Barcelona School of Economics

WORKING PAPERS

The Pick of the Crop: Agricultural Practices and Clustered Networks in Village Economies

BSE Working Paper 1426| January 2024

André Groeger, Yanos Zylberberg

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

Andre Groeger Yanos Zylberberg

January 31, 2024

Abstract

This paper studies how social networks (might fail to) shape agricultural practices. We exploit (i) a unique census of agricultural production nested within delineated land parcels and (ii) comprehensive social network data within four repopulated villages of rural Vietnam. In a first step, we extract exogenous variation in network formation from home locations within the few streets that compose each village (populated through staggered population resettlement), and we estimate the return to social links in the adoption of highly-productive crops. We find a large network multiplier, in apparent contradiction with low adoption rates. In a second step, we study the structure of network formation to explain this puzzle: social networks display large homophily, and valuable links between heterogeneous households are rare. Due to the clustered nature of networks and the dynamic, endogenous propagation of agricultural practices, there are decreasing returns to social links, and policies targeting "inbetweeners" are most able to mitigate this issue.

*Groeger: Universitat Autonoma de Barcelona, Barcelona School of Economics, andre.groger@uab.es; Zylberberg: University of Bristol, CEPR, email: yanos.zylberberg@bristol.ac.uk. We would like to thank Aranya Chakraborty, Giacomo De Giorgi, Sanjeev Goyal, Simon Hess, conference participants at the meetings of the European Economic Association, the Urban Economics Association European Meeting, the German Development Economics Conference, and seminar participants at Bristol, Hannover, and NYU Abu Dhabi for very useful comments and suggestions. Groeger acknowledges financial support from the Spanish Ministry of Science and Innovation through the Severo Ochoa Programme (CEX2019-000915-S) funded by MCIN/AEI/10.13039/501100011033, and through grant PID2021-122605NB-I00 and RYC2021-033208-I as well as from the Generalitat de Catalunya (2021-SGR-00571). Zylberberg acknowledges funding by the Natural Environment Research Council (NERC), the National Foundation for Science and Technology Development (NAFOSTED) and the British Academy through the Newton Hydrometeorological Hazards (NE/S003061/1), and the Humanities and Social Sciences Tackling Global Challenges programme (TGC/200149). We thank Pham Khanh Nam and Truong Dang Thuy for supervising the survey and sharing the household data, and we thank Laurence Hawker and Jeffrey Neal for sharing flood and elevation data. The usual disclaimer applies.

Agricultural productivity is low in developing economies (see, e.g., Gollin et al., 2014a,b), and there is disparity in (measured) yield and agricultural practices across farms (see, e.g., Restuccia and Santaeulalia-Llopis, 2017; Gollin and Udry, 2021; Adamopoulos et al., 2022).[1] In Vietnam, for instance, there is a large *crop productivity gap* observed across farms with different portfolios of agricultural commodities. Some crops—coffee, cashew nuts, rubber or pepper, which are typically grown on low-quality, rugged land—are more than twice as productive as traditional, staple crops—rice, maize, cassava—, and this difference is not explained by inputs (land, soil quality, capital, labor, irrigation, fertilizers, herbicides, etc.) or by the general skills of farm managers. In spite of this "premium" and extensive public efforts to promote investment in high-productivity crops, adoption remains limited: only a minority of households have a land parcel on which they grow such tree crops. This low incidence could reflect technological barriers to adoption (e.g., skills), frictions in allocating production factors (credit constraints, frictional land markets), or imperfect information transmission from adopters to others within villages. The present study focuses on the latter: what is the role of local networks in explaining the (relative lack of) crop adoption?

This paper identifies how the *structure* of social networks affects the dynamic adoption of high-return agricultural practices.[2] The novelty hinges on observing both agricultural production within delineated land parcels and the comprehensive social network of repopulated villages in rural Vietnam.[3] The delineation of land and geo-location of homes allow us to account for all geographic differences across farmers (e.g., soil quality, ruggedness, flood risk) and for their relative proximity whether at home or on their agricultural plots. The network data allows us to reconstruct a full, dynamic social structure, including indirect linkages of higher order. The nature of village formation—through staggered population resettlement and long-distance migration—allows us to isolate exogenous variation in network formation: the exact timing of arrival across settlers strongly predicts home locations within the few streets that compose each village; and home proximity (within 100 meters) increases the likelihood to form a link. Our em-

[1]The literature has studied under-mechanization, input quality, selection induced by rural-urban migration, or returns to scale as possible explanations for the low productivity of (most) farms. Our focus is on agricultural practices in general and crop choice in particular.

[2]Learning through networks has been shown to play an important role in the adoption of new technology by farmers of rural economies (see, e.g., Foster and Rosenzweig, 1995; Bandiera and Rasul, 2006; Conley and Udry, 2010; Duflo et al., 2011; Suri, 2011; Kala, 2017; Beaman and Dillon, 2018; Banerjee et al., 2019; de Janvry et al., 2022).

[3]Our survey is a census of four rural villages in the Central Highlands of Vietnam conducted in 2019 and in 2022 (with a panel of about 950 households and 2,700 land parcels). The survey covers: living standard measurement survey questions; the detailed geography of agricultural production through a specific module based on satellite imagery and cadastral maps; soil testing; elevation and flood risk modeling; and a module recording the whole network of family and friends. Retrospective questions are asked for agricultural production, settling within the village, and network formation.

pirical analysis uncovers an apparent contradiction: we identify a large network multiplier, and yet, the overall adoption of high-return crops remains relatively low. We show that the structure of network formation explains this puzzle: there is large homophily between connected households; adopters (resp. non-adopters) are frequently tied with other adopters (resp. non-adopters); and valuable links between heterogeneous households are rare. Importantly, we show that this homophily is the endogenous outcome of the dynamic propagation of agricultural practices through clustered networks, and policies targeting the "in-betweeners"—villagers connecting the different clusters—would be efficient.

Studying the relationship between social networks and agricultural production is challenging. A first important empirical challenge is that network links are not exogenous to agricultural practices, especially so in a context where villagers share knowledge and sometimes labor when they work on contiguous land parcels. In short, the workplace is endogenous to agricultural practices within villages. Our approach leverages the location of homes, conditional on the portfolio of land parcels, to isolate exogenous variation in network formation: these villages were part of a large resettlement program spanning about 30 years, and new settlers would typically be given or claim a land plot for residential purposes upon arrival (Evans, 1992; Hardy, 2005; Van de Walle and Gunewardena, 2001). Accordingly, the geography of residential settlement strongly reflects the timing of arrival to the Central Highlands. We exploit these weaker links—induced by the location of residence within the few streets that compose each village—to identify a causal effect of networks on agricultural practices. More specifically, we instrument a measure of *direct* exposure to tree-growing farmers based on actual network links by a residential, distance-weighted measure of exposure (living close to tree-growing farmers) while controlling for the workplace location (working close to tree-growing farmers) and various measures of connectedness within the network and within the village.[4] We also consider a measure of *indirect* exposure to tree-growing farmers based on second-order linkages, and we combine two distinct sources of network formation—shared origins and the location of homes—to isolate exogenous variation in such *indirect* exposure. We find that the local network predicts the adoption of new crops: one standard deviation in exposure increases the likelihood to adopt highly-productive crops on a given land parcel by 0.10 within a 3-year period. This is a large multiplier, e.g., to be compared with the initial incidence of 0.18 across suitable land parcels.[5] This finding is puzzling: How do

[4]We exclude family links from all our calculations.

[5]In theory, the network multiplier could result from information transmission, an easier access to resources (e.g., factors or intermediary inputs), or informal insurance/credit allowing connected farmers to incur a costly investment (even though recent research has shown that connections also transmit shocks through the network, see Kinnan et al., 2024). In support of the first channel, we document that farmers

we reconcile such a large social multiplier with low adoption rates?

The second empirical challenge consists in understanding the quantitative impact of a causal multiplier—identified using weaker, rarer links—on the endogenous dynamics of technology adoption. The limited diffusion of high-productivity agricultural practices is explained by the network structure through (i) a static argument and (ii) a dynamic argument. First, we show that a large network multiplier may coexist with a relatively low, but non-negligible, adoption rate when social networks display large homophily (A and B are quite similar in agricultural practices when they are connected). Indeed, network connections are primarily formed by family links or by work practices, and there is a high degree of resemblance between linked households. A node that is connected to treated nodes is likely to be treated already; reciprocally, a non-treated node is unlikely to be exposed to the treatment through the network.[6] In short, the random, weaker links that we use for identification purposes are not so common. For instance, in counterfactual simulations we show that randomizing network formation for households arriving in our villages from 1980 onwards would have increased crop adoption by 50%. Second, even when the *initial* allocation of treatment is random, a clustering structure implies that the treatment will primarily diffuse within localized clusters, thereby inducing (endogenously and dynamically) high degrees of homophily and lower returns to the social network. To shed light on this decreasing return to social multipliers, we consider projections randomizing the initial allocation of treatment—the distribution of agricultural practices—and/or the network structure—the social links across nodes. Randomizing social links or providing an initially more dispersed treatment would generate a higher adoption rate after 50 years. These experiments however markedly differ in their dynamic impact: a random network does not generate decreasing returns over time, while a randomized treatment would generate homophily and decreasing returns in the longer run, through an endogenous propagation within clustered networks. We conclude the analysis by discussing targeted, simple policies. Efficient policies allocate treatment to "inbetweeners", who connect different clusters of farmers, ensuring that the returns to social links do not decrease with time.

Our contribution is to relate the structure of social networks to the dynamics of crop adoption in a non-experimental setting. Our context has unique features in that we can observe entire, closed social networks. The resulting insight is novel: network links are

exposed to the treatment have higher priors about the suitability of their land to growing high-return crops, even controlling for objective indicators of soil quality or inferred land quality from the evaluation of other farmers. We do not find a very high reliance on informal insurance or informal borrowing in our context, most loans originating from rural development banks.

[6]For instance, we find that the correlation between agricultural productivity, adoption of tree crops, or land area is about 0.30 between two nodes of a network link versus 0.03 within a given village.

potentially effective at fostering the adoption of highly-productive crops; the high degree of homophily within social networks however limits their impact at any point in time; and the clustered nature of social networks implies that such homophily endogenously arises over time. Policies targeted at "in-between" villagers might alleviate this clustering of agricultural practices and accelerate their diffusion in the shorter run—as would interventions fostering the formation of random, weak network links.

Our work relates to the large literature discussing technology adoption and learning through networks (Griliches, 1957; Bandiera and Rasul, 2006; Duflo et al., 2008; 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), and more specifically, to the research discussing the structure of such networks (clustering and homophily, see, e.g., Calvó-Armengol et al., 2009; Acemoglu et al., 2011; Ferrali et al., 2020; Jackson et al., 2023), seed targeting (e.g., Akbarpour et al., 2023; Sadler, 2023), or influence maximization (e.g., Banerjee et al., 2013; Kim et al., 2015; Cai et al., 2015; Banerjee et al., 2019; Beaman et al., 2021). As in Banerjee et al. (2013), Banerjee et al. (2019), and Beaman et al. (2021), we find that centrality measures are important to target seeds which could increase propagation. However, we find betweenness centrality to perform (slightly) better, possibly due to the high level of clustering within our network.[7] We relate to Foster and Rosenzweig (1995) and Conley and Udry (2010) in that spatial proximity is instrumental to our identification of network links, but our empirical strategy also exploits overlaps across farmers of different origins (Barnes et al., 2016).

One advantage of our setting is to observe full social networks within a closed environment, i.e., four isolated villages in the Central Highlands of Vietnam, and to exploit high-quality information on both sides of a social link (in contrast with the influential research proposed in Banerjee et al., 2019; Breza et al., 2020, with many poorly-observed networks "without network data"). To our knowledge, there exist only a few papers that draw on complete network data and can relate heterogeneous network characteristics to the rate of information diffusion or technology adoption (Banerjee et al., 2013; Cai et al., 2015; BenYishay and Mobarak, 2019; BenYishay et al., 2020; Beaman et al., 2021;

[7] Akbarpour et al. (2023) discuss the complex issue of influence maximization and show that optimal seeding is not performing much better than random seeding (with more seeds) for a large set of propagation mechanisms. The basic intuition is that, with standard propagation models, random seeding will end up reaching influential nodes. In practice, treatment adoption might require several, independent information sources, thus departing from these standard propagation models and making in-between individuals more influential. Our empirical setting does not allow us to properly characterize the propagation of agricultural practices (e.g., we do not have exogenous variation predicting the number or the nature of treated friends), and thus test if the theoretical results of Akbarpour et al. (2023) would hold or not. Along the same lines, we do not have exogenous variation predicting the characteristics of friends, such that we cannot really discuss seed selection through observable characteristics (Sadler, 2023).

Chakraborty, 2022; Bandiera et al., 2023). We finally relate to the literature discussing challenges in estimating peer effects through networks (Manski, 1993; Bramoullé et al., 2009; De Giorgi et al., 2010; Bramoullé et al., 2020; Jochmans, 2023) and challenges induced by endogenous network formation (Graham, 2017; De Paula, 2020). Our identification comes from network shocks rather than an exogenous allocation of treatment; in that respect, we closely relate to Burlig and Stevens (2023), who exploit aggregate network shocks—the merge of churches in the United States—to study technology adoption among newly-formed links.

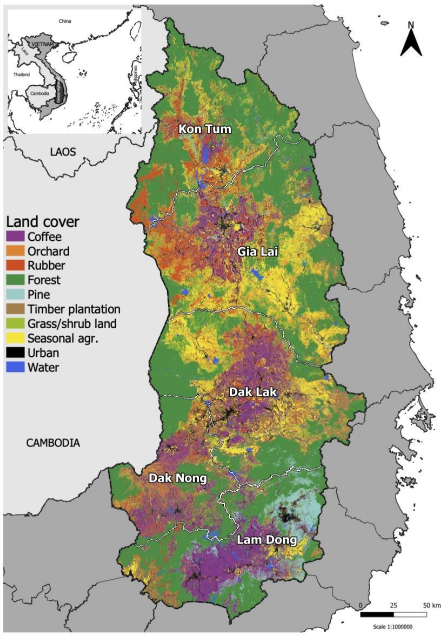
Our work also contributes to the large, nascent literature discussing agricultural productivity in developing countries (Udry, 1996; Restuccia and Rogerson, 2008; Gollin et al., 2014a,b; Chen, 2017; Restuccia and Santaeulalia-Llopis, 2017; Gollin and Udry, 2021; Adamopoulos et al., 2022). We highlight the role of crop choice in explaining low agricultural productivity, but also its dispersion across farmers of a same village. One explanation behind the dispersion of agricultural productivity could be frictional factor markets, credit as documented in numerous contexts, but also land (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) or frictions to mobility and (selection into) rural-urban migration (Lagakos et al., 2018; Adamopoulos et al., 2022). Our focus is instead on (imperfect) information transmission.

The remainder of the paper is organized as follows. Section 1 discusses agricultural practices and resettlement patterns in the Central Highlands of Vietnam. Section 2 describes our data sources, the productivity gap across agricultural commodities, and the structure of local networks. Section 3 details our main empirical strategy. Section 4 establishes the baseline result and discusses robustness checks. Section 5 rationalizes the co-existence of low adoption rates with a high social multiplier and discusses the role of clustered networks and targeted policies. Finally, Section 6 briefly concludes.

1 Agricultural production in rural Vietnam

Agricultural practices in the Central Highlands of Vietnam Many developing countries, including Vietnam, experience a rapid transformation of their economic activity with large migration flows from rural areas to urban agglomerations. Along this process, the nature of rural economic activity is expected to change with an intensification and mechanization of agriculture. The focus of our study is about these important changes in agricultural practices with: higher agricultural investment (e.g., irrigation, machines, or the use of fertilizers, pesticides, and herbicides); but mostly the adoption of high-risk/high-yield perennial crops (see, e.g., Benjamin et al., 2018). This transfor-

Figure 1. Crop diversity in the Central Highlands of Vietnam.



(a) Crop diversity

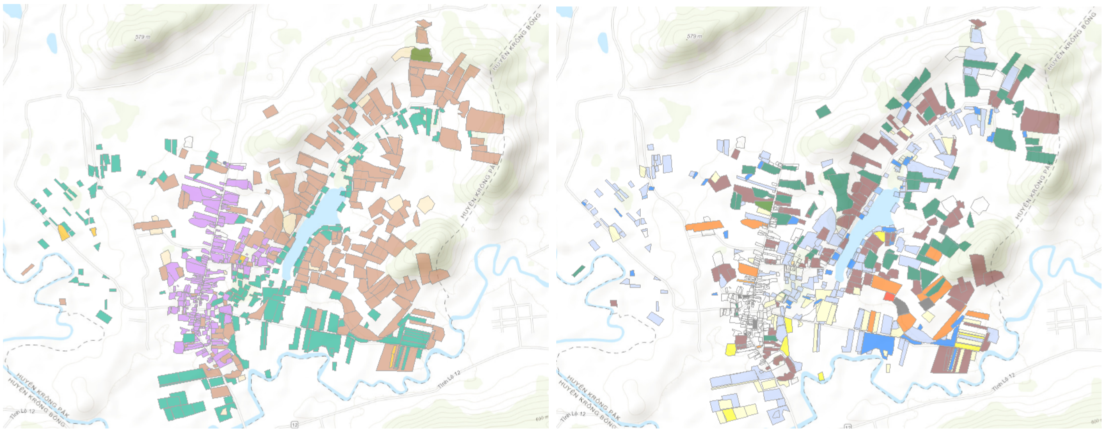
(b) Buon Ma Thuot Coffee festival

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.

mation is apparent from the wide variety of crops grown in the Central Highlands of Vietnam (see panel a of Figure 1). One reason for such changes is that agricultural investment in high-return crops for export purposes is actively promoted by policy makers at the local and national levels. The Buon Ma Thuot Coffee festival is an example of such efforts (see panel b)—cashew nuts, pepper, or flowers also have local, annual festivals.

