Endogenous Information Sharing and the Gains from Using Network Information to Maximize Technology Adoption

Manzoor H. Dar, Alain de Janvry, Kyle Emerick, Erin M. Kelley and Elisabeth Sadoulet*

December 24, 2018

Abstract

Can agents in a social network be induced to obtain information from outside their peer groups? Using a field experiment in rural Bangladesh, we show that demonstration plots in agriculture — a technique where the first users of a new variety cultivate it in a side-by-side comparison with an existing variety — facilitate social learning by inducing conversations and information sharing outside of existing social networks. We compare these improvements in learning with those from seeding new technology with more central farmers in village social networks. The demonstration plots — when cultivated by randomly selected farmers — improve knowledge by just as much as seeding with more central farmers. Moreover, the demonstration plots only induce conversations and facilitate learning for farmers that were unconnected to entry points at baseline. Finally, we combine this diffusion experiment with an impact experiment to show that both demonstration plots and improved seeding transmit information to farmers that are less likely to benefit from the new innovation.

*Dar: International Rice Research Institute, New Delhi India, m.dar@irri.org; de Janvry: University of California at Berkeley, 207 Giannini Hall, Berkeley, CA 94720-3310, alain@berkeley.edu; Emerick: Tufts University, 8 Upper Campus Road, Medford, MA 02155-6722, kyle.emerick@tufts.edu; Kelley: University of California at Berkeley, Agricultural and Resource Economics, erinmkelley@berkeley.edu; Sadoulet: University of California at Berkeley, 207 Giannini Hall, Berkeley, CA 94720-3310, esadoulet@berkeley.edu. We acknowledge financial support from the Standing Panel on Impact Assessment of the CGIAR and from USAID through the Feed the Future Innovation Lab for Assets and Market Access (AID-OAA-L-12-00001). The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the US Government. Emerick is grateful to the Institute of Economic Development at Boston University where he was a visiting scholar while part of this research was carried out.

1 Introduction

People rely on their peers for information. Research has established the existence of such peer effects across a variety of domains, ranging from learning in school to adoption of new innovations in poor countries.[1] Building on the importance of networks for transmitting knowledge, recent work has sought to answer the question who should be chosen as the initial recipients of information in order to make that information proliferate throughout the network? The proven methods for selecting these optimal entry points or "seeds" include collecting the full network in a community (Beaman et al., 2015) or asking a smaller number of individuals for the right people to spread information (Banerjee et al., 2018b). These approaches rely on the structure of the social network being roughly fixed and stable over time. Typically, agents pass information only amongst their connected peers.

Another strategy — and one that has received less attention in the literature — is to consider the exchange of information between individuals without existing social ties, and ask what can be done to encourage individuals to seek out information from outside their networks? Is it possible for the policymaker to intervene in a way that encourages people to seek information from new sources? Or, are information-sharing relationships sufficiently rigid to necessitate the selection of optimal seeds within the network? In this paper we focus on the efficacy of a commonly used technique in agricultural extension, namely the use of demonstration plots that showcase the features of a new seed variety relative to traditional varieties. We find evidence suggesting that this low-cost approach to encouraging learning can effectively substitute for the sometimes difficult task of finding the optimal entry points in the network.

We arrive at this conclusion using two related experiments spread across 256 villages in rural Bangladesh. The first experiment contrasts the two approaches detailed above to spreading information about a new rice variety called BRRI Dhan 56 (or BD56 for short).[2] We introduced the new variety to five farmers, referred to as "entry points" throughout the remainder of the paper, in a random subset of 192 villages. We then cross randomized villages across two treatments: (1) the selection criteria for entry-point farmers and (2)

[1] A non-exhaustive subset of research in this area includes peer effects on academic performance (Sacerdote, 2001), purchases of financial assets (Bursztyn et al., 2014), adoption of improved sanitation in developing countries (Guiteras, Levinsohn, and Mobarak, 2015), the decision of whether to purchase crop insurance (Cai, de Janvry, and Sadoulet, 2015), and the adoption of agricultural technology (Foster and Rosenzweig, 1995; Munshi, 2004; Bandiera and Rasul, 2006; Conley and Udry, 2010).

[2] BD56 necessitates a large change in the production process. In short, the variety is ready for harvest a full month before the common variety being grown in the rainy season. As a result, the technology provides enough time for a third crop to be grown in between the rainy and dry-season rice crops. This large change in the cropping system — going from producing two rice crops to producing a third crop in between — enhances the potential for social learning about BD56 and its attributes.

the demonstration method, aimed at encouraging other farmers from the village to seek information. The different treatment cells are described visually in Figure 1.

In terms of the selection of entry points, we randomized villages across three different selection methods. In the first 64 villages we randomly selected farmers (as a benchmark). In the next 64 villages we relied on the local knowledge of agricultural extension officers (known as "sub-agricultural officers" or SAO's) to identify farmers who would be most effective at demonstrating the features of a new rice variety. In the remaining 64 villages, we ranked farmers according to farm size and selected the five largest farmers.[3] We opted for these selection mechanisms because they can be implemented at relatively low cost, which is attractive from a policy perspective. Indeed, neither approach requires an expensive survey nor detailed data collection efforts.

In the second treatment, we randomized whether farmers were asked to set up demonstration plots or not. This treatment was cross randomized with the selection of entry points detailed above. This means that 32 villages out of the 64 assigned to a particular "entry-point" selection arm were also assigned to demonstration plots. There the team assisted farmers in setting up "head-to-head demonstrations", which involved cultivating BD56 alongside a counterfactual seed variety of the farmer's choosing. We provided two markers to make the comparison across the two varieties more visible — one reading "BD56" and the other listing the name of the chosen alternative variety — to be placed in the two fields (see Figure 2 for an example).[4] We refer to this treatment as "demonstration plots" throughout the remainder of the paper. The demonstration plots indicate to other farmers that the entry point is comparing the attributes of the new variety to a known one. This serves as a mechanism to focus the attention of other farmers on a new learning opportunity.[5] In the remaining 32 villages that were not assigned to demonstration plots, the entry points used BD56 on one of their plots and were provided with a single "BD56" marker. It follows that the demonstration plots must do more than broadcast information about the existence of BD56 because farmers in these comparison villages also labelled their field with a marker.

These two approaches to increasing knowledge transmission rely on very different assumptions about how information diffuses in networks. The improved selection of entry points, using either large farmers or those hand picked by SAO's, seeks to test scalable approaches to

[3] We found during piloting that large landholders often act as opinion leaders during focus group discussions about agriculture. Moreover, other farmers seemed to look to large farmers to learn about new technology. These casual observations led to the inclusion of the arm where farm size was used to target entry points.

[4] This method of demonstration plots is common, especially by private sector seed companies seeking to promote and demonstrate the attributes of their new seed varieties.

[5] We didn't provide any further assistance (such as inputs or advice) with the actual cultivation of the two plots. This was purposeful to lessen the costs of the demonstration plots treatment.

finding the right influential farmers in the network. Importantly, the notion of optimal seeds is designed to exploit the network as it exists at baseline, and does not consider that the intervention itself may change the network. In contrast, the demonstration plots are meant to spark interest and potentially induce communication beyond existing network links. The demonstration plots can do so by either 1) capturing attention and thus increasing the demand for information or 2) providing a more precise signal to entry points and therefore increasing their supply of information to others.

This experiment delivers four main results. First, and using only baseline information, we verify that the large and SAO-selected entry points are far more central in the village networks, and are thus well positioned to spread information to a larger number of farmers. At baseline our survey teams visited all farming households in each village — making a total of almost 22,000 visits — and posed the question "which farmers in this village do you regularly speak to about rice farming?"[6] Using these data, we observe that the average entry point in the random villages is connected to 4.6 other farmers. This increases sharply to 8.2 and 9.1 connections for entry points in SAO and large farmer villages, respectively. Similarly, the eigenvector centrality of entry points increases by 47 percent under SAO selection and 80 percent with large-farmer selection. Based solely on the network measured at baseline, both treatments would therefore be expected to increase the spread of knowledge.

Despite these noticeable differences in network centrality, our second result is that demonstration plots with random farmers create just as much additional knowledge as entering with large and SAO-selected farmers. After the harvest of BD56 — and the sowing of the next crop — we conducted a survey with 10 random farmers in each village. In addition to knowledge about the existence of BD56 and its basic attributes, we also collected information on reported conversations about BD56. Using large farmers as entry points increases awareness by 7.4 percentage points or 12.3 percent in villages without demonstrations. Similarly, entry points selected by extension agents increase awareness by 6.7 percentage points (11.2 percent) in villages without demonstrations. However, providing seeds to more central farmers provides no additional benefits for knowledge diffusion when we introduce demonstration plots.

At the same time, setting up demonstration plots increases awareness by 7.2 percentage points (12 percent) with random entry points. Noticeably, the impact of demonstration plots under random selection is the same magnitude as the impacts of improved entry-point selection in the absence of demonstration plots. The results remain similar when looking

[6] Chandrasekhar and Lewis (2016) show that measures of network centrality are misleading when estimated using only a sample of nodes within the network. Our approach of fully characterizing the network by sampling each household in the village alleviates this concern.

at the number of reported conversations about BD56: demonstration plots with random farmers induce conversation by about the same amount as introducing seeds with more central farmers. Given the ease of setting up demonstrations, i.e. simply placing two signs in adjacent fields, our experiment offers insight into how interventions that attract attention to the fact that there is something to be learned can substitute for seeding with more central entry points in networks.

Third, we then move on to investigate why the demonstration plots are effective. The results are consistent with the idea that demonstration plots cause farmers to learn from people outside their network. Focusing on the 64 villages with random entry-point selection, we observe strong peer effects on knowledge. Farmers that were randomly connected to an additional entry point are 7.7 percentage points more likely to know about BD56. This average effect differs meaningfully between villages with and without demonstration plots: an additional connection with an entry point has no effect in demonstration villages, but it increases awareness by 13.5 percentage points (22 percent) in non-demonstration villages. In other words, the demonstration plots lead to information exchange outside of baseline networks and therefore greater transmission of knowledge. In doing so, the demonstration plots entirely eliminate peer effects. The same pattern of results again appears when considering conversations about the technology.

Also consistent with this network interpretation, we find that demonstration plots were most effective for the farmers that are the least connected in the baseline information network — where connectivity is measured by eigenvector centrality. A plausible explanation of this result is that the demonstration plots induced these less connected farmers to endogenously seek information about BD56.

