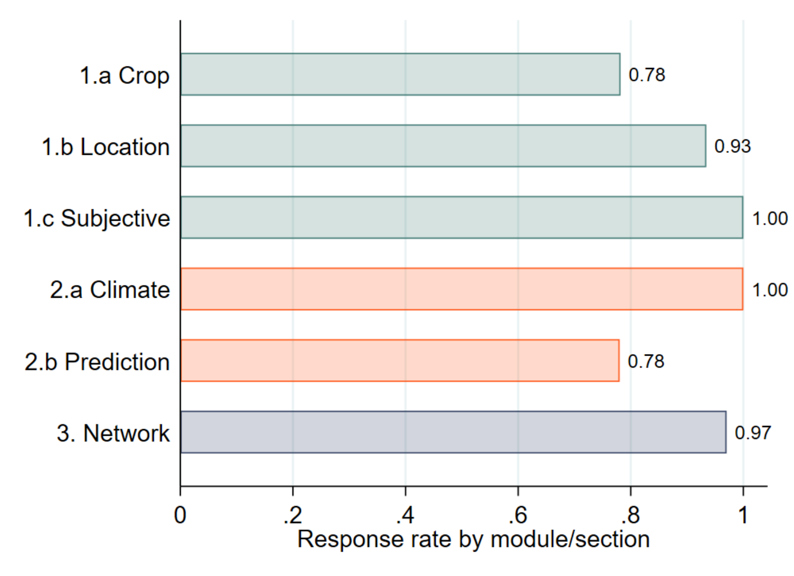
(a) Climate change

(b) Strategies

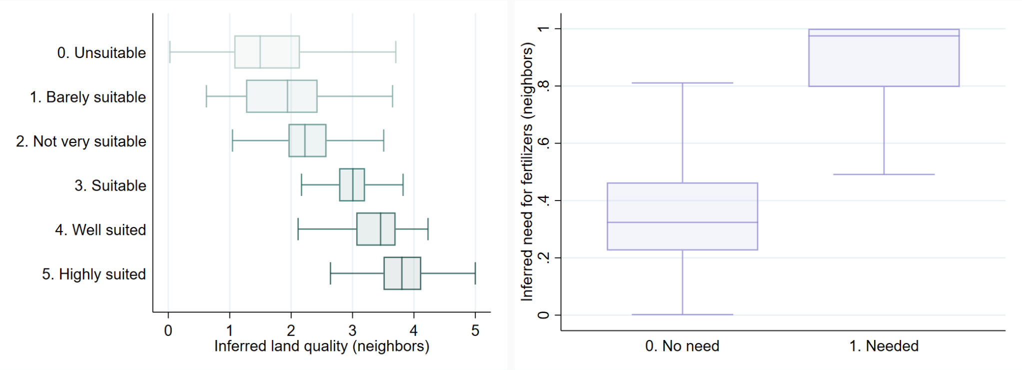
Notes: The left panel reports the share of respondents having experienced or expecting a change in agricultural conditions (flooding, droughts, water pollution). The right panel reports the share of respondents declaring considering a certain strategy.

Back

A long questionnaire...



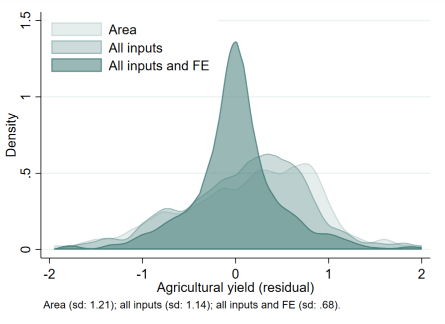
Subjective evaluations



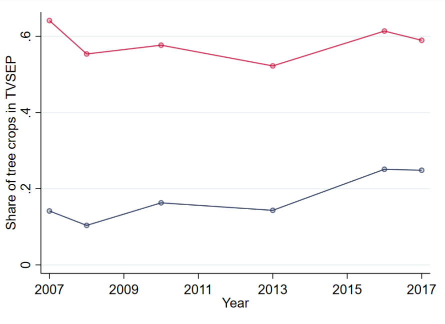
(a) Land quality (b) Fertilizers

Notes: Panel (a) shows the correlation between the farmer's evaluation of the quality of their own land parcels versus the others' evaluation for the same land parcel. Panel (b) shows the correlation between the farmer's and the others' evaluation of the need for fertilizer usage.