This diversity in cropping patterns is visible within villages: agricultural households of the Central Highlands of Vietnam hold numerous, small, and geographically dispersed land parcels and they grow different crops on each of these parcels. For instance, the 950 households of our sample grow the following crops: mung bean, cassava, rubber, coffee, rice, cashew nuts, fodder maize, corn, sweet potato, fragrant rice, pepper, sugarcane, or kale. We illustrate the local dispersion of land usage and agricultural practices in Figure 2 where we display land use in panel (a) and crop type in panel (b) for one of our villages—named Village 3 to ensure the anonymity of surveyed households. First, one can see that there is spatial clustering in land use: a part of the village is the residential area (in purple); a flat area is typically used for annual agriculture (often wetlands); and rugged terrain is typically used for perennial crops. Second, there remains large diversity in cropping patterns, even at a disaggregated level: households grow rice (different shades of blue corresponding to different varieties), coffee (brown), cashew nuts (green), pepper (gray), rubber (dark orange), or maize (yellow). Finally, note that the diversification of crops is also visible within households and that a small fraction of land is used for pasture or aquaculture.

Figure 2. Crop diversity (and land use) within villages.



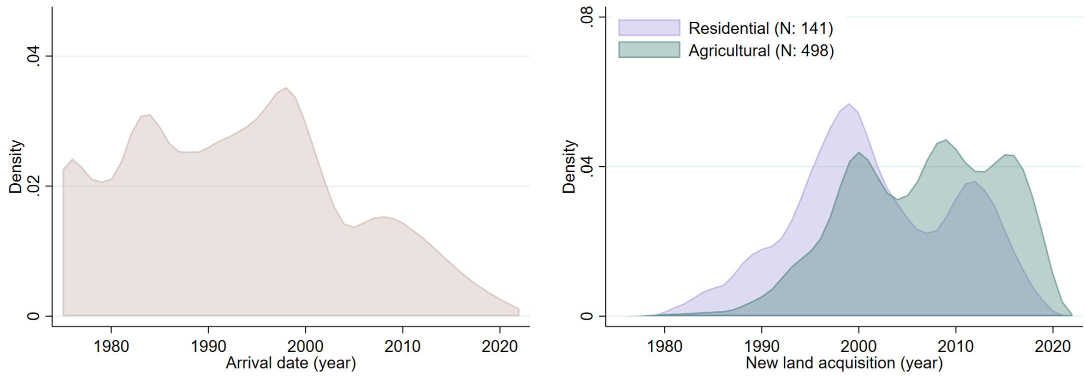
(a) Land use (b) Crops

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).

Population resettlement and land markets in Vietnam The Central Highlands of Vietnam were populated by ethnic minorities before 1954. Between 1954 and 1991, successive, centralized programs were implemented to relocate ethnic Vietnamese, mostly from the North of Vietnam. The migration waves did not stop in 1991, but slightly changed in nature: they became more decentralized, with settlers attracted by the opportunity to grow cash crops and by the "available" land. Some assistance was granted to migrants settling in through formal programs; others had to clear forest or use land supposed to be confined to public use. Our villages were typically formed through successive, individual migration spells of Northern families: only 2% of (intergenerational) households were already present before 1975; settlers arrived gradually between 1976 and 2012 (as shown in Figure 3) and mostly from the provinces of Ha Tinh (23%), Quang Nam (18%), Lang Son (11%), Hung Yen (10%), Cao Bang (9%), and Thai Binh (8%). We will see in Section 2 that the geography of settlement within our villages relates to the timing of migration, rather than to the origin of migrant households.

These settlement waves coincided with a transformation of Vietnam from a central-

Figure 3. Settlement and land acquisition in our villages.



(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).

ized economy to a market economy with a liberalization of land markets. The nature of land acquisition within our villages reflects this mix between formal resettlement programs, informal settlements, and later land transactions: About 18% of *residential* land parcels were allocated by the government; about 28% were bought; and the rest was "claimed", mostly, and inherited. Only two thirds of residential land parcels are associated with a formal land use right certificate—a red booklet. The acquisition of new *agricultural* land has been less "exogenous": many land parcels were acquired through formal means in recent times (see panel b of Figure 3).

2 Data sources, descriptive statistics and empirical strategy

This section describes our household survey and provides descriptive statistics about agricultural practices (the "crop productivity gap") and the structure of networks (notably, the nature of social links and their "homophily").

2.1 Data sources

We exploit a panel household survey conducted in 2019 and 2022, both in September, in 4 villages of the Central Highlands. These four villages—that we will label Village 1, Village 2, Village 3, and Village 4 for ethical purposes—are located in the Lak District, Krông Bông District, Krông Pac District, and Ea Súp District within the Dak Lak province. The survey covers about 950 households and 4,000 individuals in each wave with about 95% of households/individuals observed across the two waves. Most of the survey follows the World Bank high-quality standards and is similar to the main household surveys in Vietnam, i.e., Vietnam Household Living Standard Survey (VHLSS) and Thailand Vietnam Socio Economic Panel (TVSEP). An average questionnaire would take 3 hours and would cover about 1,000 questions. The survey however has the following unique features that we describe briefly below and in greater length in Appendix A.1: a comprehensive land geo-location module (with soil testing and subjective evaluations from peers); and a network module.

The research requires the precise geo-location and delimitation of land parcels. The first novelty of the survey is its land module: land plots and assets of each household are precisely geo-located using a novel procedure based on the recognition and drawing of land parcels on satellite images augmented by cadastral boundaries; and production data is then matched at the parcel and crop levels. The crop module of the household survey records the labor inputs and the expenditures related to the different activities along the crop cycle (e.g., preparation, seeding, harvesting). From this detailed account of agricultural activities at the crop/parcel level, we create the following aggregate measures of input at the parcel level and at the household level: the area cultivated for each crop; the irrigated area for each crop; the hours of work (hired labor, family labor, exchange labor) for each crop; the expenditures for each crop.[8] We also create measures of output for the different crops: quantities (kg), yield (kg/acre), sales, crop income (subtracting expenditures and adding personal consumption).

We exploit the geo-location of land parcels to better characterize land quality. More specifically, we complement the previous data about land with: (i) subjective evaluations of land quality by farmers themselves and by their peers (about 130 parcels are randomly assigned for respondents to "evaluate", in the spirit of Galton, 1907), input requirements, and suitability to grow different crops by respondents and by their peers; (ii) prospective agricultural strategies on each land parcel; (iii) soil testing in about 300 parcels for acidity and nutrients (then interpolated over the whole village); (iv) flood risk; and (v) soil bulk density (fine earth), the soil organic carbon content, and the elevation, slope and ruggedness to account for the local topography.

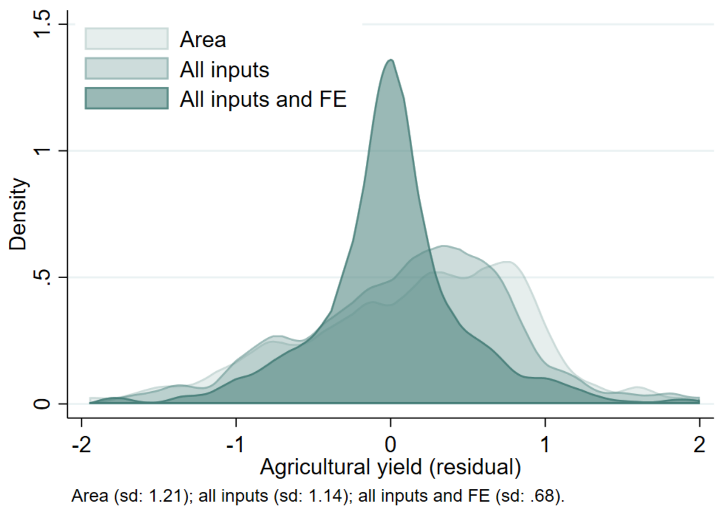
The second novelty is to record all network linkages between households, their possible usage (e.g., the motivation, strength and nature of network links), and their actual usage linked with other survey modules (e.g., previous credit linkages, informal insurance network, participation in labor exchange arrangements, land transactions). We then match the recorded links with their own survey records to observe all survey vari-

[8] The list of crops that we record is the following: areca nut, bamboo, cajeput 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.

ables on both sides of a network "edge"—an advantage of the comprehensive coverage of a few villages rather than the usual household survey sampling based on a few units scattered across space. Another advantage of survey sampling is that we can characterize the whole network of links within the village, and thus interpret missing links as proper zeros. A corollary is that we can construct indirect linkages across households (higher-degree edges: "A knows B who knows C who knows D, etc."), network measures of centrality, and characterize closed sub-networks.

Finally, the survey allows to capture risk mitigation strategies (access to credit, non-agricultural activities, remittances, informal insurance networks, labor exchange), and income/consumption patterns in addition to standard household characteristics. The panel dimension is instrumental, as it allows to capture changes in land usage between the two survey waves, as well as shocks through a specifically designed shock module and separate shock-specific modules (e.g., geo-localized floods and the effect of COVID).

Figure 4. The agricultural productivity gap.



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.

2.2 The productivity gap across agricultural commodities

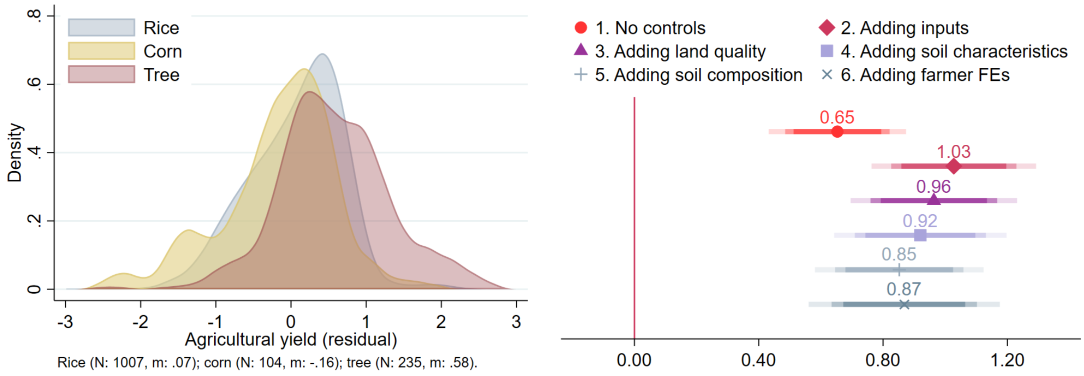
This section documents one source of heterogeneity in agricultural productivity: a crop productivity gap, observed across parcels producing different agricultural commodities.

The agricultural productivity gap Consider a household i growing a certain crop c on a given parcel. Letting y_{ic} denote agricultural revenue and $\mathbf{x}_{ic}$ denote the vector of production factors, i.e., hours provided by family members or casual farm workers, expenditures (e.g., on fertilizers and other intermediate inputs), expenditures on capital, and cultivated land area, we isolate the agricultural productivity gap across households and crops by estimating,

$$\ln y_{ic} = \ln z_{ic} + \ln f(\mathbf{x}_{ic})$$

where f is a Cobb-Douglas production function with constant parameters and the estimation residual, $\ln z_{ic}$, captures a household/crop Total Factor Productivity. We display the distribution of this residual in Figure 4 where we also show the distribution of such residuals when controlling for: area as the only input (lighter green); all inputs; and all inputs *and* farmer fixed-effects (darker green). Figure 4 shows that there is large variation in agricultural Total Factor Productivity, and that this variation decreases when adding farmer fixed-effects but does not disappear. We next study one source of variation underlying this dispersion: the nature of crop c.

Figure 5. The crop productivity gap.



(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.

The crop productivity gap We now decompose the agricultural productivity gap into a between-crop versus within-crop variation,

$$\ln z_{ic} = \mu_c + \varepsilon_{ic}.$$

We find that about 20% of the variance in residual agricultural productivity, $\ln z_{ic}$, is captured by the dispersion in average productivity across crops, $\{\mu_c\}$. We further shed light on these differences in panel (a) of Figure 5 where we show the distribution of residuals, $\ln z_{ic}$, for: premium crops (coffee, pepper, rubber); and rice and maize. The former are between 60% and 100% more productive than the latter.

In panel (b) of Figure 5, we directly estimate the premium for coffee, pepper and rubber (a "tree" premium) by adding a dummy, $\mathbb{1}_{c \in T}$, to the regression,

$$\ln y_{ic} = \beta \mathbb{1}_{c \in T} + \ln f(\mathbf{x}_{ic}) + \gamma \mathbf{w}_{ic} + \varepsilon_{ic},$$

where $\mathbf{x}_{ic}$ denote the vector of production factors, f is a Cobb-Douglas specification and $\mathbf{w}_{ic}$ are controls that we add sequentially: land quality (as evaluated by the household and as inferred through peer evaluations of 130 samples and spatial interpolation); soil characteristics (bulk, carbon content, and topography); soil composition (acidity and nutrients collected for 300 parcels, and inferred through spatial interpolation); and farmer fixed-effects. The premium β ranges from 0.65 to 1.03, which corresponds to a productivity boost of about 91% to 180%. In our preferred specification with all controls and farmer fixed-effects, i.e., controlling for farmer unobserved heterogeneity, the premium is 0.87, which corresponds to a productivity differential of about 138%.[9] Given this premium, a natural question is: why are households primarily using land for other purposes than growing coffee, pepper or rubber?

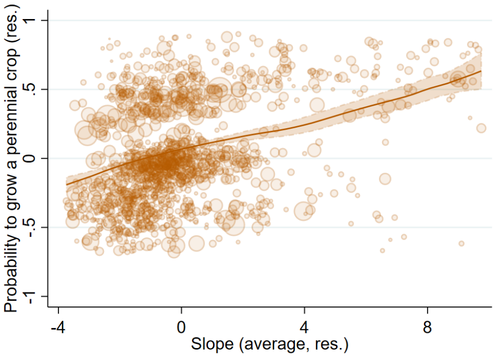
Selection We now discuss selection into the "treatment", $\mathbb{1}_{c \in T}$, or the cultivation of high-productivity crops, and show that: (i) adoption of perennial crops, and of high-productivity crops, is limited; (ii) suitable land parcels are numerous, and not of very high "quality"; and (iii) new entrants into treatment are immediately quite productive. First, only about 18% of agricultural land parcels (and less than 25% of total land area) are used for highly-productive agricultural purposes.

Second, we show in Figure 6 that the likelihood to grow a perennial crop is positively associated with ruggedness and negatively correlated with the respondent's own evaluation of land parcel quality ("How suitable is [your] land parcel for cultivation?", with a scale from 0 to 5).[10] More generally, land quality is negatively correlated with the likelihood to grow a perennial crop and explains about 10% of its variance. Adding the average slope (+) brings the R^2 to 0.32; adding chemical properties (− for organic content, + for nitrate/potash) brings the R^2 to 0.40. One corollary of this negative selection of land parcels into the treatment is that a high fraction of land parcels on which maize, mung bean, sweet potato or cassava are grown would be suitable for the cultivation of highly-productive tree crops.

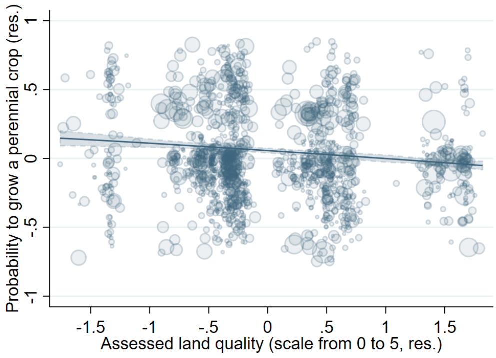
[9]We document similar stylized facts in Appendix A.2 using the household panel provided by TVSEP between 2007–2017.

[10]See Appendix A.2 for a discussion about the determinants of such evaluations.

Figure 6. Selection into the treatment: land quality, entrants and exiters.



(a) Slope



(b) Land quality

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).

Third, we can use the panel dimension to identify new entrants in the cultivation of high-productivity crops, against compliers with a high-productivity crop in both waves. We estimate the tree premium in (log) agricultural Total Factor Productivity across these two categories, and we find that the tree premium is 20% lower and less precisely estimated for new entrants, but the productivity boost that they enjoy compared to other crops is still very substantial (more than twice as productive).