Finally, we show that their network centralities at least partly explain why large and SAO-selected entry points lead to better information diffusion. The effects of large and SAO selection on knowledge decrease by 43 and 31 percent, respectively, when conditioning on the average degree of the entry points. In addition, the average degree centrality of entry points is itself positively correlated with BD56 awareness. Conditioning on average degree is not a perfect test — since the particular network measure that should "knock out" the treatment effects depends on the specific model of diffusion.[7] Nonetheless, the result is consistent with the idea that diffusion via network links partly explains why the entry-point treatments were successful.

While we mainly consider effects on knowledge transmission, we also observe uptake

[7] Degree is a suitable measure if we think of diffusion models with few periods, the probability of passing information to connected friends is high, and the farmers with the largest degrees are sufficiently spread out in the network.

when BD56 was made available for sale at subsidized prices. The results on seed adoption are noisier, but qualitatively consistent with our observations on knowledge.

Our second experiment allows us to further test whether demonstration plots (or alternative seeding strategies) are more likely to deliver information *to the farmers most likely to benefit from the new technology.* Returning to Figure 1, we also randomly selected 64 control villages where we provided a new long-duration rice seed to a set of farmers identified using the same criteria as in the 192 BD56 villages. This experiment allows us to characterize the impact of short-duration rice on agricultural practice and profits. In particular, the main benefit of the short-duration seed is the ability to grow a third crop in between the rainy and dry-season rice crops. Using the machine learning methods developed in Chernozhukov et al. (2018), we estimate a mapping between baseline covariates and the treatment effect of BD56 on the number of crops grown. This heterogeneity index serves as a prediction of which farmers are the most likely to benefit from BD56 by increasing cropping intensity.

Using this heterogeneity index with the 192 villages in our first diffusion experiment, we find suggestive evidence that both demonstration plots and our selection treatments increase knowledge and conversations only for the farmers that have *below-median* expected treatment effects of BD56 on the number of crops grown. In other words, intervening to increase knowledge diffusion in networks may only be effective for the farmers with lower expected benefits from adopting an innovation.[8] These findings point to an important consideration for research that studies alternative mechanisms for increasing diffusion of products that have heterogeneous benefits. Combining diffusion experiments with standard impact evaluations allows the researcher to estimate treatment effect heterogeneity and use that to measure which diffusion strategies reach the people most likely to benefit from an innovation, even without a strong prior on which observables drive the heterogeneity.[9]

We then go on to show how a simple diffusion model, when amended to allow for formation of new links with entry points, explains our pattern of results. In the model farmers can either become informed by receiving information flowing through the structure of existing links, or by actively communicating to "form new links" with entry points. Farmers in the worst position to learn from the network, i.e. those that are the least connected to entry points, benefit from having the opportunity to engage directly with entry points. The

[8] An alternative explanation is that the diffusion treatments succeeded in informing farmers that were the least likely to benefit from BD56, and therefore prevented their adoption. This explanation is less consistent with the positive (but noisy) point estimates we observe when estimating the effects of the diffusion treatments on seed adoption.

[9] Rigol, Hussam, and Roth (2017) is the closest example where the authors use machine learning methods to estimate treatment effect heterogeneity for microfinance in India. They then show that using community information on the returns to microfinance is more effective than the machine learning algorithm when applied to the set of observables in their baseline data.

demonstration plots treatment appears to deliver these benefits by inducing unconnected and more isolated farmers to learn directly from entry points.

Turning to literature, theoretical work has considered information transmission when network communication is endogenous (Acemoglu, Bimpikis, and Ozdaglar, 2014; Calvó-Armengol, Martí, and Prat, 2015), but empirical work in this area is more scarce. Mobius, Phan, and Szeidl (2015) show that having conversations is correlated with possessing the correct information in a field experiment amongst college students, but that frictions still exist in the diffusion of information. Chandrasekhar, Golub, and Yang (2016) make an important contribution by considering one particular friction arising with endogenous social learning: the need for information can reveal low skill, therefore creating a stigma effect that represents part of the costs of information seeking. Banerjee et al. (2018a) show evidence consistent with this same idea during India's recent demonetization. In particular, villagers were less willing to seek information about demonetization rules when everybody knew that information was provided widely and thus seeking information signals an inability to process one's own information. Our experiment delivers insights on the potential to intervene to induce communication, overcome some of these frictions, and therefore facilitate social learning.

Looking for ways to induce communication and learning builds on an active literature that treats networks as fixed and asks how to find the optimal entry points within these networks. These studies include demonstration of agricultural inputs (Beaman et al., 2015; Beaman and Dillon, 2017), diffusion of information about microfinance (Banerjee et al., 2013), the diffusion of health products (Kim et al., 2015), and information on how to capitalize on a financial opportunity or the uptake of vaccines (Banerjee et al., 2018b).[10] There are a couple of limitations with this approach. First, it can be difficult to identify the most central entry points in a social network without fully eliciting the network, which may be cost prohibitive.[11] Second, the efficacy of finding better entry points for seeding information likely depends on the underlying structure of the social network or the specific model of diffusion (Centola, 2010; Valente, 2012; Golub and Jackson, 2012). Akbarpour, Malladi, and Saberi (2018) show that in many network structures the benefits from seeding information with a slightly larger number of agents outweigh the benefits of identifying the most central individuals. Our experiment considers scalable methods of entry-point selection as a benchmark and shows that their benefits can also be obtained by interventions that trigger social learning.

[10] In addition to selection, other work has looked at different ways of making entry points communicate more, including compensation (BenYishay and Mobarak, 2015) and training (Kondylis, Mueller, and Zhu, 2017).

[11] Banerjee et al. (2018b) show how to overcome this difficulty by asking a sample of villagers who are the important people for diffusing information.

Focusing on agricultural development and policy, learning frictions are one of the frequently proposed reasons why farmers do not adopt new technology.[12] Agricultural extension serves as the policy tool to improve learning. In practice, the standard method of agricultural extension attempts to leverage social learning by seeding information with contact farmers (entry points) and relying on information diffusion through social networks (Birkhaeuser, Evenson, and Feder, 1991; Anderson and Feder, 2007; Kondylis, Mueller, and Zhu, 2017). Our experiment provides evidence on how demonstration plots can offer a substitute for the policymaker when improved selection of these contact farmers is difficult to implement.

The remainder of this paper is organized as follows. Section 2 discusses the design and implementation of the experiment, data collection, and basic characteristics of the sample. Section 3 discusses each of our individual results, focusing on how selection and demonstration plots influence learning, and on understanding what drives the effectiveness of the two approaches. Section 4 outlines one simple theoretical framework that is consistent with our results. We then provide an overview and discuss the implications of the findings in the final section.

2 Overview of the Experiment

In this section we review the details of the randomized control trial: the sampling strategy, the experimental design, and the data collection activities. We conducted the study in 11 sub-districts (upazilas) scattered across 3 districts of Rajshahi division, consulting with the Department of Agricultural Extension to identify upazilas that were suitable for the rice variety being introduced.[13] We randomly selected 20% of the villages with no more than 150 households, resulting in a final sample of 256 villages. This includes the 192 villages that received the new BD56 rice variety in the diffusion experiment, as well as 64 control villages that received the longer duration rice variety for the impact evaluation experiment. This village-level randomization was stratified by upazila.

BD56 has two key features. First, it requires less water, allowing farmers to save on supplemental irrigation fees and preserving groundwater resources. Second, it matures approximately 25 days earlier than other varieties commonly planted in the area, providing farmers with the option of harvesting and selling an additional crop between the two rice seasons (Aman and Boro rice seasons last from June/July to October/November and December/January to April/May, respectively). The 64 control villages received a long duration

[12]In addition to learning, numerous studies highlight a wide range of explanations, including behavioral biases, profitability, and risk (Duflo, Kremer, and Robinson, 2011; Suri, 2011; Karlan et al., 2014; Emerick et al., 2016; Cole, Giné, and Vickery, 2017).

[13]See Figure A1 for the location of the 11 upazilas included in the study.

rice variety called BRRI Dhan 51 (BD51) which was chosen due to its similarity to the most popular variety at baseline.[14]

2.1 Experimental Design

Figure 1 summarizes the experimental design for the diffusion and impact evaluation experiments detailed below. The design for the diffusion experiment consists in the cross-randomization of two dimensions of treatment across 192 villages: the method used to select the five seed recipients, and the implementation of "demonstration plots" by these entry-point farmers.

For the first dimension, the villages were subdivided into three groups of 64 villages each. The types of farmers we selected to receive the seeds (the "entry points") differed based on the specific treatment arm that village was assigned to. In the first group of 64 villages, the seeds were distributed to five farmers selected at random. In the second group, we ranked farmers by landholding sizes and distributed 5kg minikits of BD56 seeds to the top five farmers. In the third group of 64 villages, we asked the Sub-Agricultural Officer (SAO) to identify five farmers in the village that would be effective at demonstrating the new variety and provided BD56 seeds to them.

Within each group of 64 villages, we then selected 32 villages to receive additional assistance setting up demonstration plots. We included this additional dimension in the experiment to determine how this approach to boosting diffusion rates compared with the alternative of selecting entry points that rely on the underlying social networks to spread information. To this end, we asked farmers to select a counterfactual variety that they would like to plant beside the new BD56 seeds we distributed. We then provided two sticks to farmers: one with the name of the new variety (BD56), and another with the name of the variety they had selected to plant beside it. We asked that farmers keep these two signs in their fields throughout the cropping season to showcase the performance of BD56 relative to the counterfactual they selected. Farmers in the remaining 32 "non-demonstration" villages received a single sign for their BD56 plot. The provision of the single sign ensures that any effect we detect in the demonstration plot villages with two signs goes beyond the attention effect of placing one sign in the field.

For the impact evaluation experiment, we selected up to 15 "counterfactual" farmers to receive equal sized amounts of BD51 seed in each of the 64 control villages. This included

[14] BD51 is released as "Swarna-Sub1" in India and several other countries. Emerick et al. (2016) show that Swarna-Sub1 is similar to Swarna, besides Swarna-Sub1 being more flood tolerant. However, our sample is not a flood prone area. We introduced Swarna-Sub1 in the control villages because Swarna is not officially released in Bangladesh (and thus not available for sale) despite being the most popular variety in our sample at baseline. More precisely, 77 percent of farmers in the village census reported growing Swarna at baseline.

five farmers with the largest landholdings, five farmers selected by the SAO, and five farmers selected at random. These sets overlapped in some cases and therefore the number of farmers per control village is occasionally less than 15. Identifying these particular farmers in control villages was necessary in order to compare how the entry points we identified in the treatment groups cultivated BD56 relative to the longer duration counterfactual we distributed in the control villages. In the rest of the paper we refer to these counterfactual farmers as "entry-points" as well.