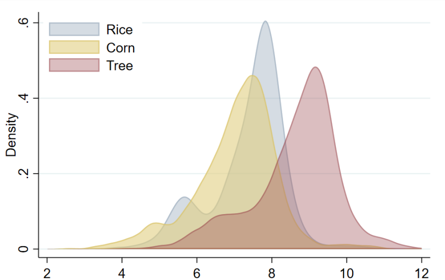
Land quality	(1)	(2)	(3)	(4)
Bulk	-0.0074	-0.0082	-0.0085	-0.0087
	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Carbon	0.1175	0.1807	0.1916	0.1935
	(0.0285)	(0.0308)	(0.0308)	(0.0307)
Slope	-0.0483	-0.0365	-0.0345	-0.0348
	(0.0089)	(0.0096)	(0.0096)	(0.0095)
pH (acidity)		-0.7080	-0.6406	-0.5746
		(0.3340)	(0.3327)	(0.3312)
OMts (organic content)		1.0029	0.9722	1.0253
		(0.4509)	(0.4486)	(0.4465)
Nts (nitrogen)		-0.3039	-0.3406	-0.4299
		(0.2974)	(0.2959)	(0.2948)
P_2O_5dt (phosphate)		-0.1389	-0.1428	-0.1591
		(0.0918)	(0.0914)	(0.0910)
K_2Odt (potash)		0.1477	0.1516	0.1381
		(0.0983)	(0.0978)	(0.0973)
Flood			-0.0127	0.0005
			(0.0609)	(0.0606)
Drought			-0.1211	-0.1076
			(0.0691)	(0.0689)
Water pollution			0.3140	0.3288
			(0.0591)	(0.0592)
Links				0.0700
				(0.0180)
Trust				0.0671
				(0.0174)
Interaction				0.0018

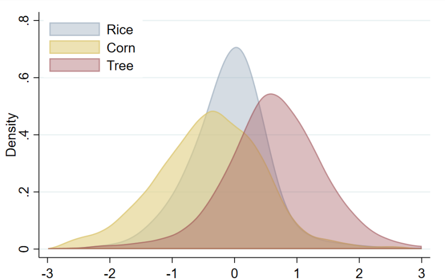


(a) Agricultural productivity gap

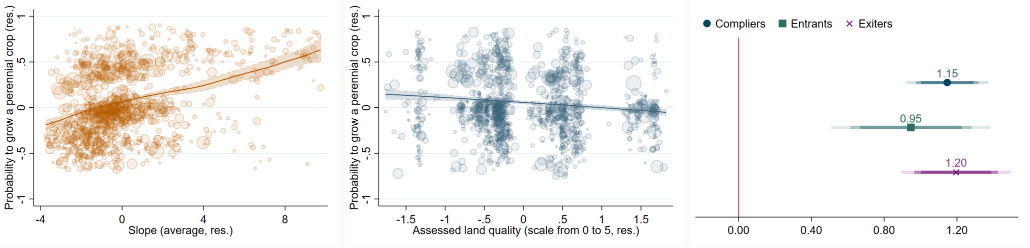


(b) Adoption of tree crops





A crop productivity gap: selection.

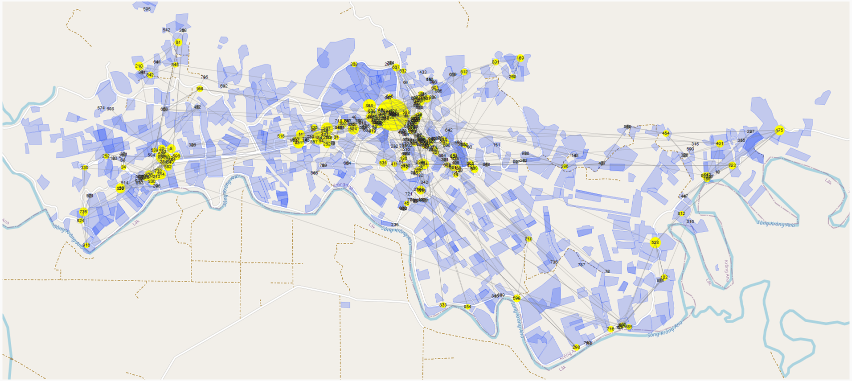


(a) Slope (b) Land quality (c) Entry/exit

Notes: Panel (a) (resp. b) shows the correlation between the likelihood to grow a perennial crop and average slope (res. assessed land quality). The different measures are residualized by soil chemical characteristics, and beliefs about climate change, network connections and village fixed-effects (in both panels). Panel (c) shows the tree premium in (log) agricultural TFP for: compliers (land parcels with trees in 2019 and 2022), exiters and entrants. Note that exiters and entrants include changes in land use over the same land parcel or changes in land ownership.