In summary, households appear to under-invest in highly-productive crops. The objective of the paper is to explain the (relative lack) of adoption by imperfect information transmission through local networks. We describe these social networks next.

2.3 The structure of local networks

This section provides statistical evidence about: the structure of networks; the motivations, origins and strength of links; and the degree of homophily across social links.

Structure of social networks The identification of the household network relies on an integrated procedure throughout the questionnaire. An early section would help establish a preliminary list of contacts ("From time to time, most people discuss important matters with other people. Looking back over the last year, who are the people with whom you discussed matters important to you? Who would be the people with whom you would discuss an important matter in the near future?"); the list would then be updated as the interview goes along if new contacts are mentioned by the respondent (when relevant, e.g., villagers within the same labor exchange group, villagers involved in a large transaction with the household, co-workers, hired labor or employers/employees). For each contact, we record their name, age, gender, phone number (last 6 digits), and a description of their relationship with the different household members (formation, frequency, strength, trust, and type of information exchange).

Table 1. Structure of social networks.

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.

We end up with about 4,000 directed links between households,[11] of which 2,900 can be located within our villages. This corresponds to 3 (directed) linked households for each respondent—a small number explained by various factors: our procedure mostly elicits strong links; most villagers are self-employed; and our villages were newly populated from different migration waves and diverse regions. We construct an undirected network within each village by connecting any two households where at least one of the two mentions the other as a link, and we report the following statistics in Table 1: 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.[12] The four villages have a similar average connectedness across nodes and typically have a few important nodes (including the

[11]While we do observe links between individuals, we ignore this dimension in most of our analysis and treat households as the decision unit and any linkages between two individuals from different households as the existence of a link between the two households.

[12]We rely on undirected networks for two main reasons: (i) analyzing propagation throughout a directed network is challenging, especially with a very limited number of linkages; and (ii) we think that

village leader), but these connections are heterogeneously allocated: Village 1, which is more scattered across space, has disconnected sub-graphs, with three of them grouping more than 10 households. The village has fewer nodes with high centrality coefficients, and the village leader is less central than in other villages. By contrast, villages 2 and 4 have slightly fewer connections on average, but they both have influential nodes bridging possible gaps between isolated clusters. Our analysis in Section 5 will discuss the importance of such nodes.

Motivations, origins and strength of links Our villages are isolated communities, usually associated with dense networks and few external contacts. In panel (a) of Figure 7, we show that 40% of contacts are motivated by some family connections, 26% of contacts are considered as friends, 53% are labeled as neighbors, and a negligible proportion of those contacts have a direct work or credit relationship with the household.[13] These relationships are quite tight on average, as shown in panel (b): the typical contact has been known for 15-20 years, partly driven by settlement patterns. Finally, households mostly seek advice from their contacts for technology purposes: about 20% of contacts would be sought to discuss an important matter regarding (agricultural) technology (panel c of Figure 7). We provide a more comprehensive description of social links in Appendix A.2 and show that links are stronger and tighter with family/friends than with neighbors. We will think of the latter as weak(er) links.

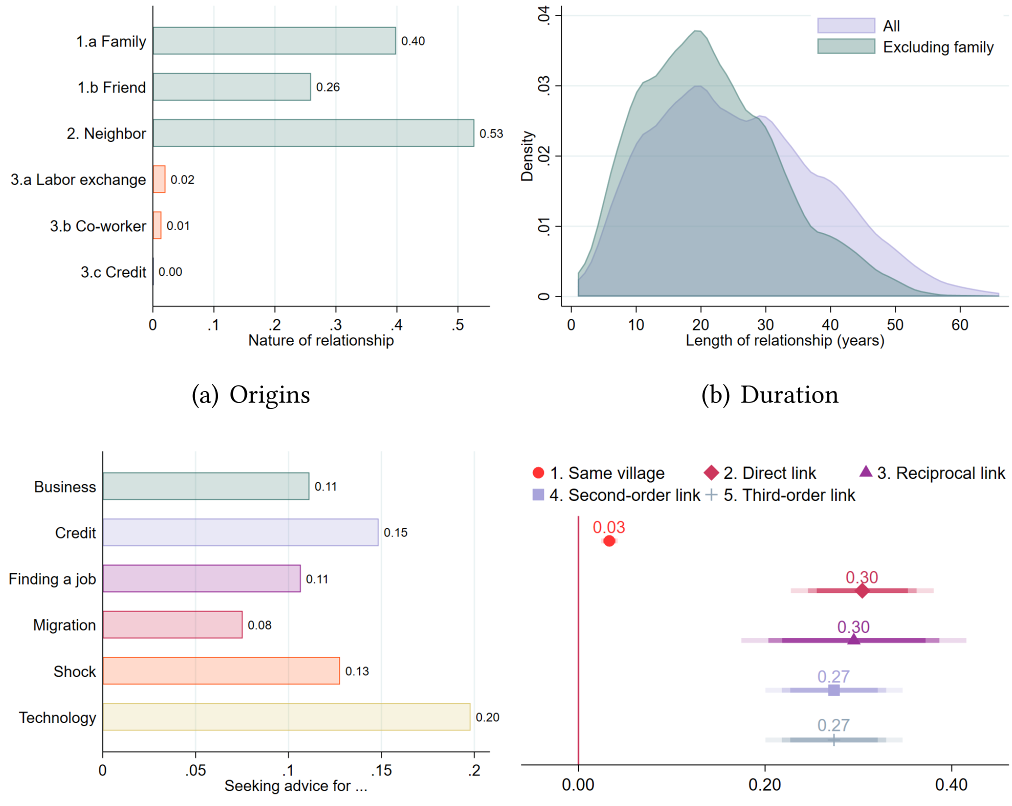
Homophily The nature of our social network (isolated villages with strong links between households) implies a high degree of homophily across two nodes. We illustrate such homophily by successively regressing a standardized outcome, w_i, for household i on: the village average ($\bar{w}_i$); its average across its direct contacts (w_i^d); its average across reciprocal links (w_i^r); its average across second-order contacts (w_i^s); and its average across third-order contacts (w_i^t). We report the estimates in panel (d) of Figure 7 for the likelihood to grow tree crops across land parcels and we leave a more comprehensive description of these correlations to Appendix A.2. One can see that the village correlation is quite small (0.03-0.04), but the correlation in agricultural patterns across a network edge is high: it is about 0.30 for direct links and not much lower for high-order linkages.

In summary, social linkages between households are infrequent, tight and strongly predicted by household characteristics—agricultural practices in particular. This pattern

a large fraction of non-reciprocal relationships are explained by under-reporting, rather than actual non-reciprocity in friendship or in information transmission.

[13]In our analysis, we will ignore family links and only use them to establish higher-order relationships between "non-kin" households. Note that these categories are not exclusive: a contact might be a friend and a neighbor for instance.

Figure 7. Motivations, origins and strength of links.



(a) Origins

(b) Duration

(c) Advice

(d) Homophily

Notes: Panel (a) shows the share of links referred to as: family, friends, neighbors, etc. Panel (b) shows the distribution of a link "duration". Panel (c) shows the share of links used for advice. Panel (d) shows the correlation between two nodes of a link in terms of growing a prime tree crop.

has two implications: it is challenging to isolate exogenous variation in the formation of social linkages; and social linkages might have limited value in terms of (novel) information transmission. We discuss the former next and leave the latter to Section 5.

3 Empirical strategy

This section proceeds in three steps. We first identify predictors of (weak) network linkages. Second, we rely on these predictors to construct measures of treatment exposure through the social network. Third, we present our empirical strategy and discuss the identification assumptions.

3.1 Predicting network links

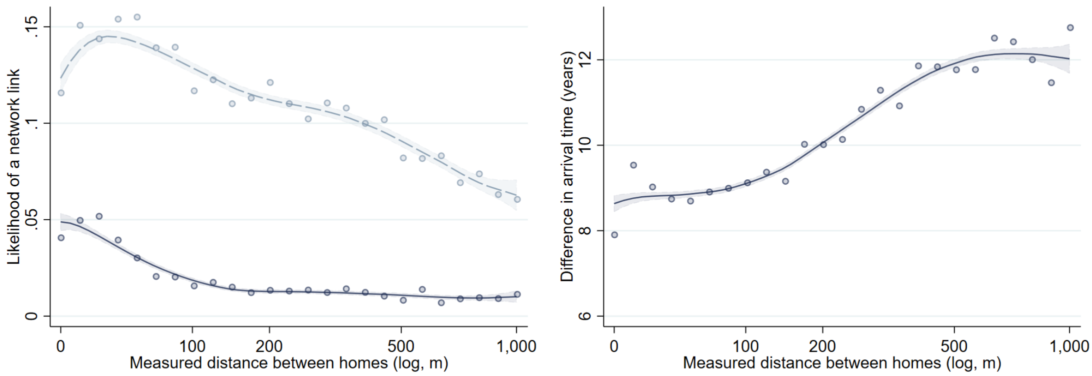
Network links within our villages are shaped by family ties, work practices, and initial settlement patterns (i.e., the origins of settlers and the location of their residential home at destination). In this section, we focus on the latter as more exogenous predictors of first- or second-order linkages between households. We will sometimes refer to these settlement-induced ties as "weak ties", as they remain quite rare and typically less homophilous than family ties and work-based ties.

In what follows, we will exploit the exact location of residential homes within villages as an exogenous predictor of first-order linkage determined by early resettlement patterns. Close neighbors will be more likely to be friends. Predicting second-order linkages, *independently* of first-order linkages, is more challenging. In general, if two households have a friend in common, the underlying factors behind these friendships (e.g., working in the same place, living in a certain street, or practicing certain activities) imply that the two households are likely to be friends. We will exploit a notion of indirect linkage whereby two households could have a friend in common for two *distinct reasons*: they might thus not be directly exposed to each other.

Home proximity, network links, and arrival time In Figure 8, we show the correlation across pairs of households between the existence of a first-order or second-order link and the spatial distance between their *homes* within villages. The unconditional probability for a first-order link to exist between any two households of a village (and not from the same family) is 0.016; this probability however rises up to about 0.05 for households living in very close proximity—between 0 and 100 meters, as illustrated in panel (a) of Figure 8. This distance gradient is less abrupt for second-order links: the probability of a second-order link between any two unrelated households A and B—i.e., there exists at least one household C with a first-order link to A and to B, and A and B are not directly linked—is 0.10 on average; it is however 0.06 for households distant by 1 kilometer and is gradually increasing to 0.14 for households living in close proximity.

Home proximity mostly reflects the settlement patterns of these repopulated villages. These settlement waves followed a concentric logic: the successive waves of households would locate at the fringe of the existing village borders at any point in time such that neighbors are typically households from concurrent settlement waves. We show in panel (b) of Figure 8 how residential proximity correlates with settlement patterns: the average difference in arrival time is about 12 years for households distant by 1 kilometer against 8 years for households living in close proximity. In Appendix B.1, we further show that such settlement patterns do not induce spatial homophily in household ob-

Figure 8. 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.

servable characteristics, e.g., educational attainments or land size.[14] They do however affect agricultural practices on the agricultural land parcels owned by these neighbors, as we will see next. Finally, we demonstrate in Appendix B.1 that residential neighbors are equally likely to be of different origins than any random pair of villagers: timing of arrival does explain location within the village, the origins of settlers not so much.

The role of origins and indirect linkages Even if the origins of different households do not play a role in the location of their residence within villages, these origins do play some role in the formation of network linkages. The probability for a first-order link to exist between any two households of a village (and not from the same family) is 0.022 when they originate from the same province versus 0.013 when they do not. The equivalent probabilities for second-order links are respectively 0.129 and 0.087.

We can combine the two distinct sources of network formation—shared origins and the location of homes—to predict second-order linkages.[15] Consider two households A and B. We define indirect linkages between A and B as the number of households C that are either in close proximity of A and from the same origins as B, or in close proximity

[14]Even if residential settlement was quasi-random, residential proximity could induce some returns to scale in agricultural production (e.g., through the delivery of material). We cannot reject the existence of such an effect, although we expect it to be limited: returns to scale should mostly materialize with proximity between agricultural parcels—a variation that we will include in the baseline controls.

[15]One reasonable alternative would be to leverage inter-generational splits of households, and the relocation of the younger generation within the village.

of B and from the same origins as A.[16] Indirect linkages should be specifically predictive of second-order links for pairs of households that are neither in close proximity of each other nor from similar origins. Conditioning on the spatial and origin proximity between A and B and the number of neighbors and settlers with the same origins, we find that each such indirect linkage adds 0.023 to the probability of actually sharing a second-order link. In the next section, we rely on spatial proximity as a predictor for first-order links and on indirect linkages as a predictor for second-order links.

3.2 A treatment through the network

We now hinge on the previous network links and their exogenous predictors to define measures of exposure and potential exposure to a treatment. Consider a household i, Φ_i its portfolio of land parcels including the residential place, and $p \in \Phi_i$ the index of land parcels. In our terminology, 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 (rubber, pepper, coffee or cashew nuts) on the parcel, and to 0 otherwise. The social network within the village can be summarized by subsets of nodes for each household i: a set of family-linked households $\mathcal{F}_i^1$; a set of non-family-related, yet directly-linked households $\mathcal{L}_i^1$; a set of second-order linked households who are not directly linked $\mathcal{L}_i^2$; etc.[17]

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

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

In other words, ϑ_i^1 is the share of non-family related, yet directly-linked households growing a high-return tree crop on a certain parcel. The previous measure, and its derivatives (e.g., with second-order linkages, ϑ_i^2), will be based on actual linkages.

We construct similar exposure, but predicted by residential proximity,

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

where $\mathcal{D}_i^h$ is the set of households j whose home falls within 100 meters of household i's

[16] In our baseline specification, we will define close proximity as having residences within 100 meters from each other.

[17] We exclude direct family linkages from our analysis for two main reasons: (i) our instrument hinges on the intuition that households do not know each other before moving within the village, and co-settlement patterns induced by larger families moving into a village would violate this assumption; and (ii) we are not able to distinguish these earlier family links from the later, endogenous linkages arising from marriages.

home, or by indirect proximity,

$$\theta_i^{ho} = \frac{\sum_{j \in \mathcal{I}_i^{ho}} \max_{p \in \Phi_j} T_{pj}}{\sum_{j \in \mathcal{I}_i^{ho}} 1},$$

where $\mathcal{I}_i^{ho}$ is the set of households j who are indirectly related to household i, i.e., there is a household k that is either in close proximity of i and from the same origins as j, or in close proximity of j and from the same origins as i.

The previous measures interact the allocation of treatment within the village with the network structure or its predictors. In practice, we construct various indicators capturing these dimensions in a separate manner. For instance, we construct the density of parcels with high-return perennial crops around the various parcels owned by the household to control for the exposure to the treatment that is not mitigated through the network. To capture the relative position of a household in the network, irrespective of the allocation of treatment, we consider the number of first-order linkages, of second-order linkages, and indicators of network centrality (betweenness, closeness, eigenvector centrality, clustering). We finally construct the density of potential linkages for a given household, i.e., the number of households in their immediate proximity or the number of settlers from similar origins.

3.3 Empirical strategy

We rely on two empirical strategies to estimate a network multiplier to treatment adoption. The first approach is cross-sectional in essence, in that it explains treatment adoption in 2022 using the allocation of treatment in 2019 combined with the social network at the time. The second approach is a pseudo-panel approach exploiting the timing of treatment adoption for each treated parcel, the time of formation for each network linkage, and the staggered arrival of households in the village.

Identification The rationale behind our identification is the following. Identifying social multipliers or peer effects within networks is challenging (Bramoullé et al., 2020). The literature has either considered exogeneity in treatment or exogeneity in the allocation of peers. Using our previous notations, the exposure ϑ_i^1 is a combination of agricultural practices in the village (a treatment allocation, $\{T_{pi}\}_i$) and social linkages (an allocation of peers, $\{\mathcal{L}_i^1\}_i$). In theory, identification with random treatment is possible if the network is independent from the treatment. Our context is such that both hypotheses are unlikely to hold: it is hard to isolate exogenous variation in the adoption of crops, and network formation is likely to be shaped by treatment in our "uncontrolled" setting.

Our approach thus needs to extract exogeneity in the allocation of peers.

Social linkages between households are usually not random: the unobserved motivations for a contact to adopt certain agricultural practices might be shared by the household, which would induce a spurious correlation between the first-order exposure, ϑ_i^1, and outcomes of interest. We leverage the variation in settlement patterns discussed in Section 3.1 to extract: (i) exogenous variation in the allocation of (first-order) peers through the close proximity of homes within the village, and (ii) exogenous variation in the allocation of (second-order) peers of peers through our indirect linkages combining origins with timing of arrival to ensure intransitivity. Identification relies on the hypothesis that both variations are indeed quasi-random—conditional on controlling for the geography of agricultural production and the local density of networks.