2.2 Timeline and Data Collection

Census with network information of each village

Figure 3 presents a complete timeline of the study. We began by performing a complete census of each village in March 2016. Our field team surveyed 21,926 households over the period of three months. Villages have between 14 and 184 households, 86 on average. We administered a short questionnaire to each member of the village, asking about their agricultural production (landholdings size, fertilizer use, production, and varieties sown), and their social networks (the name of the person they considered to be the best farmer, and the names of up to 10 farmers they turn to for advice on rice cultivation). We use these data to identify the largest farmers in each village, select the random entry-points, compute network statistics, and to forecast heterogeneous impacts of BD56 as a function of observable covariates.

Seed distribution

The distribution of 5kg minikits to the farmers selected to be entry-points in all 256 villages took place in early June 2016, in time for the Aman season. A total of 1,795 farmers were reached, achieving a response rate of 99 percent.[15] During these visits, the field team carefully explained the features of the seeds being distributed, and provided farmers with calendars to record the dates crops were sown, harvested, irrigated and applied with inputs. We also supplied farmers in the treatment villages with sticks and cards to place in their fields as a way of demonstrating to other farmers the variety they were planting. We briefly visited farmers 6 weeks later to make sure that we answered any remaining questions they had about the seeds, and to verify that the sticks were properly displayed in the fields.

Information diffusion survey

In April 2017 we visited all treatment villages, and randomly selected 10 additional farmers

[15] The total number of entry points is five per BD56 village, or 960, and up to 15 counterfactual farmers in the BD51 control villages. All 960 BD56 entry-points were reached, as well as 835 counterfactual farmers.

to conduct a small survey to determine if they had any knowledge of the BD56 seeds that were distributed to the entry points 9 months earlier. Specifically we asked whether they had heard about the new variety, which farmers they spoke to, and whether they could articulate some of the variety's key features. We use this information to assess how the diffusion of BD56 knowledge differs based on the village's treatment status.

Seed sale

In early June 2017, we visited each treatment village in our sample to sell 5kg and 2kg bags of BD56 seeds at subsidized prices. The field team called a select sample of farmers in each village to inform them about the date and time of the seed sale. The sample included the original minikit recipients, and the ten randomly selected farmers who were surveyed about their BD56 knowledge 2 months prior. The field team travelled to each village on the pre-determined date, and set-up their truck in the middle of the village (often at the local market) with a large sign indicating that the new BD56 seeds were available for purchase and that the main benefit of the seed is that its shorter duration allows for an additional crop to be grown. They recorded each sale that was made in a tablet. While we did not record the identity of the buyer, the survey provides a measure of BD56's diffusion within the village. Unfortunately, we ran out of seeds before traveling to all of the villages, and hence this information is only available for 168 of the 192 villages.

Agricultural surveys of entry-points.

In addition to the main sources of data described above, we conducted a series of agricultural surveys with the entry-points in all 256 villages. The goal of these surveys was to rigorously establish the impact of BD56 on agronomic practices, cropping intensity, and annual income. A baseline survey was administered at the time of the minikit distribution in June 2016. Important outcomes of interest included area cultivated, plot-level information on crops sown, inputs, and production volumes. We asked farmers to provide this information for 3 of their plots, which we selected randomly when farmers had more than 3.

This survey was followed by three other rounds in order to fully characterize annual production. In January-February 2017, we collected detailed information about the recently harvested Aman rice crop, and recorded whether farmers were planting a second post-Aman (known as the Rabi season) crop (midline 1). The survey asked specifically about seed variety choice, planting methods, and production at the plot level. In April 2017 we collected additional information about the Rabi crop, and the final Boro rice crop (midline 2). Finally, in August 2017, we asked farmers about their crop production levels during the Boro rice season, as well as their crop choice for the 2017 Aman rice season in order to gauge their propensity to re-plant the new rice variety (BD56) that we had offered them the previous year (endline). All surveys successfully reached the initial 1,795 farmers, except for two farmers missing in the April survey.

2.3 Baseline Characteristics from Census Data

Table A1 presents summary statistics from the household census and verifies randomization balance.[16] Our sample consists primarily of farmers cultivating long-duration rice varieties (only 1.17% of treatment farmers and 2.4% of control farmers planted short-duration varieties in the 2015 crop cycle). While approximately 35% of farmers only grow rice throughout the season, a non-negligible share grow a Rabi crop (including wheat, potato, pulses, onion, and garlic). Finally, farms in the sample are small: average area sown with Aman rice (the main crop) is approximately 1.33 acres.

We asked farmers to provide the names of up to 10 farmers they talked to about rice farming during last Aman season. We define two farmers as being connected or being peers if either name the other among the farmers they talk to. We use this information to create various centrality statistics including degree centrality (the number of connections a person has), eigenvector centrality and betweenness centrality (the number of times a node acts as a bridge along the shortest path between two other nodes). Figure 4 displays the distribution of these three measures for the entire sample of households interviewed during the census. Farmers have on average 4 connections with whom they talk to about rice cultivation, though we see a strong right tail with some farmers having up to 26 connections. The distributions of eigenvector and betweenness centrality display similar patterns with long right tails: while most farmers have a few connections some have a disproportionately high number. The long right tails in the distributions provides some initial evidence that the agricultural information networks do have some highly central farmers that in theory would serve as more effective entry points.

3 Results

We now provide our main results — starting with differences between entry points in their social network status. We then seek to understand how entry points demonstrated the attributes of BD56, by comparing them to the farmers growing long-duration rice in the control villages. Building on these results, we focus on how the demonstration plots affect awareness

[16]Table A2 further shows randomization balance for the sample of entry points, for the impact evaluation experiment.

and how these effects compare to the selection treatments. We argue that the efficacy of the demonstration plots is being driven by their ability to induce new communication links. In contrast, additional evidence suggests that the standard diffusion model in networks explains much of the effectiveness of our selection treatments. Finally, we combine our different data sources to test whether the treatments deliver information to farmers most likely to capture BD56's main benefit by increasing cropping intensity.

3.1 The network centrality of entry points

Recent literature has tested various mechanisms for identifying the most central nodes in social networks. These mechanisms include both direct elicitation of the entire social network (Beaman et al., 2015; Kim et al., 2015) or trying to infer centrality by asking a sample of individuals who is suitable for diffusing information (Banerjee et al., 2018b). Compared to eliciting the entire network, our two methods of selecting entry points are generally less demanding in terms of data; entering with large farmers requires only administrative data on farm sizes and SAO-based selection requires only a short interview with an agricultural extension agent. Yet, despite their ease of implementation, there is no guarantee that either of these two methods will deliver the entry points that are theoretically optimal for diffusion. We first check this using our social network survey.

The main measures of network centrality all differ noticeably between random entry points, those identified by SAO's, and the largest five farmers in the village. Table 1 displays average characteristics for the 960 entry points across the 192 BD56 treatment villages. The average farmer that was selected randomly has about 4.6 connections with other farmers in the village. The entry points selected by SAO's have an average of 8.14 connections and the five largest farmers in each village have an average degree of 9.04. Eigenvector centrality of entry points increases by 47 percent when selected by SAO's and 80 percent when chosen as the five largest farmers in the village. Banerjee et al. (2013) introduce diffusion centrality of farmer i as a measure of the expected number of times that farmers obtain a piece of information that was introduced with i.[17] Comparing to random entry points, average diffusion centrality of SAO-selected entry points is higher by 0.64 standard deviations and the diffusion centrality of large entry points increases by 0.87 standard deviations. Figure A2 in the appendix shows the cumulative distribution functions of the different network centrality measures across the selection treatments. Most importantly, the increases in centrality for

[17] This measure requires as parameters the number of periods for the diffusion process and the probability that an informed agent passes information to their social connections. We set the number of periods to 5 and the information passing probability to 0.5. We also normalize the measure by subtracting the village-specific mean and dividing by the village-specific standard deviation.

large and SAO farmers occur throughout the distributions.

These findings deliver an important verification that our experiment compares demonstration plots to meaningful methods of selecting entry points. Put differently, the SAO and large farmers in theory would be suitable for demonstrating technology in order to spread awareness. The high centrality of large and SAO farmers — relative to random entry points — is comparable to other mechanisms to selecting entry points that have been tested in the literature. For instance, Banerjee et al. (2018b) find that the median diffusion centrality of people identified by other villagers as suitable for spreading information is larger than that of other villagers by around 0.5 to 1 standard deviations. We find that relative to random targeting, larger farmer targeting would increase median diffusion centrality by around 0.51 standard deviations (Figure A2).

Besides network centrality, Table 1 also shows how several other observable characteristics vary across the different types of entry points. There are two notable observations. First, SAO-selected entry points tend to be larger farmers: farm size increases by about 5.9 bigha (3 bigha = 1 acre) for SAO entry points relative to random farmers. We show in Table A3 that controlling for farm size reduces substantially the gaps in network centralities between SAO-selected and random farmers. Put differently, extension officers could use knowledge of farm size in selecting entry points, and this explains part of the reason why SAO-selected entry points are more influential in networks. The ability of extension agents to select influential entry points contrasts with Beaman et al. (2015) who find that Malawian extension agents possess little information on the optimal entry points within social networks. The sharp correlation between farm size — an easily observable characteristic — and network centrality offers one possible explanation for the greater ability of extension agents in our sample. This phenomenon is visually evident in Figure 5 where we show the network structures for 6 randomly selected villages with either SAO or large-farmer selection. All three of the large-farmer villages in the top panel of the figure have at least one relatively larger farmer that is well connected in the network. Focusing on the SAO-selection villages in the bottom panel, both village 99 and 224 have medium or larger size farmers that are central in the network and were selected as entry points by the SAO.

Second, our door-to-door census asked each household to list the "best farmer" in the village. A random farmer is only named about 0.79 times while SAO entry points are named 5.27 times, and the five largest farmers in each village 6.4 times. These numbers further suggest that our selection treatments identify entry points that are both better networked and that other farmers consider to be knowledgeable about agriculture.[18]

[18] The number of nominations as the best farmer is correlated with network centrality. It explains 44 percent of the variation of degree centrality, 32 percent for eigenvector centrality, and 41 percent for betweenness

3.2 How do entry points demonstrate the technology?

In this section, we use the impact evaluation experiment to explore how BD56 affect farmers' cultivation practices and profits.