High-value crops are grown on "lower-quality" land parcels (!):

+ land quality ($-$, $R^2 = 0.10$),
+ slope ($+$, $R^2 = 0.32$),
+ chemical properties ($-$ for organic, $+$ for nit./potash, $R^2 = 0.40$).

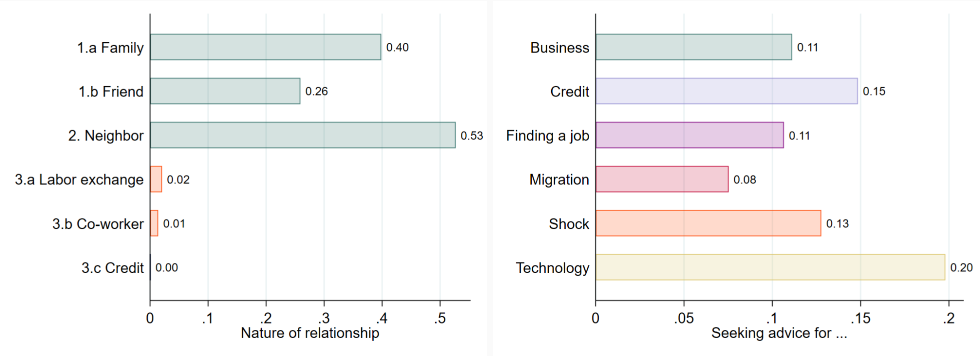


Notes: The figure displays the location of houses (yellow circles) and land parcels (blue polygons) in village 1. The size of circles indicate the number of times the members of a given household is mentioned as a contact by another respondent, and the arrows illustrate the spatial distributions of these links.

Social links form spatial clusters:

(i) households of a same extended family are more likely to live nearby,
(ii) neighbors are more likely to exchange labor,
(iii) spatial proximity reduces the communication costs.

The structure of networks: motivations, origins and strength of links Back.



(a) Origins

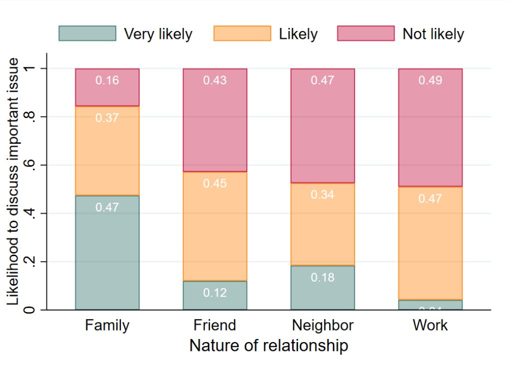
(b) Advice

Notes: Panel (a) shows the share of links referred to as: family, friends, neighbors, etc. Panel (b) shows the share of links used for advice.

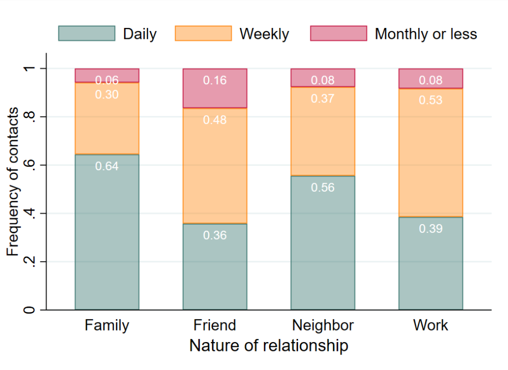
Structure of social networks Back:

Network statistics	Village 1	Village 2	Village 3	Village 4
Degree	5.117	4.808	5.549	5.023
	3.305	*3.005*	*3.132*	*3.143*
	[40]	[33]	[32]	[35]
Betweenness	0.009	0.013	0.011	0.012
	0.015	*0.027*	*0.017*	*0.022*
	[0.225]	[0.349]	[0.211]	[0.298]
Sub-networks	8	4	2	3
Large sub-networks	3	1	1	1
Observations	324	193	213	215