We further need to impose parametric assumptions, as illustrated in Equation (1). Our choice of functional form for the exposure to the treatment allows us to rely on recent advances in the estimation of shift-share designs. We can write indeed that,

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

such that the treatment of other villagers, $\max_{p \in \Phi_j} T_{pj}$, is a shift and the shares are the normalized linkages from household i to other households j: $1/(\sum_{j \in \mathcal{L}_i^1} 1)$ for peers within $\mathcal{L}_i^1$, and 0 otherwise. As previously discussed, we do not think that shifts are exogenous (thus making approaches based on random shocks irrelevant in our setting, e.g., Adão et al., 2019; Borusyak et al., 2022). We however extract exogenous variation in the shares with the following shift-share instrument,

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

where we assume that the proximity of homes—the share—is exogenous (Goldsmith-Pinkham et al., 2020).

Our model is close to the workhorse linear-in-means model used in empirical work on peer effects (see Bramoullé et al., 2020; Boucher et al., 2022, for a review and for a generalization, respectively). A crucial issue in such literature is the separate identification of endogenous peer effects—a pure impact of treatment adoption from peers in our setting—and contextual peer effects—the impact of the characteristics of peers, e.g., their general technological knowledge. This is at the heart of the famous reflection problem (whereby prospective adoption influences adoption among peers, see Manski, 1993): when all peers share the same peer effects (e.g., the same group of influential peers), endogenous and contextual peer effects cannot be identified separately. The main objective of our study is not to untangle those two effects, even though they might have different policy implications.[18] We do however consider a peer-of-peer variation, as in Bramoullé et al. (2009) or De Giorgi et al. (2010). Peer-of-peer variation (between A and B, through intermediaries C) allows to exclude the direct impact of the characteristics of B on A, and to identify the indirect impact through the actual outcomes for intermediaries.

A two-period approach Our baseline empirical strategy is at the land parcel level. Consider a land parcel p within the portfolio Φ_i of household i. We estimate,

$$y_{pi} = \alpha + \beta\vartheta_i^1 + \gamma\mathbf{X}_{pi} + \varepsilon_{pi} \tag{2}$$

where: the first-order exposure, ϑ_i^1, is instrumented by the predicted exposure through residential proximity, θ_i^h; $\mathbf{X}_{pi}$ includes the previous status of the parcel in 2019 (treated or not), parcel characteristics (area, bulk density, organic carbon content, elevation, slope, distance to the homestead), the latitude, longitude and altitude of the home location, the density of parcels (with high-return perennial crops) around the various parcels owned by the household, the number of first-order linkages, of second-order linkages, and indicators of network centrality (betweenness, closeness, eigenvector centrality, clustering), the number of households in their immediate proximity, and measures of altitude differentials with other homes in the village, and sub-network fixed effects; standard errors are clustered at the household level; and weights are adjusted such that each household contributes equally to the estimation.

Residential proximity is a good predictor of social linkages, even when conditioning for agricultural proximity and other observable characteristics of the land portfolio. We thus expect the predicted exposure through residential proximity (θ_i^h) to predict the exposure through the actual social network (ϑ_i^1). Indeed, we find that an additional standard deviation in predicted exposure through residential proximity increases exposure through the network by 0.21 standard deviations, and this effect is stable across specifications with more or less controls—see Table 2, which constitutes the first stage of our empirical strategy.

One limitation of the previous approach is that it essentially hinges on treatment adoption between 2019 and 2022, and statistical power might be limited. We thus develop a pseudo-panel approach, which exploits the variation induced by the arrival of households over 40 years, the gradual formation of network linkages, and the staggered

[18] For instance, a large endogenous peer effect could justify the implementation of targeted subsidies to crop adoption, while a large contextual peer effect would lead to the design of information programs (if information is indeed underlying the contextual peer effect).

Table 2. 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).

adoption of treatment.

A pseudo-panel approach Our second strategy reconstructs a pseudo-panel of land usage from retrospective information about the dates of land acquisition and (perennial) crop adoption.[19] We proceed in a similar manner to define presence within the settlement (from the date of arrival, as reported by the household head) and the date of formation of each linkage (from the date of linkage formation as reported in the network module). The retrospective information thus only covers the structure of the network (including the evolution of predictors of network formation) and the allocation of treatment at any point in time.

We start in $t_0 = 1980$ and construct a sequence of waves n every three years, i.e., $t_n = t_{n-1} + 3$, for which we observe the main explaining variables (ϑ_{in}^m for $m \in \{1, 2\}$), network controls, and the treatment y_{pin}. We then estimate a specification in difference,

$$y_{pin} - y_{pin-1} = \alpha_i + \beta \left(\vartheta_{in}^1 - \vartheta_{in-1}^1 \right) + \mu_n + \gamma \mathbf{X}_{pin} + \varepsilon_{pin} \tag{3}$$

where the dynamics of exposure is instrumented by the dynamics of predicted exposure

[19] One caveat is that we do not observe the previous land usage, before the planting of the current perennial crop. We assume that a parcel on which a high-return perennial crop is planted at date τ was not treated in previous periods. We also suppose that treatment is an absorbing state.

through residential proximity. The increased exposure, $\left(\vartheta^1_{in} - \vartheta^1_{in-1}\right)$, both reflects the adoption of high-return crops by linked households and the formation of new links. One advantage of the pseudo-panel approach is that we can impose some of these elements to be fixed between consecutive periods such as to neutralize the variation that is induced by changing crop practices among incumbents for instance.

4 Crop adoption and the role of weak links

This section estimates the return to local networks in crop adoption.

4.1 Baseline results

The return to local networks in crop adoption In a first step, we estimate Equation (2) in a linear 2SLS specification. We explain treatment in 2022—a dummy equal to 1 if the land parcel is used to grow high-return perennial crops—by the standardized exposure to the treatment, ϑ^1_i, as weighted by direct links between households. The latter is instrumented by the "exogenous" exposure, θ^h_i, weighted by proximity between homes. Importantly, the empirical specification conditions on the previous status of the parcel in 2019 (treated or not), soil characteristics, the location of the different household land parcels within the village, and the local density of social networks (see the footnote of Table 3 for the full list of controls).

We report the return to local networks obtained through this linear specification in Table 3. We find that an additional standard deviation in exposure (ϑ^1_i) increases the likelihood to grow a high-return perennial crop by about 0.10, whether we only control for geography within the village (column 1), we further condition on soil characteristics (column 2), or we additionally control for network density (column 3). For interpretation purposes, one standard deviation is the average gap between households who have no direct contacts growing high-productive crops and households who have only one first-order contact growing such high-productive crops. This estimate would thus correspond to a very large multiplier: high-return perennial crops are only grown on a fifth of agricultural land parcels in 2022.

One concern with the previous specification is that crop adoption is a rare event. In Table 4, we replicate the exercise of Table 3 in a probit specification based on Equation (2) and report the average marginal effects of direct exposure to the treatment (ϑ^1_i). The results are however very similar to that of Table 3.

The previous two-period approach suffers from a few other caveats: (i) crop adoption is not frequent such that the estimation relies on about 50 "adopters"; (ii) the latter im-

Table 3. The return to social network—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).

Table 4. The return to social network—a probit specification.

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).

plies that the specification is under-powered to analyze the role of second-order linkages; and (iii) the previous identification is cross-sectional in essence and does not exploit the variation induced by recent arrivals to the village versus the variation induced by incumbents (possibly contaminated by a reflection problem). For these reasons, we exploit a pseudo-panel of network formation and dynamic crop adoption in the next exercise.

A pseudo-panel approach We now exploit the dynamics of crop adoption, exposure to the treatment, arrival time in the village and network formation by constructing our treatment and exposure variables every 3 years from 1980 to 2022. In 1980, about 18% of contemporary land parcels were already cultivated by the few existing settlers, with about than 2.5% of those growing high-return perennial crops (on 13 parcels). In 2022, about 650 parcels are growing high-return perennial crops.

Table 5. 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).

To explain the staggered crop adoption across these land parcels, we estimate a specification in difference (see Equation 3) whereby crop adoption among incumbents, $y_{pin} - y_{pin-1}$, is explained by the evolution of exposure to the treatment, $\vartheta^m_{in} - \vartheta^m_{in-1}$, where we fix the allocation of treatment in wave $n-1$. In other words, the dynamics of exposure, $\vartheta^m_{in} - \vartheta^m_{in-1}$, is only allowed to evolve due to the formation of new links in the village, e.g., as triggered by the arrival of new households. This object is instrumented by the dynamics of the exogenous exposure, $\theta^h_{in} - \theta^h_{in-1}$ or $\theta^{ho}_{in} - \theta^{ho}_{in-1}$, where we fix the allocation of treatment in wave $n-1$. We consider three variations for the exposure variable: (i) a general exposure to the treatment, ϑ^0_{in} combining linkages of different orders;[20] (ii) the first-order exposure to the treatment, ϑ^1_{in}; and (iii) the second-order exposure to the treatment, ϑ^2_{in}, for households that are not directly exposed to the treatment, i.e., for whom $\vartheta^1_{in} = 0$.

Table 5 shows that a change in exposure to the treatment triggered by changes in the composition of neighbors does affect crop adoption (column 1). An additional standard deviation in general exposure (ϑ^0_{in}) increases the likelihood to adopt a high-return perennial crop between waves by about 0.014. This effect could be driven by direct exposure to the treatment through peers or indirect exposure through peers of peers. In column 2, we find that an additional standard deviation in first-order exposure (ϑ^1_{in}) increases treatment adoption by about 0.051 (an estimate directly comparable to the ones presented in Table 3). In column 3, we exclude all households with *some* first-order exposure to the treatment and consider second-order exposure instrumented by indirect linkages as the dependent variable: we find that an additional standard deviation in second-order exposure (ϑ^1_{in}) increases treatment adoption by about 0.029—a non-negligible effect, albeit imprecisely estimated.

These estimates are slightly lower than those obtained through the estimation of Equation (2). We verify in Appendix B.2 that this is not a byproduct of the longer time coverage and the inclusion of periods with limited adoption (e.g., before 2006). In what follows, we calibrate our social multiplier on the estimate of column (2), i.e., $\beta = 0.051$—possibly a lower bound, which however remains substantial. Before discussing the coexistence of a large network multiplier with a relatively low adoption rate, we provide a sensitivity analysis of the main estimates and a brief exploration of possible mechanisms.

4.2 Robustness checks and possible mechanisms

Robustness checks We now briefly describe a series of robustness checks and leave the more detailed evidence to Appendix B.2. First, we discuss identification concerns and the exact specification of strategy (2). We report the OLS regressions and show that the estimates are then slightly positive, albeit very close to 0. Our preferred interpretation is that there is a high degree of homophily within networks, the status of households

[20] We define the general exposure to the treatment, ϑ^0_i, as follows,

$$\vartheta^0_i = \frac{\sum_{m=1}^{3} \sum_{j \in \mathcal{L}^m_i} \delta^m \max_{p \in \Phi_j} T_{pj}}{\sum_{m=1}^{3} \sum_{j \in \mathcal{L}^m_i} \delta^m}.$$

While the previous formula appears complicated, it simply consists in applying a discount to higher-order linkages: a discount factor δ, set to be equal to 0.5, applies to second-order linkages compared to first-order linkages, and the discount factor δ applies to third-order linkages compared to second-order linkages.

is very correlated with their neighbors', and we observe very few changes in cropping patterns as induced by the average network link. In the baseline specification, however, we exploit the smaller subset of weaker links which can bring change. We also provide a specification with additional controls to better condition on the unobserved heterogeneity that is possibly correlated with residential proximity to treated households (e.g., land tenure, soil acidity/organic content, land quality, and household characteristics). We further provide a sensitivity analysis with alternative instruments/treatments.

Second, we exploit the empirical specification (3) to better characterize the timing of adoption. We add lags and leads in changes of exposure to capture a possible delay in crop adoption and to test that future settlement patterns are orthogonal to previous treatment. We also estimate our social multiplier across different periods of interest, and across settlers with different tenure in the village. Our findings indicate that the social multiplier remains quite similar over time, but more recently arrived households appear more responsive—even though our estimate is robust to excluding the very freshly-arrived households.

Table 6. 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 3. 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).

Possible mechanisms To investigate the channels through which exposure to the treatment may affect farmers, we estimate a specification similar to Equation (2) (see, e.g., the last column of Table 3), but we replace the dependent variable by: (i) a subjective evaluation of land quality; (ii) the likelihood to declare the land parcel suitable for the cultivation of any of the main high-return crops (coffee, rubber, pepper, and cashew nuts); (iii) a subjective evaluation of land needs (fertilizers, pesticides, and herbicides); (iv) a subjective evaluation of climate risk (flood, drought, and water pollution); and (v) a declared willingness to consider growing high-return crops on the land parcel. We report the outcome of these estimations in Table 6.

Exposure to the treatment appears to affect beliefs about land suitability for high-return crops (see column 2) and whether the respondent has seriously considered growing high-return perennial crops on the land parcel (see column 5). Both effects are sizable: for instance, an additional standard deviation in exposure increases the likelihood to consider the land parcel as suitable by 0.21—an effect as large as the mean of the variable. This effect is consistent with uninformed farmers over-estimating the requirements for growing certain crops through the lack of first-hand experience (and consistent with the first-hand experience of neighbors being instrumental in treatment adoption, see Foster and Rosenzweig, 1995; Conley and Udry, 2010). The perfect example is the cashew-growing tree: the plant has low water and soil requirements; it can survive reasonably long periods of drought, can be planted on moderately fertile soil, does not require much input, and can accommodate steep gradients. Pepper or coffee are also not very demanding in terms of input or in terms of soil characteristics.

5 Network structure and the (limited) impact of weak links

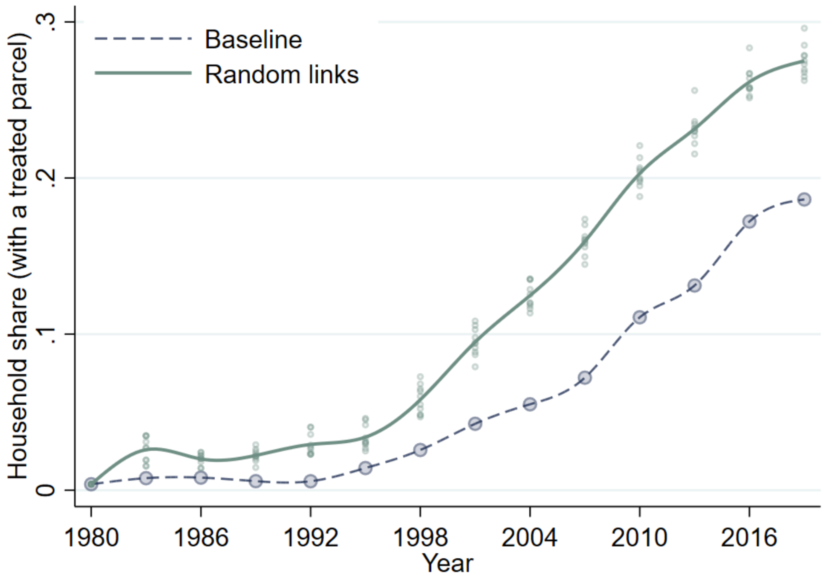
How do we reconcile a sizable network multiplier with the original puzzle? This section shows that: the puzzle is explained by homophily across the average link; such homophily endogenously arises from clustered networks; the returns to social links tend to decrease with time; and policies targeting "inbetweeners" are most able to mitigate this issue.[21]

Homophily and the propagation of treatment Treatment adoption in the village appears limited, given the size of the multiplier associated with a first-order linkage to treated households. One reason could be the high homophily within the network: households that are connected to treated households typically grow highly-productive crops already, and households in need for exposure to the treatment are typically not exposed through their network of friends and neighbors. We documented such homophily in agricultural practices across linked households in panel (d) of Figure 7.

In order to shed light on the role of homophily in tempering the impact of social mul-

[21]Throughout this section, our back-of-the-envelope approach ignores strategic considerations from households: the social network is taken as given and fixed over time, and the adoption of highly-productive agricultural practices is modeled as a diffusion process. A costly alternative would be to consider counterfactual scenarios within a structural model of crop adoption (and possibly linkage formation).

Figure 9. Network structure and the propagation of treatment—randomizing past exposure.