3.2.1 Take up

During the first midline survey we asked farmers whether they planted the seeds provided to them. We do not find any evidence of differential adoption rates (Table A4). Focusing exclusively on treatment villages, we further investigate whether take up varies across different types of entry points. We find that adoption rates remain fairly consistent across treatment arms, albeit slightly higher among large farmers assigned to the demonstration plots (Column 3). This last finding would suggest that the potential impacts of demonstration plots should be greater among large farmers because of the higher adoption rates, a result working against our main findings presented later in the paper.

3.2.2 Cultivation practices and profitability

The intervention affected cropping systems. Farmers planting BD56 harvested those fields 25 days earlier than farmers sowing BD51 (in late October rather than mid November) (Table A5 and Figure A3). Treatment farmers used this additional month between their two rice crops to increase the likelihood of planting a post-Aman (Rabi) crop. On average, BD56 plots were 27.8 percentage points more likely to be sown with the Rabi crop than BD51 plots (Table A5). Mustard, pulses, and potatoes were the most frequent short-season Rabi crop induced by the treatment.

Importantly for knowledge diffusion, this change in cropping systems is heterogeneous by type of entry point. The treatment effect for growing the Rabi crop is 17 and 11 percentage points higher for large and SAO farmers, respectively (Table A6). Because growing additional crops is such a visible activity, this offers a potential mechanism as to why knowledge diffuses faster with large and SAO selection — a possibility we investigate in Section 3.4.

While the BD56 treatment led to a sharp increase in cropping intensity, we still observe that 46 percent of the BD56 plots were left fallow in between the two rice crops.[19] In addition, BD56 naturally leads to lower yields given its shorter duration: the yield of BD56 plots was 31 percent lower than that of the longer duration BD51 plots. We also discovered that BD56

centrality.

[19]There are a number of explanations including that farmers were not prepared to grow an additional crop when the rice matured much earlier than anticipated (despite being told of the duration when receiving the seeds), an inability to access land with plows when it is surrounded by maturing rice, and lack of access to capital for planting an additional crop.

fetched a slightly lower market price, which farmers attributed to less familiarity by millers. In combination, profits during the Aman season were lower by 4,576 taka for BD56 plots, or around 44 percent (Table A7).

The average gain in profit from Rabi cultivation is 1,436 taka (60 percent). Assuming all of these benefits come from the extensive margin of growing the crop, BD56 led to an increase in Rabi profits of 5,241 taka for farmers that complied by growing the additional crop afforded by the treatment. This suggests that the technology was profitable among the subset of farmers who fully complied by planting the Rabi crop, but was not profitable on average since not all farmers capitalized on this main benefit of the technology.

3.3 How does knowledge diffuse across treatments?

Do demonstration plots increase awareness about new technology? If so, how do these effects compare to those generated by improved selection of entry points? The follow up information survey in the 192 BD56 villages allows us to answer these questions. The farmer-level specification compares awareness across the six different arms of the diffusion experiment. The corresponding regression is

$$\begin{aligned} aware_{ivs} = \beta_0 + \beta_1 RandomDemo_{vs} + \beta_2 SAONoDemo_{vs} + \beta_3 SAODemo_{vs} \\ + \beta_4 LargeNoDemo_{vs} + \beta_5 LargeDemo_{vs} + \alpha_s + \varepsilon_{ivs}, \end{aligned} \tag{1}$$

where the dependent variable is an indicator for whether farmer i in village v and upazila s is aware of BD56, $SAONoDemo_{vs}$ is an indicator for villages with SAO selection and no demonstration plots, and the remainder of the variables are defined analogously. As in all of the analysis, we include strata (upazila) fixed effects and cluster standard errors at the village level.

The results in Table 2 deliver three insights. First, the demonstration plots increase knowledge when cultivated by randomly selected farmers. Specifically, the rate of awareness increases by 7.2 percentage points when random farmers grow BD56 side by side with a chosen comparison variety. Sixty percent of farmers were knowledgeable of BD56 in control villages, meaning that the treatment effect of demonstration plots amounts to a 12 percent effect. The large rate of awareness in $RandomNoDemo$ villages (the omitted category) shows that information diffuses, even under the benchmark where entry points are random and demonstration plots are absent.

Second, the demonstration plots have no effects when cultivated by the better-connected large and SAO farmers. The estimates of β_2 and β_3 are nearly identical, meaning that the demonstration plots didn't spread knowledge with SAO selection. Similarly, the estimates of β_4 and β_5 indicate that adding demonstration plots failed to increase awareness with large-farmer entry points.

Third, the effect of demonstration plots with random farmers is roughly the same magnitude as the effects of entering with large and SAO farmers. As we would expect based on their network connections, entering with large and SAO-selected farmers increases awareness. Amongst non-demo villages, SAO selection increases awareness by 6.7 percentage points (11.2 percent) and entering with the largest farmers increases awareness by 7.4 percentage points (12.3 percent). These effects are quite similar and statistically indistinguishable from the effect of demonstration plots with random entry points. Moreover, the demonstration plots eliminate the effects of targeting more central farmers. Specifically, the estimates of β_1, β_3, and β_5 are indistinguishable.

Columns 2 and 3 of Table 2 show effects on the number of conversations farmers reported having about BD56. While somewhat noisier, these data are also consistent with the demonstration plots creating just as much conversation as the improved selection of entry points. The demonstration plots led to 0.12 more conversations per farmer about BD56 when entry points were selected randomly (col. 2), almost all of which with entry points (col. 3). This 14 percent effect is similar to the effect on knowledge reported in column 1. Focusing on the mean outcomes, the average respondent in the control villages reported 0.84 conversations about BD56, 0.72 of which were with the entry points, and by difference 0.12 were with any of ten other randomly selected farmers that each farmer was asked about. Importantly, the reported conversations should not be interpreted as the total number of conversations about BD56, but rather the number of conversations with the 15 farmers asked about in our survey.[20]

The data on conversations help to rule out an alternative explanation where the side-by-side comparison — and two markers in the field — was more effective at simply broadcasting information on the existence of BD56. Instead, the demonstration plots caused farmers to engage in social learning, rather than just learn about the new technology from the sign in the field.

3.4 What mechanism explains the effects?

We highlight one mechanism which makes the demonstration plots work. Once allowing for endogenous communication across network links, the demonstration plots make people pay attention to farmers outside of their immediate network. This mechanism is consistent with the entry-point selection effects being eliminated by demonstration plots. The 64 villages

[20]We collected information on conversations between the respondent and the five entry points as well as 10 randomly selected other farmers for each village.

with random entry-point selection allow us to further consider this mechanism. Within these villages, the number of entry points in a farmer's network is as good as randomly assigned when conditioning on the total number of connections of that farmer.[21] As a result, the average effect of being connected to an additional entry point can be estimated with

$$aware_{ivs} = \beta_0 + \beta_1 Entry\ Point\ Peers_{ivs} + \beta_2 Total\ Peers_{ivs} + \alpha_s + \varepsilon_{ivs}, \qquad (2)$$

where $Entry\ Point\ Peers_{ivs}$ is the variable measuring how many of the five entry points farmer i is connected to and $Total\ Peers_{ivs}$ is the network degree of farmer i. Under our framework the demonstration plots should make baseline relationships less important for awareness, i.e. cause a decrease in β_1.

The average peer effects are indeed consistent with social learning. Column 1 in Table 3 shows that being connected to an additional entry point increases awareness by 7.7 percentage points, i.e. about 12 percent. More interestingly, the second column shows that this social learning only exists in villages without demonstration plots. An additional connection to an entry point increases knowledge by 13.5 percentage points without demonstration plots, but when entry points set up head-to-head demonstrations this effect goes down significantly to only 1.9 percentage points. In addition, the demonstration plots only increase learning for farmers that were unconnected to entry points at baseline. Demonstration plots increase awareness by 11.3 percentage points for farmers having no baseline connections to entry points. The effect disappears with just one connection to an entry point as the coefficient on the interaction term is nearly identical to the coefficient on the demonstration villages indicator. The similarity of the effects points to how the demonstration plots substitute for social connections to entry points: the increased awareness generated by demonstration plots is nearly the same as the peer effects in non-demo villages. Lastly, measuring connections with a binary variable for being connected to at least one entry point does not change the results (columns 3 and 4).

Table A8 shows the analogous results where the dependent variable is instead the number of reported conversations between respondents and entry points. The results are consistent with those on knowledge. The demonstration plots induced conversations between entry points and farmers that were outside of their baseline information networks.

Finally, the demonstration plots were only effective for the least networked farmers. We investigate this by limiting to the random-entry-point villages and estimating

$$aware_{ivs} = \beta_0 + \beta_1 Demo_{vs} + \beta_2 Eigenvector_{ivs} + \beta_3 Eigenvector_{ivs} * Demo_{vs} + \alpha_s + \varepsilon_{ivs}, \quad (3)$$

[21] Miguel and Kremer (2004) use a similar strategy when estimating spillover effects from deworming in Kenya.

where $Eigenvector_{ivs}$ is the baseline eigenvector centrality of farmer i. The estimated coefficient on the interaction term β_3 measures whether the demonstration plots had a differential effect for farmers that were more central at baseline. Intuitively, the more connected farmers have a number of ways (both direct and indirect) to find out about BD56. In contrast, the least central farmers are the most likely to benefit from making new connections to entry points.

Table 4 shows that the effect of demonstrations varies significantly according to the farmer's eigenvector centrality. The coefficient on the interaction term (β_3) is negative and precisely estimated. The magnitude of the coefficient is large. Going from the 10th to the 90th percentile of the eigenvector centrality distribution causes the effect of demonstrations to go from 0.165 to -0.022. Put another way, demonstration plots become ineffective for the most central farmers in the network.

This evidence favors the explanation that demonstration plots substitute for social learning in networks. This substitution also offers an explanation for why the identities of entry points become irrelevant when they cultivate demonstration plots. Demonstration plots effectively cause farmers to pay attention, thus "turning off" peer effects and eliminating the need to rely on information being transmitted from entry points to other farmers. Rather, the demonstration plots induce information to flow to people that would have otherwise been excluded from peer-to-peer social learning.