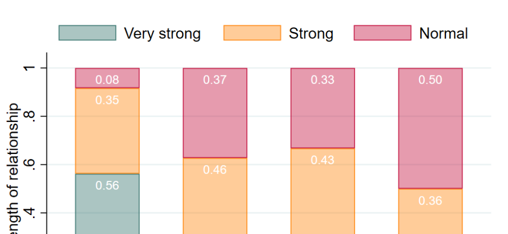
Notes: A unit of observation is a household in 2022. These statistics are computed within the undirected network generated through all recorded contacts between households of a same village. For each undirected network (corresponding to a village), we report the following statistics: the average number of edges for each node (their average *degree*); the heterogeneous centrality of nodes (the average and standard deviation of the *betweenness* centrality measure); and the number of closed sub-graphs (the total number of *sub-networks*, and the number of *large sub-networks* with more than 10 nodes). The betweenness centrality of a node is the number of shortest paths drawn between any two pairs of villagers that passes through the node. In other words, a high betweenness indicates that the node is an instrumental link between many pairs of villagers.

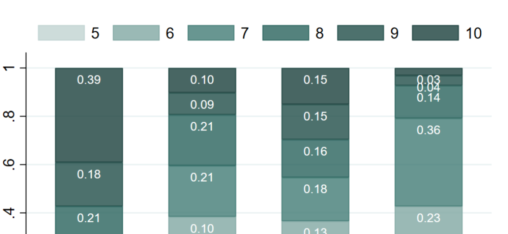


(a) Discuss subjects

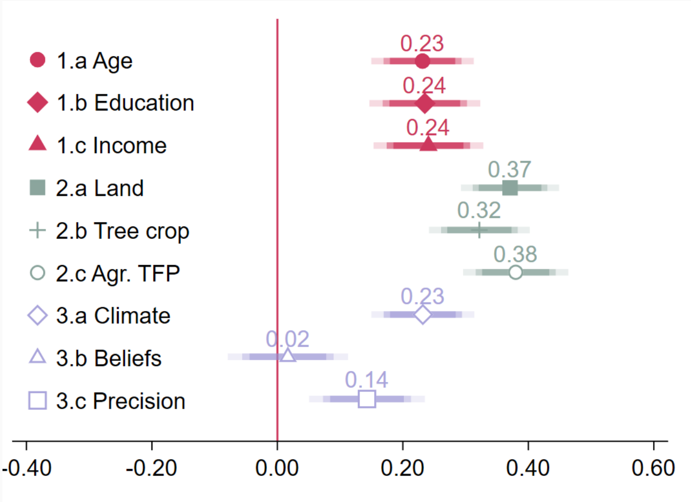


(b) Frequency





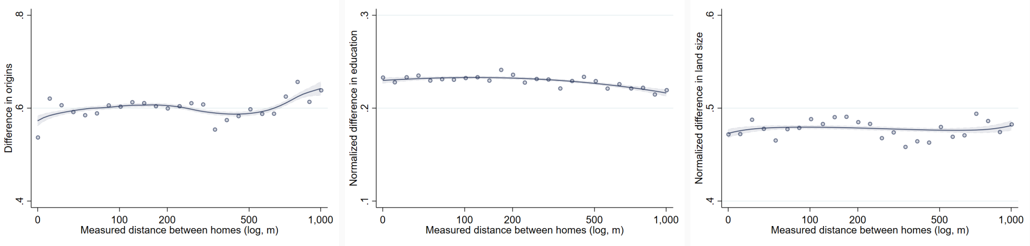
The network formation: homophily.



Notes: This Figure shows the correlation between an edge and the connected edges weighted by the link “proximity”, and for a set of selected variables.

Social linkages might have limited value in terms of (novel) information transmission

Home proximity and homophily:



(a) Origins (b) Education (c) Land size

Notes: Panel (a) shows the correlation between proximity in origins (same province) and distance between homes across all pairs of villagers. Panels (b) and (c) show the correlations between proximity in household characteristics (education of head, land holdings) and distance between homes across all pairs of villagers. The estimated coefficients are respectively: 0.007 [0.002] (panel a); -0.004 [0.001] (panel b); -0.001 [0.001] (panel c).