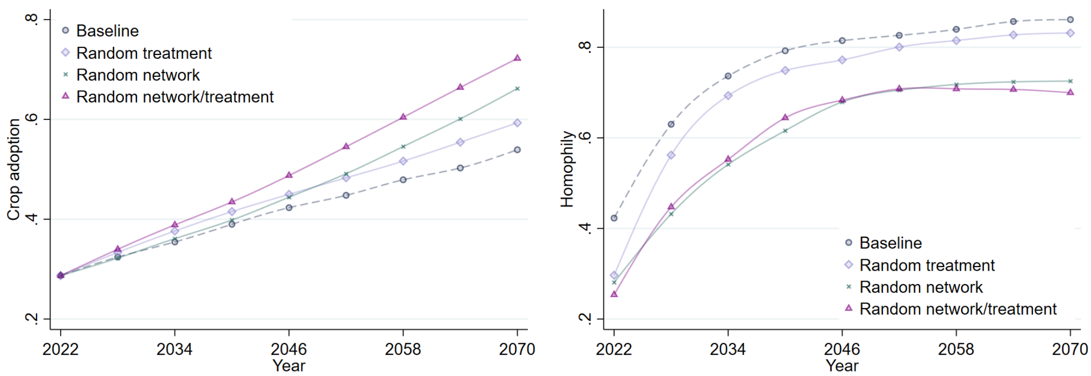


Notes: This Figure compares the actual share of households with a treated parcel (i.e., growing coffee, rubber, cashew nuts, or pepper) to counterfactual shares between 1980–2019. The counterfactuals are based on: (i) the wave-specific predicted likelihood to adopt and treatment exposure at that time, 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.

tipliers, we simulate crop adoption within actual and counterfactual social networks in the past (1980–2019). In this backward projection, we use our estimated Equation (3) to predict treatment propagation within the village, accounting for the timing of arrival and the staggered formation of linkages. We do so in two scenarios: (i) using the actual network formation $\left\{\mathcal{L}_i^m\right\}_{i,m}$ to provide a benchmark prediction, and (ii) using re-sampled, random linkages (with their timing) across pairs of households while keeping the same overall network density at any given time. We find that randomizing network linkages would have generated a faster adoption of high-return agricultural practices in 2019, as shown in Figure 9. High-return crops were planted on at least one parcel for about 18% of households in 2019 (blue, dashed curve). Counterfactual scenarios with random linkages all produce a higher adoption rate: on average, high-return crops would be planted on at least one parcel for about 27% of households in 2019 (green dots).

In theory, homophily arises for two distinct reasons: (i) exogenous factors underlying *both* the formation of linkages and the adoption of treatment; and (ii) the endogenous propagation of treatment *through the network*. First, the geographic location of land parcels or the socio-economic characteristics of households (e.g., their education, the other income sources such as remittances, their wealth) do predict agricultural practices and network formation, at least to some extent. Our empirical strategy described in Section 3 was designed to neutralize such variation. Second, the distribution of treatment across farmers might exhibit strong homophily through the social network, even when the initial treatment allocation is initially random. To understand this argument, consider a clustered network, i.e., a network where the following property holds: if there exists a household C with a link to A and to B, then A and B are more likely to be connected. In such a clustered network, treatment would primarily propagate within some clusters; and the return to social links would then decrease over time.[22] We quantify this latter effect in the following section.

Figure 10. 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 7. Note that we randomize agricultural practices in 2022 such as to keep the same exact incidence for each village as in the baseline.

[22] Almost all social networks exhibit some form of clustering: the formation of links is usually affected by underlying factors that are not idiosyncratically distributed in the population. When these correlated factors directly predict treatment adoption, their distribution across households generates clustered networks with strong homophily—as in our example (i). In our context, for instance, farmers hold agricultural parcels in certain parts of the village; their land portfolios jointly determine the formation of social links and the adoption of certain cropping practices. When these correlated factors do not directly predict treatment adoption, the dynamic propagation of agricultural practices through the network would generate clustering.

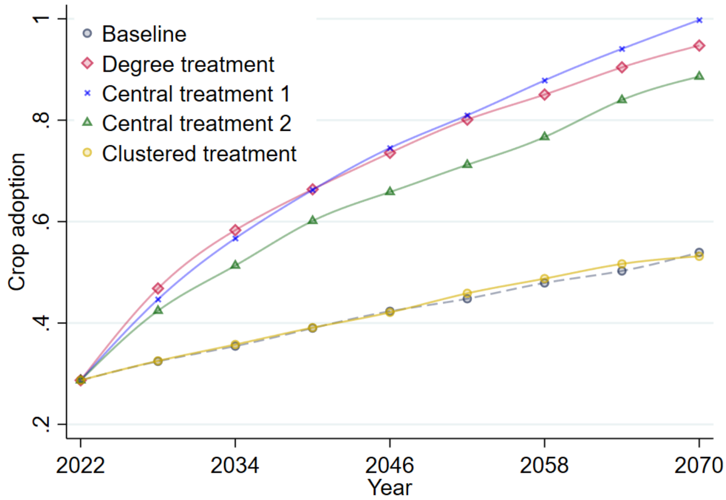
Homophily, clustering, and the propagation of treatment To better understand the importance of network structure, we leverage our previous estimates to predict crop adoption forward between 2022–2070. In these counterfactual experiments, we assume that the network structure, represented by the set of m-order linkages, $\left\{\mathcal{L}_i^m\right\}_{i,m}$, remains constant over time.[23] Households will however change cropping patterns across simulations and over time, thereby providing a different distribution of treatment within the village and a different exposure to the treatment for households who have yet to adopt the highly-productive crops. We consider four scenarios: (i) a baseline projection with the actual network structure and distribution of agricultural practices in 2022; (ii) randomized agricultural practices in 2022; (iii) randomized network linkages; and (iv) randomized agricultural practices in 2022 and network linkages.

We report the outcome of our forward projection in Figure 10. Panel (a) illustrates the evolution of the share of households with a treated parcel (i.e., growing coffee, rubber, cashew nuts, or pepper). In the baseline, high-return crops would be planted on at least one parcel for about 52% of households in 2070, with a slowdown of adoption over time. Randomizing the initial treatment distribution would lead to a 60% adoption rate in 2070, with a negative inflection over time. Randomizing exposure through random links also generates a higher adoption rate in 2070 (about 65%), as shown in Figure 10; randomizing links allows for a more convex pattern in crop adoption: the impact of weak links does not fade away so much over time. Randomizing both the network structure and the initial treatment distribution would lead to a 72% adoption rate in 2070, without any negative inflection over time. Panel (b) provides an explanation behind the differential dynamic returns observed across simulations: homophily in agricultural practices is initially high in the baseline and markedly increases in the first 25 years. After 2050, the correlation in agricultural practices across network linkages is around 0.85. Reshuffling the initial distribution of agricultural practices lowers homophily in the short run. The correlation in agricultural practices across network linkages however swiftly increases in the first 25 years, and homophily becomes almost as high as in the baseline. The crucial underlying factor is the clustered structure of networks: in the experiments with randomized networks, homophily rises at a much more steady pace and remains below 0.70, even in the medium-long run.

Clustered networks significantly limit dynamic social multipliers in treatment adoption. There are, however, mitigation strategies: treatment could be targeted to those few important in-between nodes within the network. We study such targeted policies next.

[23] We implement this exercise between 2022–2070, rather than retrospectively for the following reason: households would arrive in the village between 1980–2019 and new links are formed, such that the social network would need to evolve.

Figure 11. 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.

Targeted policies We conclude the analysis by discussing policy and how to target the most relevant households with respect to their network position. Maximizing propagation is notoriously hard (and not so important in the long run under some propagation mechanisms, see, e.g., Akbarpour et al., 2023), and we cannot identify a type-specific network multiplier (Sadler, 2023). For these reasons, we focus on the role of network nodes with high centrality coefficients (following Banerjee et al., 2013; Kim et al., 2015; Banerjee et al., 2019; Beaman et al., 2021, for instance). To this purpose, we replicate the approach described in the previous section and redistribute treatment in 2022: (T1) to connected households—with the highest network degree; (T2) to households that are bridges between other households—with the highest betweenness centrality measure; (T3) to households that are most central—with the highest closeness centrality measure; (T4) to most clustered households—with the highest clustering coefficient.

We report the outcome of projected crop adoption across these scenarios in Figure 11. We find that the baseline projection is not too different from the scenario where we allocate treatment to the most clustered household, a strong indication that the initial allocation of treatment in our villages is least prone to treatment diffusion. In stark contrast, reshuffling treatment to connected households (T1), to households most likely to form a bridge between other households (T2), or to central households (T3) would hugely accelerate treatment adoption. Interestingly, the most efficient of these experiments is the one targeting "in-betweeners" through the betweenness centrality measure: policies should be targeted at nodes that are most likely to be along the shortest paths between any two other nodes.

6 Concluding remarks

This paper documents an apparent contradiction: while there are high returns to adopting certain cropping practices, and observing such practices being adopted within one's social network leads to further adoption, actual adoption remains limited. We show that it is related to the network structure itself: the useful social linkages—that we exploit for identification purposes—across households with different practices and beliefs are rare. The homophily of networks prevents productive practices from being adopted and limits the social spillovers to adoption. We show how such homophily arises endogenously from the structure of village networks and how targeted policies might alleviate this issue.

A limit to our present study is to ignore other frictions (credit, land, insurance, technology) that could slow down agricultural transformation within rural villages. Another limit is that it considers the crop premium as fixed and ignores fluctuations in such premium: adopting coffee might be a good idea in 2022, given current coffee prices, but these conditions might change. We explore the role of fluctuations, fixed costs, technological frictions, and frictional factor markets in explaining (the lack of) changes in agricultural practices in a companion paper. Finally, village networks might propagate information, but also shocks, and those households that are connected to high-productivity, high-risk villagers might bear part of this risk (Kinnan et al., 2024).

References

Acemoglu, Daron, Asuman Ozdaglar, and Ercan Yildiz, "Diffusion of innovations in social networks," in "2011 50th IEEE Conference on Decision and Control and European Control Conference" 2011, pp. 2329–2334.

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

__ **, Loren Brandt, Jessica Leight, and Diego Restuccia**, "Misallocation, Selection, and Productivity: A Quantitative Analysis With Panel Data From China," *Econometrica*, 2022, *90* (3), 1261–1282.

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales, "Shift-share designs: Theory and inference," *The Quarterly Journal of Economics*, 2019, *134* (4), 1949–2010.

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi, "Just a few seeds more: value of network information for diffusion," Technical Report 2023.

Bandiera, Oriana and Imran Rasul, "Social Networks and Technology Adoption in Northern Mozambique," *The Economic Journal*, 2006, *116* (514), 869–902.

__ **, Robin Burgess, Erika Deserranno, Ricardo Morel, Munshi Sulaiman, and Imran Rasul**, "Social Incentives, Delivery Agents, and the Effectiveness of Development Interventions," *Journal of Political Economy Microeconomics*, 2023, *1* (1).

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson, "The Diffusion of Microfinance," *Science*, 2013, *341* (6144), 1236498.

__ **, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson**, "Using gossips to spread information: Theory and evidence from two randomized controlled trials," *The Review of Economic Studies*, 2019, *86* (6), 2453–2490.

Barnes, Michele L, John Lynham, Kolter Kalberg, and PingSun Leung, "Social networks and environmental outcomes," *Proceedings of the National Academy of Sciences*, 2016, *113* (23), 6466–6471.

Beaman, Lori and Andrew Dillon, "Diffusion of agricultural information within social networks: Evidence on gender inequalities from Mali," *Journal of Development Economics*, 2018, *133*, 147–161.

__ **, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak**, "Can network theory-based targeting increase technology adoption?," *American Economic Review*, 2021, *111* (6), 1918–43.

Benjamin, Dwayne, Loren Brandt, Brian McCaig, and Nguyen Le Hoa, "Program participation in a targeted land distribution program and household outcomes: evidence from Vietnam," *Review of Economics of the Household*, 2018, *16* (1), 41–74.

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

— **, Maria Jones, Florence Kondylis, and Ahmed Mushfiq Mobarak**, "Gender gaps in technology diffusion," *Journal of Development Economics*, 2020, *143*, 102380.

Blarel, Benoit, Peter Hazell, Frank Place, and John Quiggin, "The economics of farm fragmentation: evidence from Ghana and Rwanda," *The World Bank Economic Review*, 1992, *6* (2), 233–254.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel, "Quasi-experimental shift-share research designs," *The Review of Economic Studies*, 2022, *89* (1), 181–213.

Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou, "Toward a general theory of peer effects," Technical Report 2022.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, 2009, *150* (1), 41–55.

— **,** — **, and** — , "Peer effects in networks: A survey," *Annual Review of Economics*, 2020, *12*, 603–629.

Breza, Emily, Arun G Chandrasekhar, Tyler H McCormick, and Mengjie Pan, "Using aggregated relational data to feasibly identify network structure without network data," *American Economic Review*, 2020, *110* (8), 2454–2484.

Burchardi, Konrad B, Selim Gulesci, Benedetta Lerva, and Munshi Sulaiman, "Moral hazard: Experimental evidence from tenancy contracts," *The Quarterly Journal of Economics*, 2019, *134* (1), 281–347.

Burlig, Fiona and Andrew W Stevens, "Social networks and technology adoption: Evidence from church mergers in the US Midwest," *American Journal of Agricultural Economics*, 2023.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet, "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, April 2015, *7* (2), 81–108.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, "Peer effects and social networks in education," *The review of economic studies*, 2009, *76* (4), 1239–1267.

Chakraborty, Aranya, "Network-Based Targeting with Heterogeneous Agents for Improving Technology Adoption," Technical Report 2022.

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

— **, Diego Restuccia, and Raül Santaeulàlia-Llopis**, "The effects of land markets on resource allocation and agricultural productivity," *Review of Economic Dynamics*, 2021.

Comola, Margherita, Carla Inguaggiato, and Mariapia Mendola, "Learning about Farming: Innovation and Social Networks in a Resettled Community in Brazil," IZA Discussion Papers 14092, Institute of Labor Economics (IZA) February 2021.

Conley, Timothy G and Christopher R Udry, "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, 2010, *100* (1), 35–69.

de Janvry, Alain, Manaswini Rao, and Elisabeth Sadoulet, "Seeding the Seeds: Role of Social Structure in Agricultural Technology Diffusion," Technical Report 2022.

de Walle, Dominique Van and Dileni Gunewardena, "Sources of ethnic inequality in Viet Nam," *Journal of Development Economics*, 2001, *65* (1), 177–207.

Dercon, Stefan and Luc Christiaensen, "Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia," *Journal of Development Economics*, 2011, *96* (2), 159–173.

Duflo, Esther, Michael Kremer, and Jonathan Robinson, "How high are rates of return to fertilizer? Evidence from field experiments in Kenya," *American Economic Review*, 2008, *98* (2), 482–88.

_, _, and _, "Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya," *American Economic Review*, 2011, *101* (6), 2350–90.

Emerick, Kyle, Alain De Janvry, Elisabeth Sadoulet, and Manzoor H Dar, "Technological innovations, downside risk, and the modernization of agriculture," *American Economic Review*, 2016, *106* (6), 1537–61.

Evans, Grant, "Internal colonialism in the Central Highlands of Vietnam," *Sojourn: Journal of Social Issues in Southeast Asia*, 1992, pp. 274–304.

Fabregas, Raissa, Michael Kremer, and Frank Schilbach, "Realizing the potential of digital development: The case of agricultural advice," *Science*, 2019, *366* (6471), eaay3038.

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

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

Galton, Francis, "Vox populi," *Nature*, 1907, *75* (1949), 450–51.

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

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift, "Bartik instruments: What, when, why, and how," *American Economic Review*, 2020, *110* (8), 2586–2624.

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

_ , David Lagakos, and Michael E Waugh, "Agricultural productivity differences across countries," *American Economic Review*, 2014, *104* (5), 165–70.

_ , _ , and _ , "The agricultural productivity gap," *The Quarterly Journal of Economics*, 2014, *129* (2), 939–993.

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

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

Hardy, Andrew, *Red hills: migrants and the state in the highlands of Vietnam*, Vol. 93, NIAS press, 2005.

Hengl, Tomislav, "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," 2018.

_ and Ichsani Wheeler, "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," 2018.

Hoch, Jannis M, Jeffrey Neal, Fedor Baart, LPH van Beek, Hessel Winsemius, Paul Bates, and Marc FP Bierkens, "GLOFRIM v1. 0–A globally applicable computational framework for integrated hydrological–hydrodynamic modelling," *Geoscientific Model Development*, 2017, *10*, 3913–3929.

Jackson, Matthew O, Stephen M Nei, Erik Snowberg, and Leeat Yariv, "The Dynamics of Networks and Homophily," Technical Report, National Bureau of Economic Research 2023.

Jochmans, Koen, "Peer effects and endogenous social interactions," *Journal of Econometrics*, 2023, *235* (2), 1203–1214.

Kala, Namrata, "Learning, adaptation, and climate uncertainty: Evidence from Indian agriculture," Technical Report 2017.

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

Kinnan, Cynthia, Krislert Samphantharak, Robert Townsend, and Diego Vera-Cossio, "Propagation and insurance in village networks," *American Economic Review*, 2024, *114* (1), 252–284.

Lagakos, David, Ahmed Mushfiq Mobarak, and Michael E Waugh, "The welfare effects of encouraging rural-urban migration," Technical Report, National Bureau of Economic Research 2018.

Laskievic, Andrei Arminio, "Commodity Booms and Structural Transformation: The Role of Input Use and Land Inequality," Technical Report 2021.

Le, Kien, "Land use restrictions, misallocation in agriculture, and aggregate productivity in Vietnam," *Journal of Development Economics*, 2020, *145*, 102465.

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

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

Perego, Viviana ME, "Crop prices and the demand for titled land: Evidence from Uganda," *Journal of Development Economics*, 2019, *137*, 93–109.