In contrast to inducing conversations in endogenous social networks, the baseline network structure explains part of the effects of SAO and large-farmer selection. Any network-based model of diffusion with an exogenous network would favor targeting more central entry points. As a result, the effects of SAO and large-farmer selection would be predicted to decrease when conditioning on the average centrality of entry points. The data show exactly this. Table 5 shows that conditioning on the average degree centrality of entry points (moving from column 1 to 2) causes the effects of large and SAO selection to decrease by 43 and 31 percent, respectively.[22] In addition, the average degree of entry points is strongly correlated with knowledge diffusion. We of course can not provide a causal interpretation of this parameter. The exercise instead offers evidence that the ability of large and SAO entry points to increase knowledge is due to their more central network positions.

Beyond network centrality, the act of growing the additional crop — something more likely done by large and SAO farmers — appears to have captured attention and resulted in increased knowledge of BD56. Column 3 shows that the effects of large and SAO selection also decrease when conditioning on the number of entry points that planted a Rabi crop on

[22]We limit the data to the non-demo villages for this analysis since the selection treatments are only effective in these villages.

their BD56 plot. The effects of large farmer and SAO selection are smaller by 63 percent and 51 percent, respectively, when conditioning on both degree centrality and the choice of growing a Rabi crop.

The villages with random and SAO-based selection also include large farmers as entry points. At least one of the five largest farmers was selected as an entry point in 19 of the 64 random villages and 38 of the 64 SAO selection villages. Table 6 shows analysis where we compare villages where at least one large farmer was targeted with those having no large-farmer entry points.[23] We find similar results when exploiting this variation. In column 1, having at least one large-farmer entry point increases awareness by 7.9 percentage points in non-demonstration villages. Similar to the previous analysis, conditioning on the average network degree of entry points and the number growing the Rabi crop absorbs much of this effect (columns 2 and 3). Introducing demonstration plots makes these effects disappear. Columns 4-6 show that hitting a large farmer has no effect in demonstration villages and that the correlation between the degree centrality of entry points and awareness is much weaker with demonstration plots.[24] Consistent with the other findings, the demonstration plots decrease the relevance of the baseline social network for knowledge diffusion.

In sum, two attributes explain much of the reason why the selection treatments were effective. First, large and SAO farmers are more central in the network and therefore share information with more farmers. Second, these farmers do a better job of showcasing the benefits of new technology. Nonetheless, despite these two mechanisms, demonstration plots with random farmers are equally effective at spreading awareness about new technology.

3.5 Effects on seed purchases

Awareness about new technology is our main outcome variable. In addition to information diffusion, we also obtained data on purchases of BD56 seeds. A local NGO visited each village prior to the 2017 rainy season — a year after BD56 had been introduced in the village. The NGO set up a small stand and made BD56 seeds available to any farmer wishing to purchase. Importantly, the seeds were subsidized at a rate of 60 percent, and farmers visiting the shop were told that BD56 shortens the season, gives lower yield relative to longer duration varieties, but allows for an additional crop to be grown during the year.[25] The NGO representative explained to farmers that a third crop is needed to make BD56

[23]We only worked with SAO's to select entry points in the 64 SAO villages and the 64 BD51 villages. We therefore can't repeat this analysis using SAO-identified farmers rather than large farmers.

[24]Table A9 gives the statistical test showing that across all villages, the impact of hitting at least one large farmer is significantly smaller in demonstration villages.

[25]That year the government rate for BD56 seed ranged between 34 and 40 tk/kg. The NGO sold the seeds for 15tk/kg.

profitable on an annual basis.

Table 7 shows regression results akin to Equation (1), but at the village level where the dependent variable is either the number of farmers purchasing or the adoption rate (number of buyers divided by village size). The point estimates are much noisier, but the coefficients are sizable and the directions line up with what we observe on knowledge diffusion. About 1.7 farmers purchased seeds per village in control villages, and this increased by around 0.67 farmers (40 percent) when adding demonstration plots. The number of farmers purchasing seeds also increases in large and SAO villages. Turning to column 2, the degree centrality of entry points and the number growing the third crop are positively correlated with the number of purchasing farmers and absorb some of the selection effects. Columns 3 and 4 show that the pattern remains when considering the share of farmers purchasing, rather than the absolute number. Overall, the coefficients in Table 7 show a similar pattern to what we observe on awareness creation, despite being less precisely estimated.

3.6 Who becomes informed from the different treatments?

Combining our two experiments (diffusion and impact evaluation experiments) give a unique opportunity to test whether alternative mechanisms for encouraging information diffusion deliver information to the people expected to benefit the most from new technology. Capturing the short-duration benefits of BD56 requires growing an additional short-season crop — an action that was not universally taken by farmers in our BD56 treatment group (Table A5). We next ask two questions (1) do observable characteristics explain variability across farmers in the treatment effect of BD56 on the number of crops grown across the year and (2) do the demonstration plots or the entry-points selection strategies differentially inform the farmers who, based on these same observable characteristics, are the most likely to increase their number of crops grown when adopting BD56?

We start by using the impact experiment to estimate a function measuring the treatment-effect heterogeneity of BD56, which we refer to as the heterogeneity index. The first step is to implement the method suggested in Chernozhukov et al. (2018) to generate a linear prediction of the ATE conditional on the observed covariates from our door-to-door census, denoted as z_i. Our sample of treatment and control farmers is first divided into two samples: one "training" sample where we seek to estimate the heterogeneity index, denoted as $s_0(z_i)$, and another "validation" sample where we seek to measure whether this estimate $\hat{s}_0(z_i)$ is a significant determinant of the heterogeneous treatment effect of BD56 on the number of crops grown. First for the training sample, we estimate separate LASSO regressions for the treatment and control groups to pick which of the covariates in z predict the number of crops grown y. Using these covariates, we generate estimates of the conditional expectations of the number of crops grown as $E(y_i|D_i = 1, z_i)$ and $E(y_i|D_i = 0, z_i)$.[26] The difference between these two conditional expectations serves as the heterogeneity index, $\hat{s_0}(z_i)$.

Turning to the validation sample, we want to verify that the treatment effect varies according to this measure $\hat{s_0}(z_i)$. We do this in two ways. First, we add an interaction between the treatment indicator and $\hat{s_0}(z_i) - \bar{s}$ in a regression where the dependent variable is the number of crops grown.[27] Second, we estimate separate treatment effects for the four quartiles of the distribution of $\hat{s_0}(z_i)$. Finally, this process is iterated 100 times, delivering 100 separate sample divisions and 100 estimates of the heterogeneity index.

The observed covariates predict treatment-effect heterogeneity in the validation sample. Figure 6 shows the 100 estimates of the ATE and the linear heterogeneity term. The heterogeneous effect is almost always larger than zero, suggesting that the heterogeneity index $\hat{s_0}(z_i)$ does proxy for the true heterogenous effect of BD56 on the number of crops grown. In other words, farmers with larger values of $\hat{s_0}(z_i)$ appear more likely to increase cropping intensity if adopting short-duration rice. Figure 7 shows the separate treatment effects by quartile of the heterogeneity index. Treatment effects increase with the heterogeneity index and are largest in the top two quartiles of the distribution of $\hat{s_0}(z_i)$.

We then return to the diffusion experiment in 192 villages to test whether the treatment effects differ based on values of the heterogeneity index. We possess 100 estimates of $\hat{s_0}(z_i)$ for each of the 1,920 farmers for which we elicited knowledge of BD56. We take the median of $\hat{s_0}(z_i)$ across these 100 sample divisions and estimate whether the treatment effects of demonstration plots or entry-points selection strategies depend on this predicted heterogeneity index.

Table 8 shows the regression results. Columns 1 and 2 interact the treatment variables directly with $\hat{s_0}(z_i)$, while columns 3 and 4 use an indicator for observations with above-median values of this heterogeneity index. The estimates are noisy, but the heterogeneity index is positively associated with learning and having conversations, indicating that information is more likely to flow to farmers with higher returns when entry points are selected randomly and there are no demonstrations. For instance, farmers with an above-median heterogeneity index are 10.1 percentage points more likely to learn about BD56 in random and no demo villages (column 3). From column 4, these same farmers are expected to have .26 more conversations (about 31 percent). However, the point estimates on the interaction

[26] We use the OLS regressions with the covariates selected by the two LASSO procedures. The selected covariates can be different in the treatment and control groups.

[27] The coefficient on the treatment indicator in this regression measures the average treatment effect, while the coefficient on the interaction between treatment and $\hat{s_0}(z_i) - \bar{s}$ measures whether the heterogeneity index predicts actual treatment-effect heterogeneity.

terms between the heterogeneity index and the treatment indicators are generally negative and the coefficients on the five treatment indicators are positive and of similar magnitudes. These findings indicate that while our treatments increased knowledge by either exploiting existing network structure or triggering conversations, the gains in knowledge were concentrated amongst farmers that would be less likely to capitalize on the main benefit of BD56 if adopting.

As examples in column 4, demonstration plots increased conversations about BD56 for farmers with *below-median* predicted effects on cropping intensity by around 0.29, but had no effects for farmers who appeared more likely to increase cropping intensity if adopting BD56 (0.29 + -0.32). Similarly, seeding with SAO-selected farmers and with the largest farmers increases conversations for farmers with below-median values of the heterogeneity index, but had no effect for farmers that are above the median.

This finding sheds light on who is induced to have conversations and learn when a policymaker intervenes with either an alternative seeding strategy or demonstration plots. Put simply, the people impacted by these treatments do not appear to be those that would be the most likely to enjoy the technology's main benefit if it was randomly introduced to them. One reasonable interpretation is that conversations take place and information is obtained endogenously. Therefore, farmers with the highest returns from obtaining information are more likely to engage in social learning regardless of the dissemination strategy. Intervening to trigger the spread of information only affects those with lower returns, i.e. those who are less likely to endogenously seek information regardless of the dissemination strategy.

4 A model that rationalizes the experimental results

This section presents a basic diffusion model that allows for endogenous interaction between unlinked agents and can rationalize our experimental findings. Our goal is not to capture all of the strategic elements of network formation.[28] Instead, we opt for the simplest formulation that captures the tradeoffs introduced by our treatments. First, the policymaker can introduce information to central entry points therefore taking advantage of the existing network structure. Or, the policymaker can seek to induce more communication in a world where communication networks are endogenous. We show how the two approaches act as substitutes in a way that is consistent with our empirical findings.