Back

Adoption	(1)	(2)	(3)
Exposure (ϑ_i^1)	0.097	0.113	0.113
	(0.065)	(0.061)	(0.060)
Controls (instrument)	Yes	Yes	Yes
Controls (soil)	No	Yes	Yes
Controls (network)	No	No	Yes
Observations	2,198	2,198	2,198

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include sub-network fixed effects. The explaining variable is the standardized exposure to the treatment; the instrument is the standardized, predicted exposure to the treatment—as predicted by proximity between homes. In both cases, the exposures are computed using the allocation of treatment in 2019. The set of (instrument) controls include: the previous status of the parcel in 2019 (treated or not), the number of households in immediate proximity, the average (absolute) altitude differential with other homes in the village, the density of parcels with high-return perennial crops around the various parcels owned by the household, and the density of parcels around the various parcels owned by the household. The set of (soil) controls include: parcel characteristics (area, bulk density, organic carbon content, elevation, slope, distance to the homestead), the latitude, longitude and altitude of the home location. The set of (network) controls include the number of first-order linkages, of second-order linkages, and indicators of network centrality (betweenness, closeness, eigenvector centrality, clustering), and sub-network fixed effects. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

Back

The return to social network—a pseudo-panel approach:

Adoption ($y_{pin} - y_{pin-1}$)	(1)	(2)	(3)
Exposure ($\vartheta^0_{in} - \vartheta^0_{in-1}$)	0.014		
	(0.004)		
First-order exposure ($\vartheta^1_{in} - \vartheta^1_{in-1}$)		0.051	
		(0.022)	
Second-order exposure ($\vartheta^2_{in} - \vartheta^2_{in-1}$)			0.029
			(0.019)
Observations	23,569	20,436	7,306
F-stat	385.97	12.52	15.79

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include year fixed effects. The set of (instrument) controls include: the previous status of the parcel in the previous period (treated or not), the number of households in immediate proximity, the average (absolute) altitude differential with other homes in the village, the density of parcels with high-return perennial crops around the various parcels owned by the household, and the density of parcels around the various parcels owned by the household. The set of (soil) controls include: parcel characteristics (area, bulk density, organic carbon content, elevation, slope, distance to the homestead), the latitude, longitude and altitude of the home location. The set of (network) controls include the number of first-order linkages, of second-order linkages, and indicators of network centrality (betweenness, closeness, eigenvector centrality, clustering), and sub-network fixed effects. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

Back

The return to social network—possible mechanisms:

	Quality	Suitability	Input	Climate	Consideration
Exposure (ϑ_i^1)	0.145	0.212	0.015	0.061	0.244
	(0.198)	(0.073)	(0.059)	(0.097)	(0.099)
Observations	2,203	2,203	2,203	2,203	2,203
F-stat	16.31	16.31	16.31	16.31	16.31

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include sub-network fixed effects. The dependent variable is: a subjective evaluation of land quality (scale from 0, unsuitable, to 5) in column (1); a subjective evaluation of land suitability to grow one of the high-return perennial crop in column (2); an index of subjective evaluations of land needs from 0 to 1 (fertilizers, pesticides, and herbicides) in column (3); an index of subjective evaluations of climate risk from 0 to 1 (flood, drought, and water pollution) in column (4); and whether the farmer consider growing one of the high-return perennial crop in column (5). The explaining variable is the standardized exposure to the treatment; the instrument is the standardized, predicted exposure to the treatment—as predicted by proximity between homes. In both cases, the exposures are computed using the allocation of treatment in 2019. The set of controls is similar to that of column 3 of Table 2. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

Back

The return to social network—OLS specification:

Adoption	(1)	(2)	(3)
Exposure (ϑ_i^1)	0.016	0.014	0.015
	(0.008)	(0.007)	(0.007)
Controls (instrument)	Yes	Yes	Yes
Controls (soil)	No	Yes	Yes
Controls (network)	No	No	Yes
Observations	2,222	2,222	2,222

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include sub-network fixed effects. The explaining variable is the standardized exposure to the treatment computed using the allocation of treatment in 2019. The set of (instrument) controls include: the previous status of the parcel in 2019 (treated or not), the number of households in immediate proximity, the average (absolute) altitude differential with other homes in the village, the density of parcels with high-return perennial crops around the various parcels owned by the household, and the density of parcels around the various parcels owned by the household. The set of (soil) controls include: parcel characteristics (area, bulk density, organic carbon content, elevation, slope, distance to the homestead), the latitude, longitude and altitude of the home location. The set of (network) controls include the number of first-order linkages, of second-order linkages, and indicators of network centrality (betweenness, closeness, eigenvector centrality, clustering), and sub-network fixed effects. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

Back

A pseudo-panel approach—periods of interest and tenure in the village:

Adoption ($y_{pin} - y_{pin-1}$)	(1)	(2)	(3)
First-order exposure ($\vartheta^1_{in} - \vartheta^1_{in-1}$)	0.043	0.050	0.032
	(0.026)	(0.028)	(0.023)
First-order exposure $\times$ shorter tenure			0.060
			(0.063)
Observations	6,945	13,491	20,436
Sample	1980–2006	2006–2022	All