Restuccia, Diego and Raul Santaeulalia-Llopis, "Land misallocation and productivity," Technical Report, National Bureau of Economic Research 2017.

_ **and Richard Rogerson**, "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic Dynamics*, 2008, *11* (4), 707–720.

Sadler, Evan, "Seeding a simple contagion," Technical Report 2023.

Sampson, Christopher C, Andrew M Smith, Paul D Bates, Jeffrey C Neal, Lorenzo Alfieri, and Jim E Freer, "A high-resolution global flood hazard model," *Water Resources Research*, 2015, *51* (9), 7358–7381.

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

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

Ty, Tran Van, Kengo Sunada, and Yutaka Ichikawa, "A spatial impact assessment of human-induced intervention on hydrological regimes: a case study in the upper Srepok River basin, Central Highlands of Vietnam," *International Journal of River Basin Management*, 2011, *9* (2), 103–116.

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

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

ONLINE APPENDIX—not for publication

A Data appendix

This section provides complements to the data description of Section 2 with: a detailed presentation of our data sources; and additional evidence about agricultural practices and the structure of social networks.

A.1 Data sources

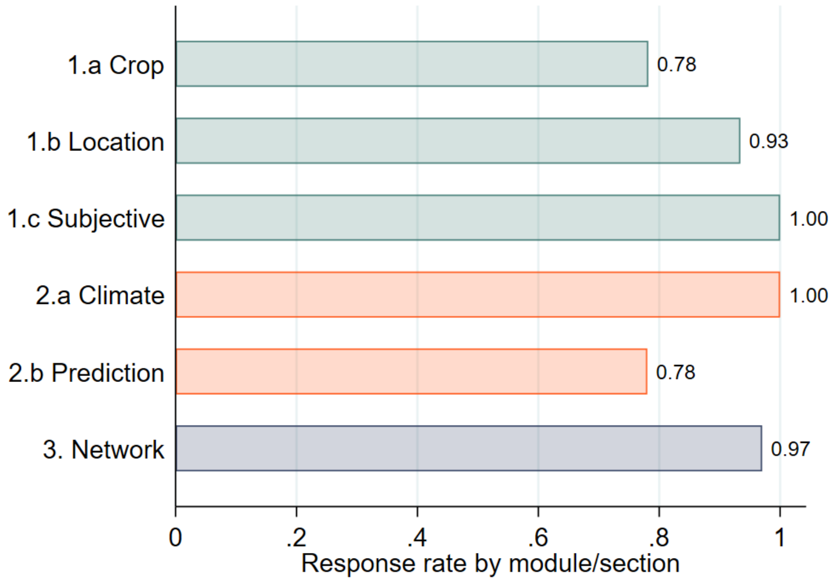
Our main data source is a survey of households in four rural villages of the Central Highlands of Vietnam.

A survey of households in rural Vietnam The first round of the survey was conducted in September 2019 as part of the project "An Interdisciplinary Approach to Understanding Past, Present and Future Flood Risk in Viet Nam", funded by the Natural Environment Research Council and the National Foundation for Science and Technology Development through the Newton Hydrometeorological Hazards (NE/S003061/1). The survey was supervised by Pham Khanh Nam (University of Economics, Ho-Chi-Minh City, EEPSEA) and Truong Dang Thuy (University of Economics, Ho-Chi-Minh City), while the questionnaire was designed by Niels Wendt (Institut für Wirtschafts und Kulturgeographie, Leibniz Universität Hannover) under the supervision of the full research team including Andre Groeger and Yanos Zylberberg. A follow-up survey was conducted in September 2022 by the same team, and financed by the British Academy through the Humanities and Social Sciences Tackling Global Challenges programme (TGC/200149).

The survey questionnaire was based on the 2017 questionnaire of the Thailand Vietnam Socio Economic Panel (TVSEP). The latter survey covers about 220 villages across three provinces of Vietnam; we randomly selected four villages of this sample, located in the Ea Súp District, Lak District, Krông Bông District and Krông Pac District, and all within the Dak Lak province. The survey was implemented by a team of 4 supervisors, about 25-30 enumerators, and the questionnaire was coded using the software Survey Solutions provided by the World Bank. The survey covers 945 households in 2019, and 950 households in 2020, with an attrition rate of about 5% between the two waves, i.e., 45 households exit the survey in 2019 and 50 new households are interviewed. Our main analysis only considers households present in both waves.

The survey has 14 modules: (1) a cover section, (2) general household information, (3) an individual roster covering demographics, education, health and family/informal transfers, (4) a network module including a labor exchange sub-module, (5) household expenditures, (6) a land (usage) module covering agriculture, livestock, and aquaculture, (7) a risk section, (8) a wage section, (9) a non-farm self-employment section, (10) a

Figure A.1. Attrition across modules.



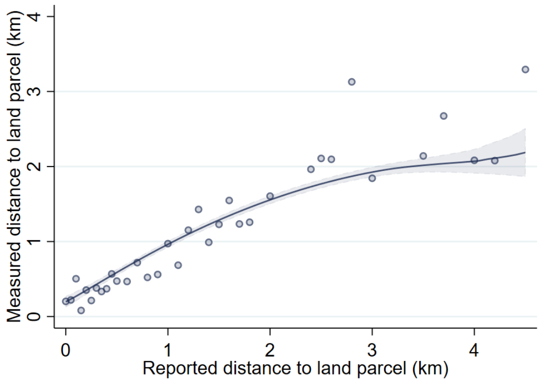
Notes: This figure reports attrition across the different (novel) modules. *1.a Crop* refers to the agricultural section; *1.b Location* refers to the geolocation of land parcels; *1.c Subjective* refers to the subjective assessment of land quality, suitability and land requirements for agricultural production; *2.a Crop* refers to the climate module; *2.b Crop* refers to an evaluation module where households would evaluate land quality, suitability and land requirements for a (randomly selected) subset of parcels in the village; and *3. Crop* refers to the network module.

credit and public transfer section covering formal and informal credit, possible defaults, public transfers and insurance, (11) an investment and assets section, (12) a section about housing conditions, (13) an interview evaluation by enumerator, and (14) a meta-section including statistics on the tablet usage and comments by enumerators, supervisors and data checkers. The latter section allowed us to perform checks during the survey and conduct additional interviews when needed. Interviewers were mostly undergraduate economics students from the local university, Tay Nguyen University (TNU). An average interview would last 3 hours and cover about 1,000 questions.

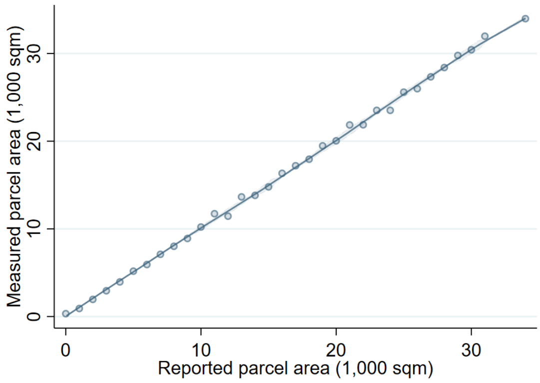
While the core of the survey is following the general standards of the literature (i.e., World Bank questionnaires), it contains more in-depth information on: labor exchange, the structure of networks within villages with a dedicated section and numerous questions linking specific peers to other questions (e.g., about informal credit), a land module with a geo-location of parcels, subjective assessment about land quality and needs, a climate change module. We show the attrition rate across the novel modules in Figure A.1 where we see that attrition is minimal, except for agricultural production (about 15-20% of households do not produce anymore) and the prediction module where households would evaluate land quality, suitability and land requirements for a (randomly selected) subset of parcels in the village.

We describe some of these novel modules in greater detail below.

Figure A.2. Validation for the geolocation procedure.



(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.

Geo-location of parcels The geo-localization of parcels is part of the land module and proceeds as follows:

- A 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 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.

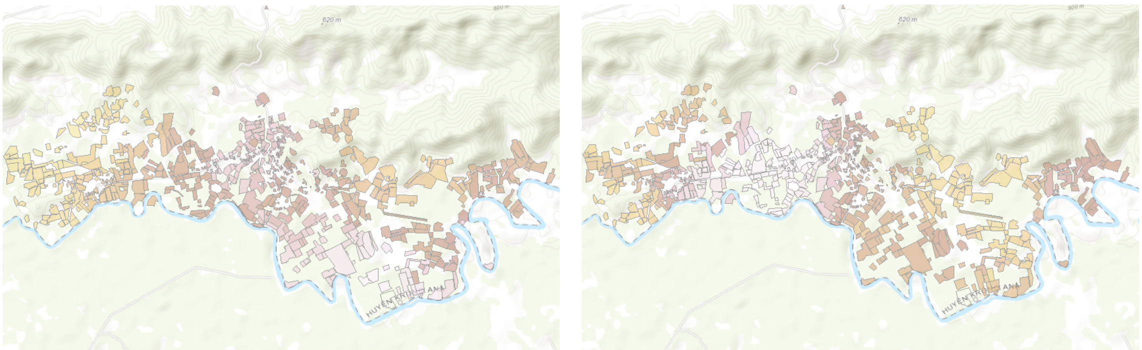
- Once the location of the land parcel is identified by the respondent on the map, the interviewer draws a polygon under the instructions of the respondent.

- Additional questions help capture possible issues with the geo-localization, e.g., how sure the respondent may be, how much help was needed etc.

The geo-location procedure is cheap and fast, it can be integrated into any survey through Survey Solutions (developed by the World Bank); attrition is very low. One question remains: how precise is it, and how does it compare to the actual boundaries of land parcels?

We report in Figure A.2 a comparison between the measured land characteristics (area and the distance to the main house) and these characteristics as reported by the respondent. There is a strong, positive relationship between the measured and reported distance, even though respondents tend to overestimate the distance to the land parcel, probably reflecting that they interpret the question as asking for the travel distance rather than the distance as the crow flies.

Figure A.3. Inferred soil quality (pH, acidity, and P_2O_5dt, phosphate).



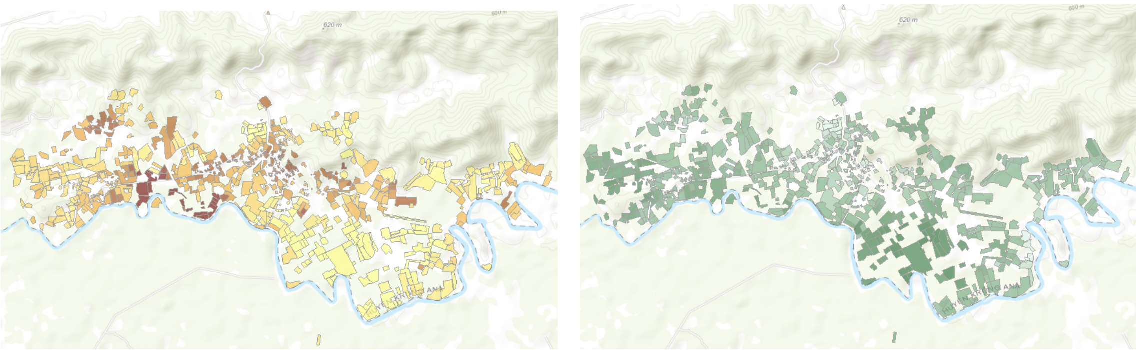
(a) pH (acidity) (b) P_2O_5dt (phosphate)

Notes: These figures report the inferred soil content from about 100 soil samples in Village 1 of our survey. We use kriging to generate a predicted map of soil characteristics from our 100 discrete samples.

Objective soil data Our survey collects subjective evaluation of land quality and soil needs. We complement this data with about 300 soil samples in 2022, and test for: pH (acidity), $OMts$ (organic content), Nts (nitrogen), P_2O_5dt (phosphate), K_2Odt (potash). The soil samples were selected from the subset of well geo-located land parcels in 2019 and chosen to best cover the different land usages (annual, perennial, non-agricultural, fallow) and to ensure a dense geographic coverage of our four villages. The output of the procedure is displayed in Figure A.3 for one of our four villages. Note that we use the inferred soil content in robustness checks—see Appendix B.2.

We also extract high-quality data on topography at a 30m precision and soil characteristics at a 100m precision from satellite imagery and derivatives (Hengl, 2018; Hengl and Wheeler, 2018). These data, shown in Figure A.4, allow us to control for objective land parcel characteristics (area, bulk density, organic carbon content, elevation, slope) in our baseline specifications—see Sections 3.3 and 4.1.

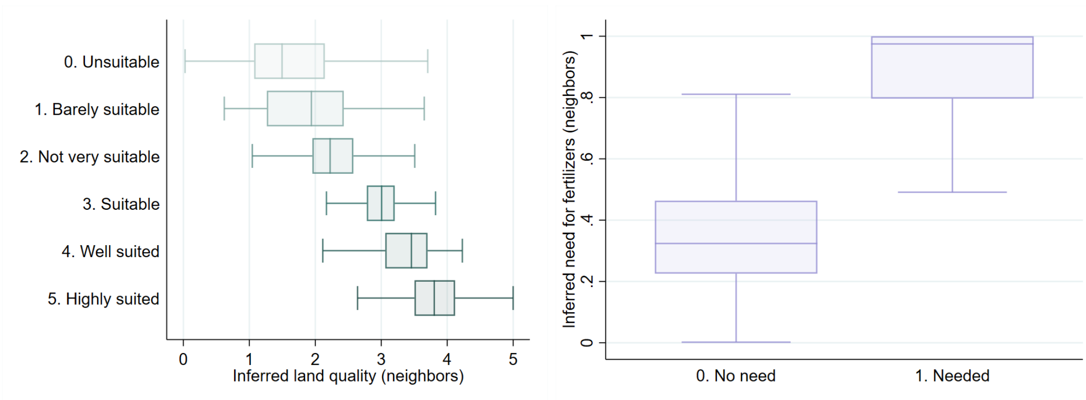
Figure A.4. Soil characteristics (bulk density, and carbon content).



(a) Soil bulk density at 0 cm depth (b) Organic carbon content

Notes: These figures report inferred carbon content and soil bulk in Village 1. 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.

Figure A.5. Comparing evaluations of land quality and need for fertilizer usage (own versus other villagers').

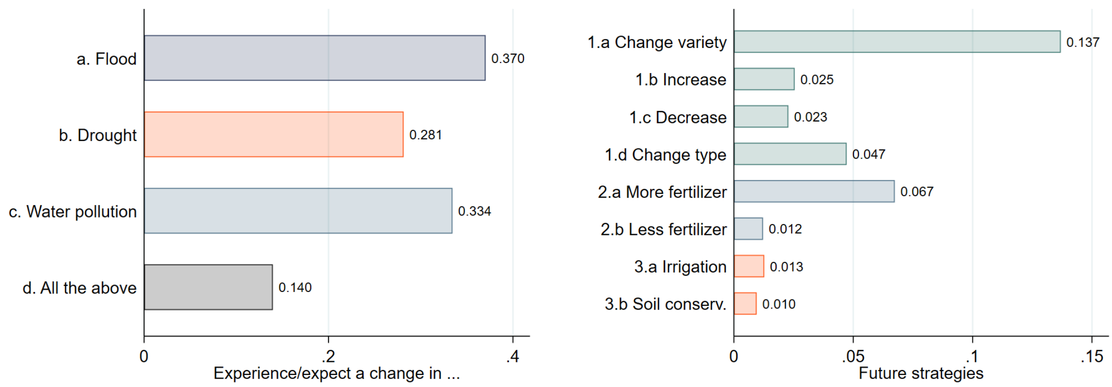


(a) Land quality (b) Fertilizers

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.

Subjective land evaluations Our land questionnaire in 2022 contains questions about land quality, suitability and requirements. More specifically, we ask about: land quality on a scale of 0 (not suitable for cultivation) to 5; crops that the respondent considers as being possibly cultivated on the land parcel—if the land parcel is suitable (land quality above 0), at least one crop should be selected; which crop the respondent has seriously considered growing on the land parcel; the need to use fertilizers, herbicides, or pesticides on the land parcel; whether the respondent is aware of the existence of soil testing kits and whether they would like their soil to be tested. We ask similar questions to the respondent about four land parcels within the village that are randomly-selected amongst 140 representative land parcels. We show in Figure A.5 a comparison between the own and other villagers' evaluations of land quality and land needs: the villagers are not in perfect agreement with each other or with the landowner, but we see a strong correlation across evaluations. Some land parcels are better than others, and this information is known to many villagers.

Figure A.6. Climate change and mitigation strategies.