[28] There are several papers looking at different aspects of how information links are formed. These range from differences between actively sharing information and passively listening (Calvó-Armengol, Martí, and Prat, 2015), the types of initial network structures that allow for efficient information aggregation as societies grow large (Acemoglu, Bimpikis, and Ozdaglar, 2014), and the stigma from seeking information when it might signal low ability or understanding (Chandrasekhar, Golub, and Yang, 2016; Banerjee et al., 2018a).

4.1 Model Environment

We consider a village with N farmers indexed by $i \in \{1, 2, ..., N\}$. The village social network is described by the $N \times N$ adjacency matrix G, where $g_{ij} = 1$ indicates that farmers i and j have an information-sharing link. In terms of our data, this feature of the model corresponds to the baseline social network module.

The policymaker first seeds technology with five entry points, indexed by $j = 1, ...5$. Each entry point is now "informed" and can then spread the information to others, as in the standard information cascade. We assume, for now, that the network is fixed and a farmer can only become informed if they receive information that emanated from an entry point. The parameter q represents the exogenous probability that any informed farmer passes information to a connected peer. The probability that a farmer becomes informed, denoted as h_i, is a function of the information-passing probability q and the length of the possible paths between i and each entry point j. Intuitively, farmers having the greatest number of shortest paths to entry points become the most likely to be informed. In contrast, a farmer that is completely isolated from entry points is not informed.

We build on this standard framework by adding the ability to form new links. In addition to passively waiting for information to arrive from entry points, a farmer can increase the probability of becoming informed by forming a link with an entry point. Suppose the cost of forming a new link is c and the probability that an entry point will pass information is p. A new link between farmer i and entry point j is denoted as $l_{ij} = 1$ if i chooses to form the link and 0 otherwise. Overall, the probability of gaining information from one of the two channels, denoted as μ_i, is

$$\mu_i = 1 - (1 - h_i) * \prod_j (1 - p)^{l_{ij}} . \tag{4}$$

Finally, we write v as the utility of being informed and normalize the utility of not being informed to 0.

4.2 The link-formation decision

The simple problem of the farmer is to choose whether to link with each of the entry points. More formally, the farmer's optimization problem is written as

$$\max_{l_{ij}} v \left(1 - (1 - h_i) * \prod_j (1 - p)^{l_{ij}} \right) - \sum_j c * l_{ij}. \tag{5}$$

The problem can be simplified to choosing the number of new contacts with entry points, denoted as m, since each entry point adds the same probability of learning p. The exact decision rule is that the farmer will link with m entry points if and only if:

$$v(1 - h_i)p(1 - p)^{m-1} > c. \tag{6}$$

The farmer seeks to connect directly to entry points for information if the costs of doing so are low (c decreases), or if they are in a poor position to obtain information via diffusion in the network (h_i is low). Increasing the probability that an entry point shares useful information (p) increases the likelihood of connecting with one entry point ($m = 1$). If p is sufficiently large, then increasing p causes the marginal benefit of linking with further entry points to decrease because the farmer is likely to obtain the information from the first entry point newly added to her network.

4.3 Consistency between the model and experimental findings

This small modification to a standard diffusion framework predicts treatment effects that are in line with our RCT results. Put differently, the two mechanisms for being informed — receiving information via the existing social network or communicating with an entry point outside the network — can explain the main impacts as well as the heterogeneity across the sample. The discussion that follows links this simple theory to our results.

Effectiveness of entering with Large/SAO farmers: By being more central in networks, large and SAO-selected farmers facilitate diffusion. Holding networks fixed, the probability h_i in equation (6) increases, causing the probability of receiving information to increase. This prediction is the obvious one that supports network-based approaches to identifying entry points when network structure remains fixed. The main effects of large farmer and SAO-based selection on knowledge, and the sensitivity of these effects to conditioning on centrality of entry points (Tables 2, 5, and 6) are all consistent with this standard mechanism.

Effectiveness of demonstration plots: We argue that the demonstration plots convey active experimentation and signal a farmer that is paying attention to how the new technology performs against its' relative alternative. This is nontrivial as Hanna, Mullainathan, and Schwartzstein (2014) show that inattention bias can hinder what farmers learn from experimentation. As a result, fellow villagers perceive that farmer to be more likely to pass useful information, i.e. the parameter p increases with the introduction of demonstration plots. The marginal benefit of forming a link with a single entry point then increases. Correspondingly, the new information link increases the likelihood of becoming informed. The main results on knowledge and reported conversations in Table 2 appear consistent with this reasoning.

Interactions between network-based selection and demonstration plots: The endogenous network mechanism posits substitutability between seeding with more central farmers and demonstration plots. Returning to equation (6), seeding with more central farmers increases the probability of learning via existing network links and therefore reduces the marginal benefit of endogenously seeking information from entry points. The findings show this exactly: demonstration plots have no effect when cultivated by more central farmers and the effect of entering with more central farmers is eliminated when the policymaker introduces demonstration plots.

Network effects: Farmers directly connected to entry points gain less from demonstration plots because the existing network connection increases the likelihood of learning through the information-diffusion mechanism. We found that baseline connections with entry points do increase knowledge, however, demonstration plots eliminate this advantage by giving a channel for unconnected farmers to learn. This is compatible with equation (6) where the benefit of making new links with entry points declines with h_i.

In sum, our experimental findings, along with the simple model, emphasize the two mechanisms that serve to boost learning in networks: optimizing the selection of entry points or inducing communication to facilitate learning. The latter mechanism has received less attention in the literature. However, our results suggest that it gives the policymaker an important alternative for making information diffuse faster in networks.

5 Concluding Remarks

We have shown experimental evidence on a new mechanism for spreading awareness about technology. Demonstration plots — where partnering farmers cultivate a new technology side by side with an existing one — increase awareness relative to a control where new technology is demonstrated on its own. We found that this relatively straightforward method of agricultural extension made an additional seven percent of farmers aware about a new seed variety. Demonstration plots raised awareness only for farmers that lacked social connections with adopters. In addition, the demonstration plots were the most effective for the farmers that were the most isolated — in terms of their eigenvector centralities — in the baseline information network. All of these results, when taken together, are consistent with a model where the effectiveness of demonstration plots is explained by their ability to facilitate learning by triggering communication in information-sharing networks.

The experiment benchmarked these demonstration plots against scalable and policy-relevant alternatives where entry points were selected strategically to increase their centralities in the social network. These were entering with the largest farmers and those hand picked by government extension agents. Indeed, these methods do increase knowledge compared to random selection. But the gains in awareness are about the same as the gains from the demonstration plots. And in contrast to endogenous communication between people without links, the selection treatments seem to be effective because of how they exploit the structure of the baseline social network and because the more central entry points did a better job of demonstrating a key benefit of new technology.

We also showed evidence on who becomes informed. The seed variety we introduced has heterogeneous benefits. Specifically, only some farmers took advantage of the early maturation by increasing cropping intensity. Applying machine-learning methods to identify the characteristics associated with taking this action, we found farmers who are the most likely to grow an additional crop when adopting BD56 are not those learning from demonstration plots. We found the same result for improved seeding strategies. This finding highlights a tradeoff: intervening to increase knowledge transmission may be ineffective for the highest return individuals who learn even in the absence of additional intervention by the policymaker.

Overall, our analysis highlights the potential for alternate mechanisms of agricultural extension, outside of those that rely on information cascades through social networks. Despite the ample evidence that networks are important for knowledge transmission, there have been few studies that compare information cascades with alternative methods of spreading knowledge.[29] Focusing on policy, it is important to consider such alternatives because policymakers may face difficulty in identifying the entry points that are theoretically positioned for the best spread of information. Either it could be prohibitively expensive to do a full network survey, or there may not be observable characteristics (such as farm size) that correlate strongly with less observable measures of network centrality. Our results show that in these contexts, improving learning can be achieved by taking small steps to capture people's attention, and get them to communicate and engage in learning from new people.

[29] Outside of agricultural technology, Banerjee et al. (2018a) is the only paper we are aware of that compares seeding information about rules for India's 2016 demonetization with broadcasting that information more widely.

References

Acemoglu, Daron, Kostas Bimpikis, and Asuman Ozdaglar. 2014. "Dynamics of information exchange in endogenous social networks." *Theoretical Economics* 9 (1):41–97.

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi. 2018. "Just a Few Seeds More: Value of Network Information for Diffusion." *Unpublished* .

Anderson, Jock R and Gershon Feder. 2007. "Agricultural extension." *Handbook of Agricultural Economics* 3:2343–2378.

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

Banerjee, Abhijit, Emily Breza, Arun Chandrasekhar, and Benjamin Golub. 2018a. "When Less is More: Experimental Evidence on Information Delivery During India's Demonitization." *Unpublished* .

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

———. 2018b. "Using Gossips to Spread Information: Theory and Evidence from two Randomized Controlled Trials." *Unpublished* .

Beaman, Lori, Ariel BenYishay, Mushfiq Mobarak, and Jeremy Magruder. 2015. "Can Network Theory based Targeting Increase Technology Adoption?" *Unpublished* .

Beaman, Lori and Andrew Dillon. 2017. "Diffusion of Agricultural Information within Social Networks: Evidence on Gender Inequalities from Mali." *Unpublished* .

BenYishay, Ariel and A Mushfiq Mobarak. 2015. "Social Learning and Incentives for Experimentation and Communication." *Unpublished* .

Birkhaeuser, Dean, Robert E Evenson, and Gershon Feder. 1991. "The economic impact of agricultural extension: A review." *Economic Development and Cultural Change* :607–650.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman. 2014. "Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions." *Econometrica* 82 (4):1273–1301.

Cai, J, A de Janvry, and E Sadoulet. 2015. "Social Networks and the Decision to Insure." *American Economic Journal: Applied Economics* 7 (2):81–108.

Calvó-Armengol, Antoni, Joan Martí, and Andrea Prat. 2015. "Communication and influence." *Theoretical Economics* 10 (2):649–690.

Centola, Damon. 2010. "The spread of behavior in an online social network experiment." *Science* 329 (5996):1194–1197.

Chandrasekhar, Arun and Randall Lewis. 2016. "Econometrics of sampled networks." *Unpublished* .

Chandrasekhar, Arun G, Benjamin Golub, and He Yang. 2016. "Signaling, Stigma, and Silence in Social Learning." *Unpublished* .

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Ivan Fernandez-Val. 2018. "Generic machine learning inference on heterogenous treatment effects in randomized experiments." Tech. rep., National Bureau of Economic Research.

Cole, Shawn, Xavier Giné, and James Vickery. 2017. "How does risk management influence production decisions? Evidence from a field experiment." *The Review of Financial Studies* 30 (6):1935–1970.