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include year fixed effects. The set of (instrument) controls include: the previous status of the parcel in the previous period (treated or not), the number of households in immediate proximity, the average (absolute) altitude differential with other homes in the village, the density of parcels with high-return perennial crops around the various parcels owned by the household, and the density of parcels around the various parcels owned by the household. The set of (soil) controls include: parcel characteristics (area, bulk density, organic carbon content, elevation, slope, distance to the homestead), the latitude, longitude and altitude of the home location. The set of (network) controls include the number of first-order linkages, of second-order linkages, and indicators of network centrality (betweenness, closeness, eigenvector centrality, clustering), and sub-network fixed effects. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

Back

The return to social network—alternative exposures and treatments:

	(1)	(2)	(3)
Exposure (ϑ_i^1)	0.060	0.133	0.112
	(0.029)	(0.088)	(0.055)
Observations	2,203	2,203	2,203
F-stat	57.60	7.12	14.62
Exposure	Incl. family	-	Treatment 2022
Instrument	-	Density	Treatment 2022

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include sub-network fixed effects. In the baseline specification, the explaining variable was the standardized exposure to the treatment; the instrument is the standardized, predicted exposure to the treatment—as predicted by proximity between homes. In both cases, the exposures were computed using the allocation of treatment in 2019. The alternative specifications are: the baseline instrument but an exposure measure which includes family links in column (1); the baseline exposure but an instrument which computes an inverse-distance weighted measure of exposure to the treatment (against the average treatment among neighbors between 0 and 100 meters in the baseline) in column (2); and both exposure and instrument calculated using the allocation of treatment in 2022 in column (3). The set of controls is similar to that of column 3 of Table 2. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

Back

The return to social network—additional controls:

	(1)	(2)	(3)	(4)
Exposure (ϑ_i^1)	0.103	0.111	0.108	0.110
	(0.052)	(0.054)	(0.052)	(0.051)
Observations	2,203	2,203	2,203	2,203
F-stat	16.14	14.93	16.26	17.28
Additional controls	Land tenure	Soil comp.	Land quality	Demographics

Notes: A unit of observation is a land parcel in 2022. Standard errors are reported between parentheses and clustered at the household level. All specifications include sub-network fixed effects. The additional controls are: dummies for each type of land tenure in column (1); soil composition as inferred from 300 soil testing samples in column (2); a subjective evaluation of land quality (scale from 0, unsuitable, to 5) in column (3); and household characteristics (age, gender of the head, number of dependents) in column (4). The explaining variable is the standardized exposure to the treatment; the instrument is the standardized, predicted exposure to the treatment—as predicted by proximity between homes. In both cases, the exposures are computed using the allocation of treatment in 2019. The set of standard controls is similar to that of column 3 of Table 2. The sample is restricted to agricultural parcels for which we are confident about their geolocation (i: observed in both waves, ii: with similar geolocation and area across waves, iii: where the household is not unsure when locating the parcel or drawing its borders).

Back

A pseudo-panel approach—timing of adoption ◂ Back:

Adoption ($y_{pin} - y_{pin-1}$)	(1)	(2)	(3)
Exposure (F, $\vartheta^0_{in} - \vartheta^0_{in-1}$)	0.005		
	(0.005)		
Exposure ($\vartheta^0_{in} - \vartheta^0_{in-1}$)	0.012		
	(0.004)		
Exposure (L, $\vartheta^0_{in} - \vartheta^0_{in-1}$)	0.003		
	(0.004)		
First-order exposure (F, $\vartheta^1_{in} - \vartheta^1_{in-1}$)		0.021	
		(0.023)	
First-order exposure ($\vartheta^1_{in} - \vartheta^1_{in-1}$)		0.034	
		(0.018)	
First-order exposure (L, $\vartheta^1_{in} - \vartheta^1_{in-1}$)		0.020	
		(0.021)	
Second-order exposure (F, $\vartheta^2_{in} - \vartheta^2_{in-1}$)			-.140
			(0.329)
Second-order exposure ($\vartheta^2_{in} - \vartheta^2_{in-1}$)			0.083
			(0.176)
Second-order exposure (L, $\vartheta^2_{in} - \vartheta^2_{in-1}$)			0.175
			(0.287)
Observations	19,090	15,944	5,574