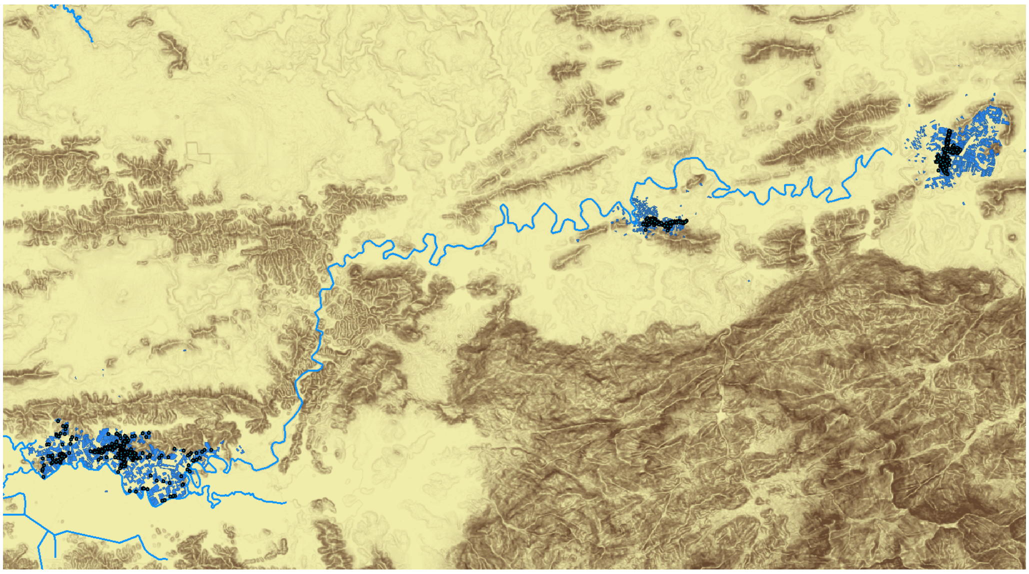
(a) Climate change (b) Mitigation strategies

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 considering a certain mitigation strategy.

Climate change and mitigation strategies We include a climate module to capture beliefs about climate change and mitigation strategies. Our questionnaire proceeds as follows: Have you experienced recent changes in [X] or do you expect [X] to increase/decrease/remain stable in the future? If so, when/where etc.? Which one of the following strategies [Y] do you expect to adopt in the future? Is adopting [Y] motivated by changes in agricultural conditions or environmental concerns? The events [X] cover droughts, flooding, and water pollution. The strategies [Y] include changing cropping patterns, changing the use of inputs, adopting more sustainable practices, etc. We summarize the most common answers in Figure A.6. Interestingly, the most common mitigation strategy is a change in crop variety.

Flood risk, elevation and flood hazard Our survey was originally designed to capture objective and subjective evaluations of flood risk. The hydro-physical modeling of flood hazard builds upon the global Flood Hazard Model (FHM) recently applied at a 90m resolution to develop global inundation probability maps (NERC grant NE/M007766/1, see Sampson et al., 2015) and currently used by the World Bank think hazard (http://thinkhazard.org/en/). In our application to the Central Highlands of Vietnam, the FHM is extended along three dimensions: the model simulates river dynamics using historical remotely sensed data and river gauge observations;[24] the model adds information on the location, volume and expected operation rules for the dams and reservoirs in Dak Lak in order to model how these structures affect discharge (Ty et al., 2011); the model finally relies on remote sensing data to infer currently unavailable information about the presence of levees and drainage channels.

Figure A.7. Slope (FloodAdaptVN project).



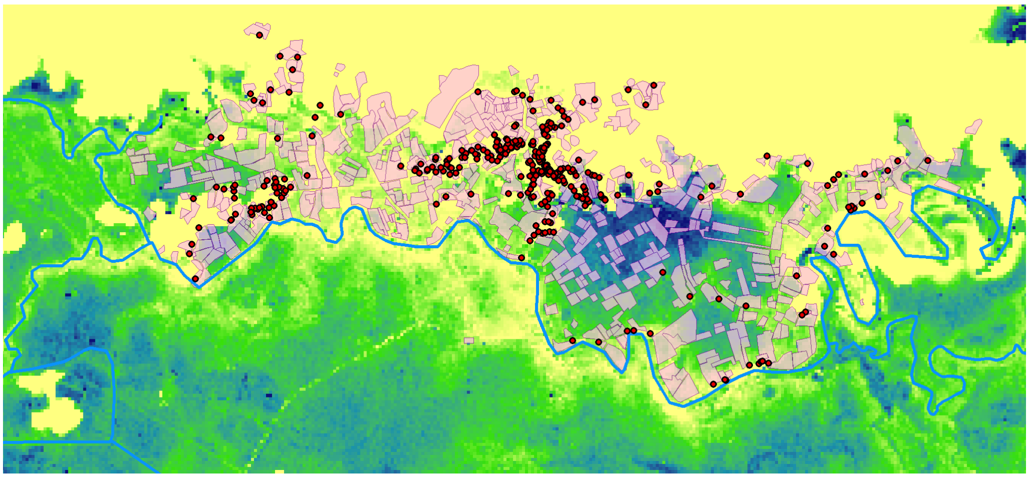
Notes: This map shows the average slope as computed from a 30m grid resolution raster. We show the location of three of our four villages: Village 1 (West), Village 2 (Center), and Village 3 (East).

The accuracy for parcel-level risk assessment crucially depends on the precision of the elevation data. Highly-precise floodplain topography limits measurement error associated with the modeling of fluvial inundations in lowland areas; high resolution and vertical accuracy also limits measurement error associated with the modeling of pluvial inundations for villages in hilly environments. In order to produce historical event simulations—past shocks—and a measure of vulnerability—flood disas-

[24]A model-interfacing tool to undertake a linking with models of river dynamics (GLOFRIM) was recently developed by Hoch et al. (2017). This tool automatically links output from different environmental models, including the PCR-GLOBWB hydrological model developed at Utrecht University and the LISFLOOD-FP hydrodynamic model developed in Bristol. PCR-GLOBWB simulates catchment runoff from atmospheric variables and is calibrated with river gauging data from the Vietnamese Hydrometeorological Data Center (HMDC). LISFLOOD-FP simulates all river routing and flood inundation processes in the Central Highlands from 1993 onwards.

ter risk—, we integrate highly-precise elevation data from the FloodAdaptVN project (https://floodadapt.eoc.dlr.de/) within the global Flood Hazard Model. The data come at a resolution of 30m and a vertical accuracy below 1 meter, and we correct for measurement error due to the vegetation cover. We illustrate the topography around three of our villages (Dak Lak, Central Highlands of Vietnam) in Figure A.7. None of our villages stand on the flank of a mountain; they are located in relatively low-land areas immediately surrounded by rugged terrain. A small but non-negligible number of land parcels are on such hilly terrain.

Figure A.8. Flood hazard in Village 1 (simulations based on 1-in-100 years events).



Notes: This map shows the dispersion of land parcels and deciles of flood risk in Village 1. 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.

We use the augmented Flood Hazard Model at a 30m resolution to construct measures of recent exposure to flooding and overall risk at the level of each agricultural parcel. We illustrate in Figure A.8 the spatial variation induced by topography and river dynamics across land parcels within Village 1. Flood hazard is computed from simulations based on 1-in-100 years events. Panel (a) shows that 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. Panel (b) shows a very different pattern induced by pluvial inundations with the most vulnerable parcels being at the flank of small mounts.

Identification of the household network The identification of the household network relies on: (i) a list of contacts with their name, age, gender, phone number (last 6 digits), and a description of their relationship with the different household members; (ii) references to these contacts when relevant along the questionnaire.

Enumerators were encouraged to establish a preliminary list and to update the list as the interview went along if new contacts were mentioned by the household. We reproduce below a translated summary of the guidelines/training provided to enumerators.

> In this section, we would like to learn about the household main contacts within the village. One difficulty of this exercise is the following: the precise questions that will define the notion of "contact" may be asked later (for instance, in Section 10). In this earlier section, you may need to anticipate those questions. If some persons have not been entered and come up as contacts in the following sections, you will need to go back to this section and add them. You can proceed as follows in order to optimize the procedure:
>
> - Ask which persons interact the most with the household for important issues and provide context: "From time to time, most people discuss important matters with other people. Looking back over the last year, who are the people with whom you discussed matters important to you? Who would be the people with whom you would discuss an important matter in the near future?" The first answers may be family, friends and immediate neighbors.
> - Once these general contacts have been entered, please collect additional names of villagers with whom the household had an economic exchange in the previous year. This would include villagers who are: (i) part of the same labor exchange group, (ii) involved in a large transaction with the household (e.g., land, house or truck/tractor), (iii) coworkers, hired labor or employers/employees, (iv) in a financial transaction with the household. These questions can be asked sequentially: "Are there other villagers than the one that you have mentioned who belong to your labor exchange group or who work with you?"; "Are there other villagers than the one that you have mentioned whom you are lending to or borrowing from?"; "Are there other villagers than the one that you have mentioned with whom you had a large transaction last year?".
> - A few additional remarks: one name may be mentioned in several occasions, please do not enter the same contact twice; please stop the respondent if you feel that he/she is providing far too many names of people that he/she does not know very well—in such case, repeat the initial question and insist that the relationship needs to be tight; you

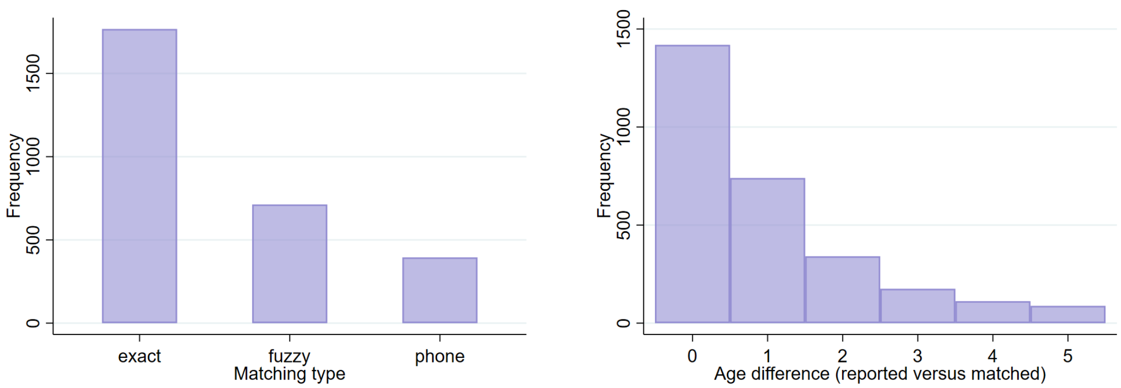
may help the respondent by providing reference periods—we are interested in people who are currently interacting with the household or who had a recent economic exchange with the household.

- Once all contact names are entered, go through the roster to enter the information for each contact separately.
- If new names are mentioned later in the questionnaire, please go back to this section and the contact information.

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; among unmatched entries, (ii) matching is then based on gender, age, and exact name matching; (iii) unmatched entries are finally matched through a fuzzy matching on names, accounting for specificities of the Vietnamese language (and frequent misspelling). The outcome of this matching procedure is about 2,900 linkages from about 4,000 reported contacts (a match rate of about 71%).

To better understand the quality of matching, we first report the match rate per village: 70% in Village 2, 67% in Village 3, 82% in Village 4, and 65% in Village 1. The difference between the last two villages illustrates a possible source of "false negatives": Village 4 is the most isolated village, while Village 1 is very close to other villages along the road; villagers may report names of contacts outside the scope of our household survey.

Figure A.9. Matching quality (method, and reported age difference).



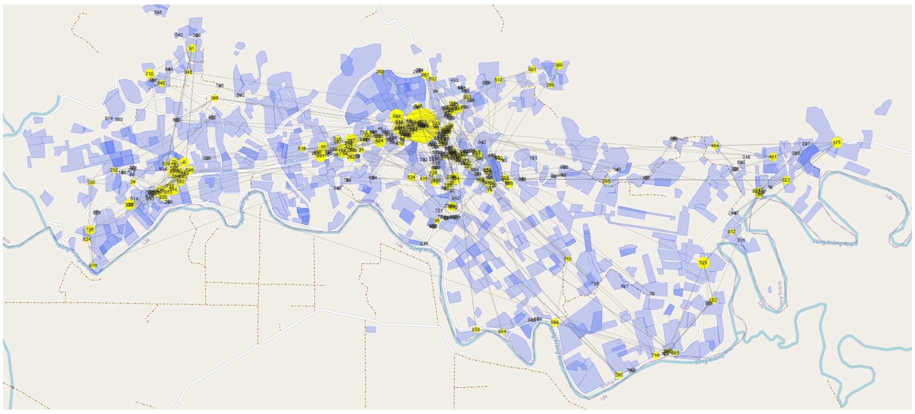
(a) Method (b) Age difference (reported versus actual)

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.

Figure A.10 provides an illustration of the structure of social networks in Village 1.

Social links follow spatial clusters to some extent. There are various explanations for this pattern: (i) households of a same extended family are more likely to live nearby (e.g., due to divided inheritance, land bequests, or to the land acquisition process); (ii) neighbors are more likely to form a labor exchange network; (iii) spatial proximity reduces the communication costs. Another marking observation is that a few households are central nodes within the village, either because of their political role (the village leader) or their economic occupation (the supermarket owners, often offering credit to villagers).

Figure A.10. An illustration of the social network in Village 1.



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: households of a same extended family are more likely to live nearby, neighbors are more likely to form a labor exchange network, and spatial proximity reduces the communication costs. A few households are central nodes within the village (the village leader, the supermarket owner).

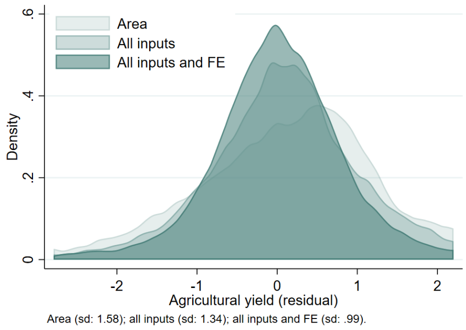
A.2 Descriptive statistics

In this section, we provide complements to Section 2.2 (agricultural practices) and Section 2.3 (social networks).

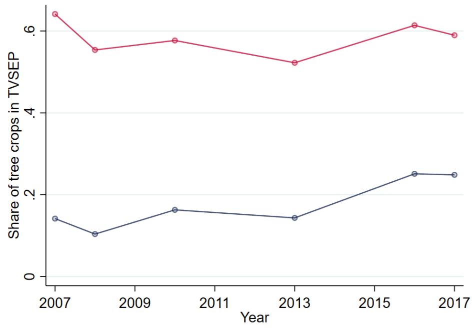
The productivity gap across agricultural commodities In Section 2.2, we document an agricultural productivity gap across farms and the importance of high-return perennial crops in explaining such a gap.

We replicate the exercise in Figure A.11 using data from the Thailand Vietnam Socio Economic Panel (TVSEP). We find that there is a large dispersion in agricultural productivity across farms (panel a), that tree crops become more prevalent over time, even

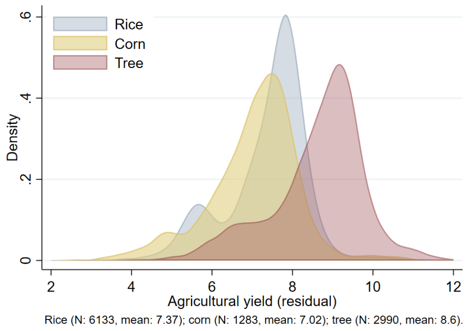
Figure A.11. The agricultural productivity gap and the tree premium in TVSEP.



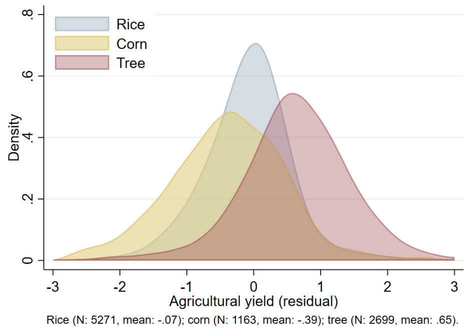
(a) Agricultural productivity gap



(b) Adoption of tree crops



(c) Returns to crops



(d) Agricultural TFP

Notes: Panel (a) shows the distribution of agricultural TFP in TVSEP, $\ln z_{ic}$, when controlling for: area as the only input; all inputs (area, labor, intermediary, capital); and all inputs and farmer fixed-effects. Panel (b) shows the incidence of tree production across agricultural parcels over time (as weighted by the number of land parcels in blue, or by the value of agricultural production in red). Panel (c) shows the crop-specific distribution of agricultural TFP when controlling for area. Panel (d) shows the crop-specific distribution of agricultural TFP when controlling for all inputs.

though adoption remains limited (panel b), and that there are large differences in agricultural yields across crops (panels c and d). We find very similar differences in agricultural yields, a "tree premium", in TVSEP as in our own survey.

Determinants of land quality One interesting aspect of our survey is to collect a subjective assessment of the farmer's own land parcels, their quality, suitability to grow certain crops, and input requirements. We can use these evaluations, which are visibly shared to some extent by other villagers (see Figure A.5), to learn about the determinants of higher or lower land quality.