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

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101:2350–2390.

Emerick, Kyle, Alain de Janvry, Elisabeth Sadoulet, and Manzoor H Dar. 2016. "Technological Innovations, Downside Risk, and the Modernization of Agriculture." *American Economic Review* 106 (6):1537–1561.

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

Golub, Benjamin and Matthew O Jackson. 2012. "How Homophily Affects the Speed of Learning and Best-Response Dynamics." *Quarterly Journal of Economics* 127 (3):1287–1338.

Guiteras, Raymond, James Levinsohn, and Ahmed Mushfiq Mobarak. 2015. "Encouraging sanitation investment in the developing world: a cluster-randomized trial." *Science* 348 (6237):903–906.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein. 2014. "Learning through noticing: Theory and evidence from a field experiment." *The Quarterly Journal of Economics* 129 (3):1311–1353.

Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Christopher Udry. 2014. "Agricultural decisions after relaxing credit and risk constraints." *The Quarterly Journal of Economics* 129 (2):597–652.

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

Kondylis, Florence, Valerie Mueller, and Jessica Zhu. 2017. "Seeing is believing? Evidence from an extension network experiment." *Journal of Development Economics* 125:1–20.

Miguel, Edward and Michael Kremer. 2004. "Worms: identifying impacts on education and health in the presence of treatment externalities." *Econometrica* 72 (1):159–217.

Mobius, Markus, Tuan Phan, and Adam Szeidl. 2015. "Treasure hunt: Social learning in the field." Tech. rep., National Bureau of Economic Research.

Munshi, Kaivan. 2004. "Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution." *Journal of Development Economics* 73 (1):185–213.

Rigol, Natalia, Reshmaan Hussam, and Benjamin Roth. 2017. "Targeting High Ability Entrepreneurs Using Community Information: Mechanism Design In The Field." *Unpublished* .

Sacerdote, Bruce. 2001. "Peer effects with random assignment: Results for Dartmouth roommates." *The Quarterly Journal of Economics* 116 (2):681–704.

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

Valente, Thomas W. 2012. "Network interventions." *Science* 337 (6090):49–53.

Tables

Table 1: Differences in baseline characteristics for different entry points

	Coefficients and SE:			
	(1) Constant	(2) SAO	(3) Large farmers	(4) p-value (2)-(3)
Network Variables:				
Degree	4.562***	3.582***	4.481***	0.473
	(0.355)	(1.042)	(0.853)	
Eigenvector centrality	0.089***	0.042***	0.071***	0.030
	(0.006)	(0.012)	(0.011)	
Diffusion centrality	-0.010	0.643***	0.872***	0.157
	(0.054)	(0.141)	(0.110)	
Betweenness centrality	164.186***	394.084***	315.640***	0.509
	(27.926)	(103.540)	(69.762)	
Household Characteristics:				
Area cultivated all seasons (bigah)	9.013***	5.865***	21.396***	0.000
	(0.658)	(1.368)	(2.689)	
Times named best farmer	0.790***	4.477***	5.589***	0.275
	(0.206)	(0.785)	(0.707)	
Log revenue per bigah	10.061***	-0.016	-0.014	0.970
	(0.057)	(0.077)	(0.075)	
Number livestock owned	3.950***	-0.008	1.968***	0.000
	(0.217)	(0.284)	(0.512)	
Number of overseas migrants	0.138***	-0.021	-0.026	0.881
	(0.031)	(0.039)	(0.037)	
Education	4.647***	1.247***	0.925*	0.536
	(0.304)	(0.464)	(0.488)	
Age	42.222***	0.712	3.594***	0.007
	(0.739)	(1.026)	(1.078)	
Tubewell owner	0.097***	0.094***	0.181***	0.108
	(0.022)	(0.036)	(0.051)	

The data are limited to the 960 selected entry points in the 192 BD56 villages. Each row is the result from a separate regression where the characteristic is regressed on a constant and indicators for SAO and large farmer villages. The omitted group is the villages where demonstrators were selected randomly (meaning the first column is the mean value for random entry points). The standard errors in each regression are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 2: Treatment effects on knowledge

	(1) Heard About	(2) Conversations	(3) Conversations w/ Entry Points
Random w/ demo	0.072*	0.116	0.106
	(0.041)	(0.086)	(0.088)
SAO no demo	0.067*	0.046	0.042
	(0.039)	(0.067)	(0.071)
SAO w/ demo	0.065	0.074	0.096
	(0.040)	(0.089)	(0.086)
Large no demo	0.074**	0.123*	0.114*
	(0.036)	(0.068)	(0.066)
Large w/ demo	0.049	0.108	0.113
	(0.044)	(0.075)	(0.079)
Strata fixed effects	Yes	Yes	Yes
Mean in Control	0.60	0.84	0.72
Number of Observations	1919	1920	1920
R squared	0.171	0.212	0.250

The data are for the 10 random farmers per village that were selected for the information survey. The dependent variable in column 1 is an indicator for having knowledge of BD56. The dependent variable in column 2 is the number of conversations the farmer had with 15 other farmers about BD56 (the five entry points and 10 randomly selected farmers). The dependent variable in column 3 is the number of conversations specifically with entry points. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels. Note, the rates of awareness are similar amongst the different selection arms in demonstration plot villages, i.e. the estimates on SAO w/ demo, Large w/demo, and Random w/demo are indistinguishable.

Table 3: Peer effects on knowledge, separate for villages with and without demonstration plots

	(1)	(2)	(3)	(4)
Peer connections w/	0.077**	0.135***		
entry points	(0.034)	(0.046)		
Peer connections w/		-0.116*		
entry points * Demonstration Village		(0.061)		
Connected to at			0.092*	0.156***
least 1 entry point			(0.053)	(0.059)
Connected to at				-0.130
least 1 entry point * Demonstration Village				(0.097)
Number of	-0.004	-0.005	-0.002	0.002
connections	(0.003)	(0.006)	(0.003)	(0.005)
Number of		0.000		-0.005
connections * Demonstration Village		(0.006)		(0.005)
Demonstration		0.113**		0.132***
Village		(0.045)		(0.045)
Strata fixed effects	Yes	Yes	Yes	Yes
Mean in Control	0.60	0.60	0.60	0.60
Number of Observations	635	635	635	635
R squared	0.186	0.199	0.185	0.197

The dependent variable in all regressions is an indicator for having heard of BD56 amongst the 10 randomly surveyed farmers per village. The data are limited to the 64 villages where entry points where chosen randomly and peer effects can therefore be causally identified. The variable *Peer connections w/ entry points* is the number of entry points (from 0 to 5) that the farmer is connected with while *Connected to at least 1 entry point* is an indicator variable for being connected to at least one of the entry points. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 4: Effects of demonstration plots as a function of baseline network centrality

	(1)	(2)
Demonstration Village	0.072* (0.041)	0.179*** (0.057)
Eigenvector Centrality * Demo		-1.067*** (0.381)
Eigenvector Centrality		0.935*** (0.325)
Strata fixed effects	Yes	Yes
Mean in Control	0.60	0.61
Number of Observations	639	517
R squared	0.183	0.202

The dependent variable in both regressions is an indicator for having heard of BD56. The data are limited to the 64 villages where entry points were chosen randomly. *Eigenvector Centrality* is the baseline network centrality of the respondent. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 5: Effects of entry-point treatments when conditioning on observable attributes of entry points

	(1)	(2)	(3)	(4)
SAO no demo	0.068*	0.047	0.056	0.033
	(0.038)	(0.039)	(0.039)	(0.040)
Large no demo	0.076**	0.043	0.064*	0.028
	(0.035)	(0.035)	(0.035)	(0.036)
Average degree of entry points		0.006***		0.006***
		(0.002)		(0.002)
Number entry points growing rabi crop			0.020*	0.022*
			(0.011)	(0.011)
Strata fixed effects	Yes	Yes	Yes	Yes
Mean in Control	0.60	0.60	0.60	0.60
Number of Observations	960	960	960	960
R squared	0.173	0.179	0.176	0.182

The data are for the 10 random farmers per village that were selected for the information survey and are limited to the 96 villages without demonstration plots. The dependent variable in all regressions is an indicator for being aware of BD56. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 6: Effects of having at least one large-farmer entry point on BD56 knowledge

	Non-Demo Villages			Demo Villages		
	(1)	(2)	(3)	(4)	(5)	(6)
At least 1 large entry point	0.079**	0.048	0.035	-0.017	-0.024	-0.025
	(0.032)	(0.034)	(0.035)	(0.037)	(0.040)	(0.038)
Average degree of entry points		0.005**	0.006***		0.003	0.003
		(0.002)	(0.002)		(0.003)	(0.003)
Number entry points growing rabi crop			0.022*			0.020
			(0.012)			(0.013)
Strata fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Mean in Control	0.62	0.62	0.62	0.68	0.68	0.68
Number of Observations	960	960	960	959	959	959
R squared	0.175	0.179	0.182	0.175	0.176	0.180

The data are for the 10 random farmers per village that were selected for the information survey. The dependent variable is an indicator for being aware of BD56. Column 1-3 are for the villages without demonstration plots and columns 4-6 are for the demonstration villages. *At least 1 large entry point* is an indicator for villages where one of the five largest farmers was selected as an entry point. The mean in the control group is defined as the mean awareness rate in villages where none of the entry points were large farmers. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table 7: Treatment effects on seed purchasing behavior

	Number of farmers		Share of village	
	(1)	(2)	(3)	(4)
Random w/ demo	0.673	0.613	0.00622	0.00493
	(0.813)	(0.807)	(0.00916)	(0.00917)
SAO no demo	0.697	0.392	0.0116	0.00809
	(0.694)	(0.690)	(0.0107)	(0.0104)
SAO w/ demo	0.116	-0.154	0.00411	0.00135
	(0.615)	(0.640)	(0.00744)	(0.00774)
Large no demo	0.272	-0.171	0.00817	0.00325
	(0.535)	(0.621)	(0.00782)	(0.00868)
Large w/ demo	0.866	0.515	0.0195^{**}	0.0147
	(0.641)	(0.654)	(0.00946)	(0.00914)
Average degree of entry points		0.0642^{*}		0.000567
		(0.0382)		(0.000507)
Number entry points growing rabi crop		0.166		0.00291
		(0.156)		(0.00214)
Strata fixed effects	Yes	Yes	Yes	Yes
Mean in Control	1.679	1.679	0.019	0.019
Number of Observations	168	168	168	168
R squared	0.085	0.106	0.091	0.106