In Table A.1, we regress the land quality index (from 0 to 5) on various soil (objective) characteristics, beliefs of the farmer about climate risk and social integration of the farmer within the village. First, we find that soil characteristics (bulk density, carbon content or slope) are strong predictors of land quality: a high-quality land should have

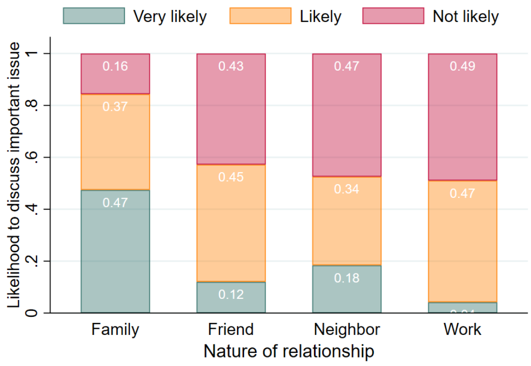
Table A.1. Explaining land quality evaluations.

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
				(0.1298)
Observations	2,406	2,406	2,406	2,406

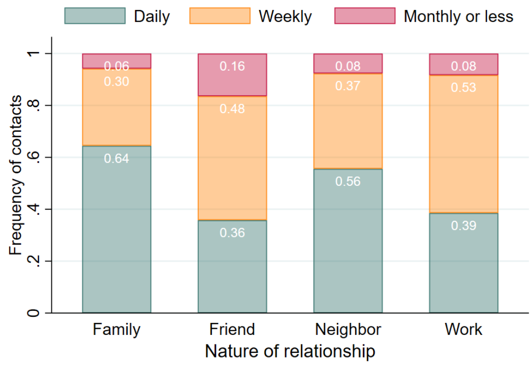
lower bulk density, higher carbon content and be less rugged. Second, soil composition (acidity, phosphate, potash) also has some predictive power. Third, higher land quality positively correlates with the identified risk of water pollution from farmers. The explanation most likely relies on the nature of high-quality land parcels: those are typically irrigated parcels, whereby fertilizer, pesticide, herbicide usage from other farmers would generate spillovers. Fourth, social integration in the village appears to play a role: the number of social links and the average trust correlate positively with land quality. One reason could however be underlying characteristics of the farmer, e.g., more or less optimistic in general.

The structure of local networks We document in Section 2.3 the motivations, origins, and strength of social links within villages. In Figure A.12, we provide further evidence about the strength of relationships as a function of their initial nature (whether peers know each other through family ties, spatial proximity, etc.). Note that these categories are not mutually exclusive, such that a peer can be a neighbor and a friend. We find that important issues are much less likely to be discussed with “work colleagues” and much more likely to be discussed within families (panel a); this pattern is the same

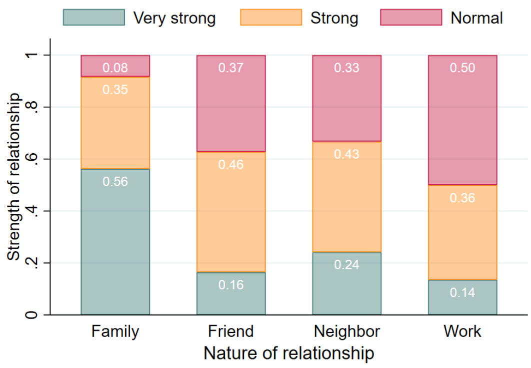
Figure A.12. Network "usage": motivation and strength of links.



(a) Discuss subjects



(b) Frequency



(c) Strength

(d) Trust

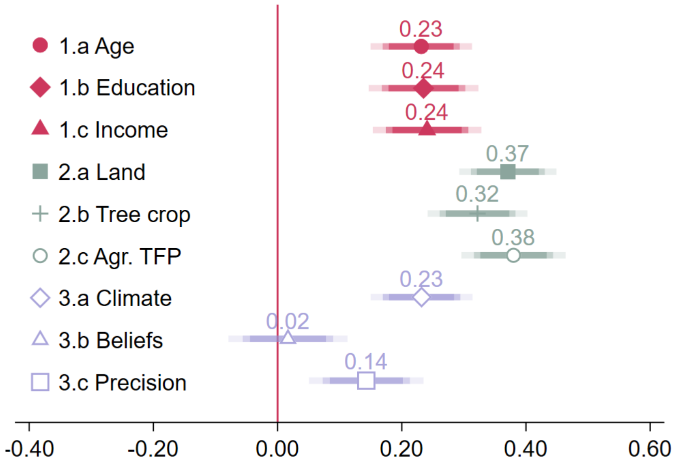
for a subjective evaluation of the relationship strength (from 0 to 10, see panel c). Families and neighbors interact very often (respectively 64 and 56% of family-based links and neighbor-based involve daily interactions). Finally, family links are characterized by a higher level of trust between contacts (panel d). In summary, family ties are very strong; our study however ignores those ties and identifies social multipliers from the weak(er) links between neighbors.

The second important observation made in Section 2.3 is that there is a high degree of homophily within the social network. In other words, there is a high correlation between the respondent's characteristics and activities and that of their peers. For each characteristics x_i of a respondent i, we define the "multilateral" peer characteristic as follows,

$$X_i = \frac{\sum_{m=1}^{3} \sum_{j \in \mathcal{L}_i^m} \delta^m x_j}{\sum_{m=1}^{3} \sum_{j \in \mathcal{L}_i^m} \delta^m}.$$

The multilateral characteristics is a weighted average of the peers' characteristics where

Figure A.13. Homophily (multilateral computations).



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.

we apply a discount $\delta = 0.5$ to higher-order linkages. We report the correlation between x_i and X_i in Figure A.13 for household characteristics (age, education, income), agricultural practices (land holdings, the growing of high-return perennial crops and agricultural Total Factor Productivity cleaned for input/factor usage), and beliefs (about climate change and about land quality/requirements for other land parcels in the village). We find that linked households are closer in demographics, in agricultural practices and even in their beliefs about climate change (3.a) or their capacity to evaluate the land parcels of others in the village (3.c).

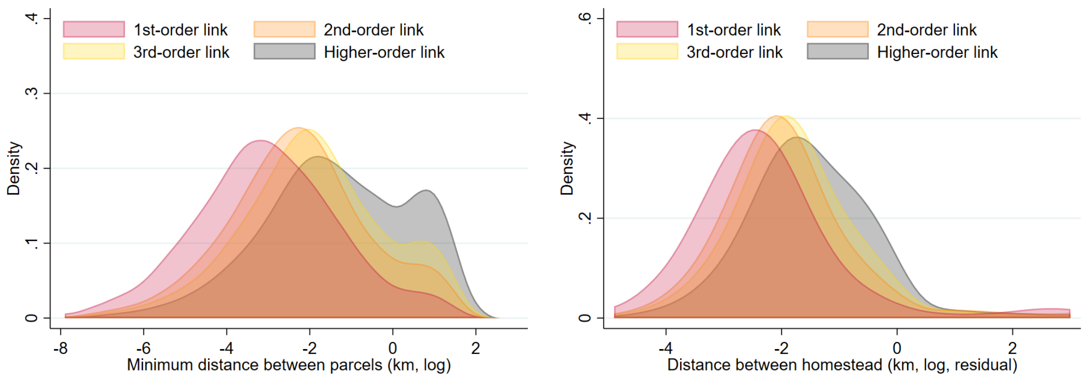
B Complements to the empirical analysis

This section provides complements to the empirical analysis. More specifically, we discuss identification (see Section 3 of the paper), and we provide a sensitivity analysis of our main results (as briefly summarized in Section 4.2).

B.1 Identification

We first shed additional light about the role of residential proximity in fostering (exogenous) social links.

Figure B.1. Network links and spatial proximity (land parcels, and homes).



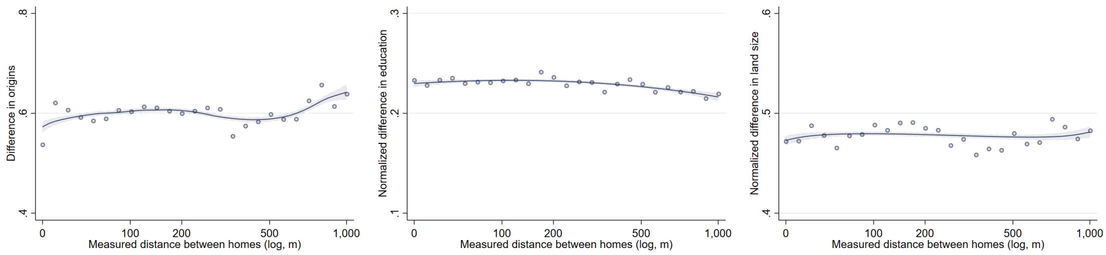
(a) Land parcels (b) Homes (contr. for land)

Notes: Panel (a) shows the distribution of the minimum distance between land parcels for links of different orders; panel (b) shows the distribution of the minimum distance between homes for links of different orders, conditioning on the minimum/average/maximum distances between land parcels.

In Figure B.1, we display the distribution of distances between homes and land parcels of two households connected by: a first-order linkage (red), a second-order linkage (orange), third-order linkage (yellow), and a higher-order linkage (blue). Panel (a) displays the minimum distance between non-residential land parcels; and panel (b) shows distribution of distances between homes, once conditioning on the distance between non-residential land parcels. We see that both the proximity between land parcels and residential proximity foster more direct connections between pairs of households. Panel (b) illustrates that residential proximity is very predictive, even when controlling for the spatial distribution of other land parcels. This is the first argument of our empirical strategy: residential proximity predicts social links.

The second argument of our empirical strategy is that residential proximity predicts exogenous social links because it reflects the timing of arrival in these repopulated vil-

Figure B.2. 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).

lages. We illustrate the role of this factor in Section 3.1: neighbors are more likely to be part of the same settlement wave. In Figure B.2, we show that home proximity does not generate further (observable) resemblance between households: they are as likely to be of the same origins (province within Vietnam) as any other random pairs of households (panel a); the normalized differences in education (panel b) or land (panel c) are that of any other pairs of households.

Finally, we provide further evidence that residential proximity is associated with a much lower auto-correlation in agricultural practices than agricultural proximity. In Figure B.3, we nest the adoption of a high-return perennial crop at the level of a parcel and the average across all parcels owned by a household at the level of homesteads. We see that land use is very correlated across land parcels that are scattered around the different villages. By contrast, residences are concentrated along a few streets and display far less homophily in treatment.

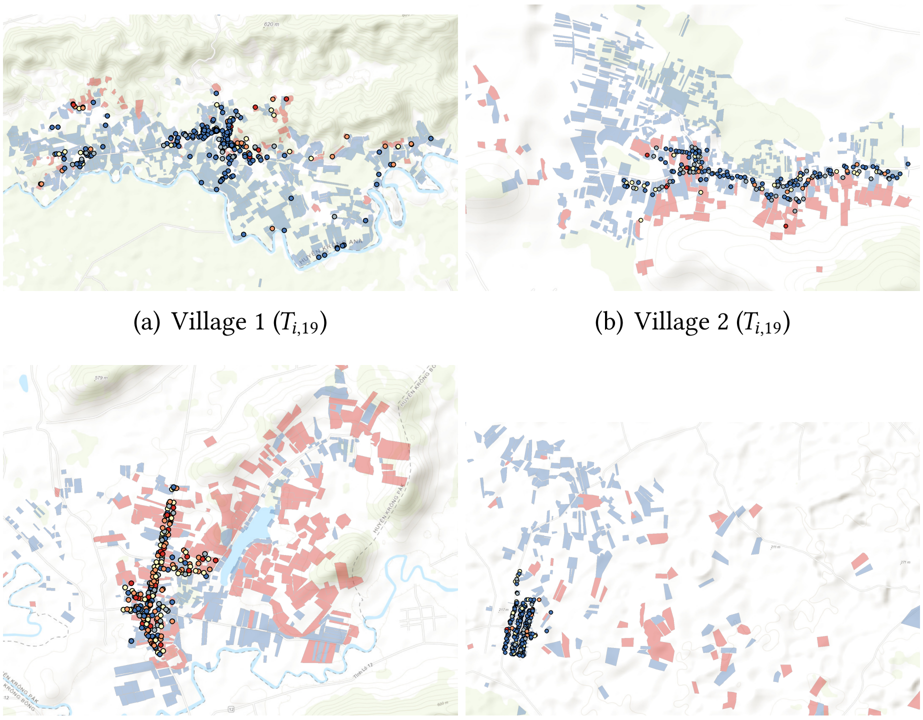
B.2 Robustness checks

This section provides a sensitivity analysis of our baseline results.

The return to local networks in crop adoption In Section 4.1, we estimate Equation (2) through a 2SLS specification with the proximity-weighted exposure to treatment as the instrument.

We report in Table B.1 the OLS counterpart to Table 3. We find an additional standard deviation in exposure (ϑ_i^1) is associated with a 0.015 higher likelihood to grow a high-return perennial crop. This estimate is about a sixth of the 2SLS estimate. One interpretation for these differences is the correlated nature of agricultural practices on

Figure B.3. Distribution of treatment across villages.



(a) Village 1 ($T_{i,19}$) (b) Village 2 ($T_{i,19}$)

(c) Village 3 ($T_{i,19}$) (d) Village 4 ($T_{i,19}$)

Notes: This Figure shows the treatment $T_{j,19}$ nested at the land parcel level and nested at the "homestead" level across our four villages.

both sides of a network edge: strongly-linked households are more likely to have similar cropping patterns in 2019, such that the average network link is unlikely to trigger some changes between 2019 and 2022. The instrument however hinges on weaker links, featuring starker differences in agricultural practices across households.

In Table B.2, we probe the sensitivity of our findings to additional controls. More specifically, we consider our preferred baseline estimate (column 3 of Table 3) and further condition for: land tenure in column (1); soil composition in column (2); a subjective evaluation of land quality from the respondent in column (3); and household characteristics in column (4). The main estimate remains between 0.103 and 0.111.

Our baseline specification hinges on weighted measures of exposure to the treatment where: the latter is defined in 2019, we exclude family links, and the instrument only considers households within 100 meters of each other's homes. In Table B.3, we consider variations of the main exposure measure and instrument: we consider an exposure measure which includes family links in column (1); we construct the instrument as an inverse-distance weighted measure of exposure to the treatment in column (2); and we calculate both exposure and instrument using the most recent allocation of treat-

Table B.1. 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).

Table B.2. 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 3. 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).

ment (2022) in column (3). Adding family links slightly reduces the estimate, possibly reflecting the homophily in agricultural practices observed across family ties. The other specifications deliver similar results to our baseline.

Table B.3. 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 3. 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).

Table B.4. A pseudo-panel approach—first-stage estimation.

Exposure	$\vartheta_{in}^0 - \vartheta_{in-1}^0$	$\vartheta_{in}^1 - \vartheta_{in-1}^1$	$\vartheta_{in}^2 - \vartheta_{in-1}^2$
Proximity instrument	0.244	0.059	
	(0.012)	(0.016)	
Indirect linkages			0.122
			(0.030)
Observations	23,569	20,436	7,306

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).

A pseudo-panel approach We now provide complements to the pseudo-panel approach discussed in Table 5. We first show the first-stage estimates in Table B.4 for our three different measures of exposure (and two distinct instruments). First, a time-varying increase of additional standard deviation in predicted exposure through residential proximity increases weighted exposure through the network, $\vartheta^0_{in} - \vartheta^0_{in-1}$, by 0.24 standard deviations (column 1). The same increase raises first-order exposure to the treatment, $\vartheta^1_{in} - \vartheta^1_{in-1}$, by 0.06 standard deviations (column 2). Second, our indirect-linkage instrument—two households have a friend in common for two *distinct reasons*—predicts second-order exposure to the treatment, $\vartheta^2_{in} - \vartheta^2_{in-1}$ (column 3).

Table B.5. A pseudo-panel approach—timing of adoption.

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

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).

One advantage of the pseudo-panel is to better exploit the timing of adoption against the timing of network formation. In Table B.5, we replicate the exercise performed in Table 5, but we add leads and lags of the dependent variable—thus adding leads and lags of the respective instruments. Our first-stage keeps sufficient statistical power in columns (1) and (2), but not in column (3)—we will thus ignore the estimates for the second-order exposure to the treatment. We find that "contemporary" changes in exposure to the treatment affects adoption within the same time frame. Importantly, the forward estimates are smaller and less precisely estimated.

Table B.6. 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 × 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).

In our last robustness checks, we explore treatment heterogeneity across periods of interest and tenure in the village. First, treatment adoption among existing parcels is higher after 2006 than before. For instance, the probability for each (existing) parcel to be converted to high-return perennial cropping between two waves (three-year period) is about 0.01 before 2006 versus 0.024 afterwards. In the first two columns of Table B.6, we consider specification (2) in Table 5 and divide the sample between 1980–2006 and 2006–2022. We find similar multipliers across the two periods. Second, tenure within the village might play a role in the pace of adoption. We interact the first-order exposure ($\vartheta^1_{in} - \vartheta^1_{in-1}$) with a dummy equal to 1 if the household has arrived within 15 years of wave n (15 years is the median across our different waves). We find that the social multiplier is higher for more recently-arrived households. Nonetheless, our estimate is robust to excluding the very fresh arrivals (e.g., before 5 years, with a coefficient of 0.041 and a standard error of 0.019).