The data are from seed sales that were carried out for each village prior to the 2017 rainy season. We are missing data for 24 of the 192 villages because the seed supply ran out before those villages could be completed. The dependent variables are the number of farmers purchasing BD56 seeds (columns 1-2) and the share of farmers purchasing (columns 3-4). Robust standard errors are in parentheses. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

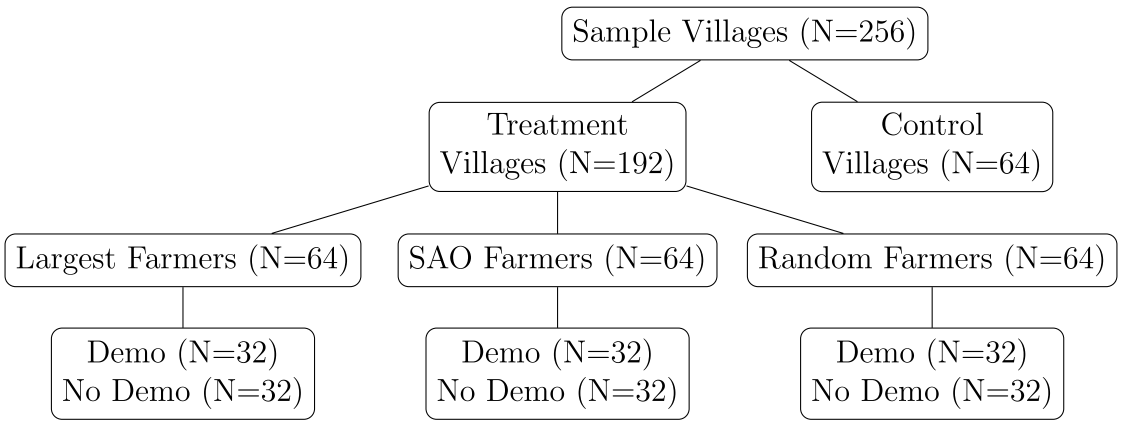
Table 8: Heterogeneous effects on knowledge and conversations by predicted impact of BD56 on number of crops grown

	Linear Heterogeneity		Effect Above Median	
	(1) Heard About	(2) Conversation	(3) Heard About	(4) Conversations
Random w/ demo	0.082	0.247*	0.121**	0.294**
	(0.069)	(0.131)	(0.060)	(0.128)
SAO no demo	0.073	0.096	0.089	0.125
	(0.072)	(0.115)	(0.060)	(0.094)
SAO w/ demo	0.096	0.184*	0.111*	0.253**
	(0.064)	(0.110)	(0.057)	(0.099)
Large no demo	0.148**	0.247**	0.148***	0.284***
	(0.061)	(0.101)	(0.053)	(0.101)
Large w/ demo	0.094	0.190	0.123*	0.265***
	(0.083)	(0.116)	(0.068)	(0.096)
Heterogeneity	0.034	0.236	0.101*	0.261**
	(0.123)	(0.195)	(0.057)	(0.103)
SAO no demo *	0.001	-0.145	-0.014	-0.111
Heterogeneity	(0.164)	(0.282)	(0.072)	(0.115)
SAO w/ demo *	-0.096	-0.369	-0.072	-0.337**
Heterogeneity	(0.143)	(0.227)	(0.077)	(0.148)
Large no demo *	-0.230	-0.386*	-0.123*	-0.276**
Heterogeneity	(0.152)	(0.227)	(0.072)	(0.135)
Large w/ demo *	-0.136	-0.252	-0.124	-0.274**
Heterogeneity	(0.192)	(0.240)	(0.081)	(0.119)
Random w/ demo *	-0.020	-0.430*	-0.078	-0.318**
Heterogeneity	(0.161)	(0.241)	(0.077)	(0.141)
Strata fixed effects	Yes	Yes	Yes	Yes
Mean in Benchmark	0.60	0.85	0.60	0.85
Number of Observations	1910	1911	1910	1911
R squared	0.174	0.216	0.175	0.219

These regressions test whether the different treatments increase knowledge and spark conversations more (or less) for farmers that are predicted to have the largest impact of of BD56 on the number of crops grown. Columns 1 and 2 show linear heterogeneity where the treatment indicators are interacted with $\hat{s_0}(z_i) = E(y_i|D_i = 1, z_i) - E(y_i|D_i = 0, z_i)$ and columns 3 and 4 partition the sample into farmers that are above and below the median in the distribution of $\hat{s_0}(z_i)$. For each farmer we calculate $\hat{s_0}(z_i)$ as the median value across the 100 sample divisions in Figures 6 and 7. The dependent variable in columns 1 and 3 is an indicator for having knowledge of BD56. The dependent variable in columns 2 and 4 is the number of conversations the farmer had with 15 other farmers about BD56. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Figures

Figure 1: Experimental Design



Notes: Figure shows a schematic representation of the experimental design. The 192 BD56 treatment villages were divided into three groups for entry-point selection: random selection, relying on the five largest farmers, and selecting those indicated by the ag. extension officer (SAO). Demonstration plots were set up on half of the 64 villages within each of these arms.

Figure 2: A visualization of the demonstration plot in comparison to the control group

A: Demonstration Plot | B: Control Plot

Notes: Panel A on the left shows an example demonstration plot. The plot on the left side is the BD56 plot while the plot on the right is the popular longer duration variety Swarna. Panel B on the right shows an example from the comparison villages where farmers were only given one marker to denote the BD56 plot.

Figure 3: A timeline of the experiment and data collection

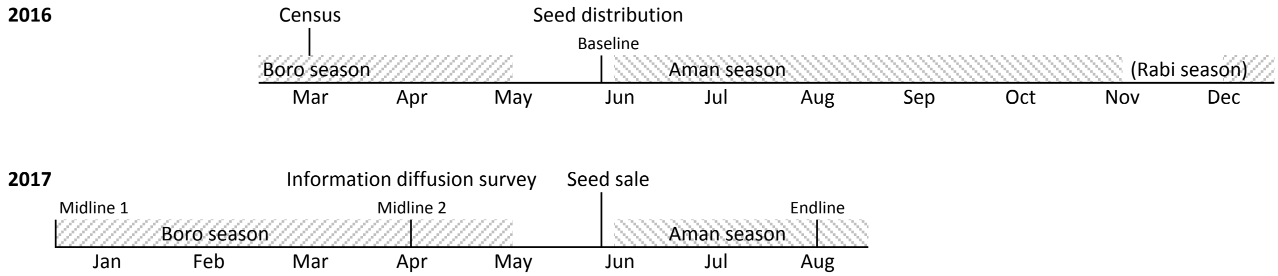
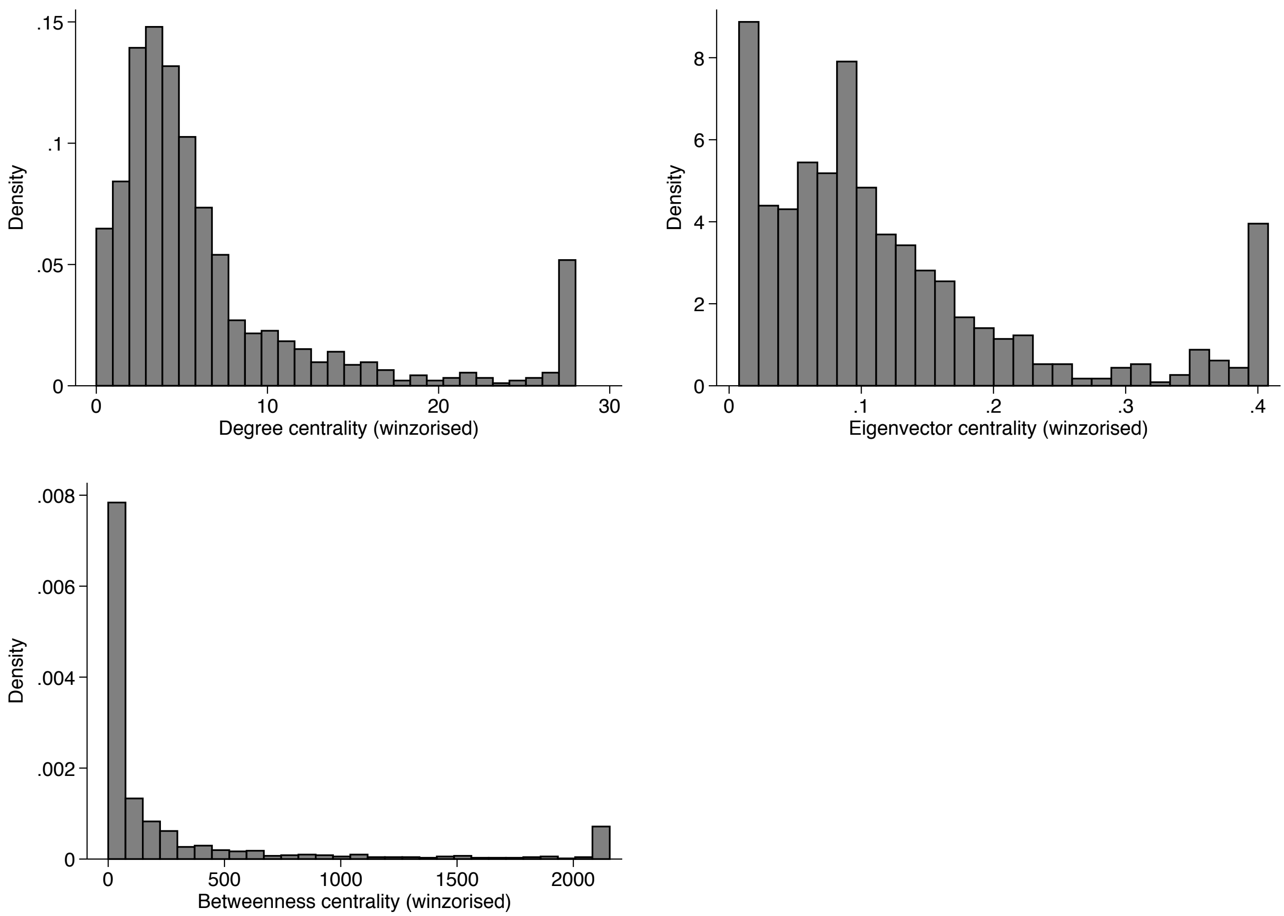
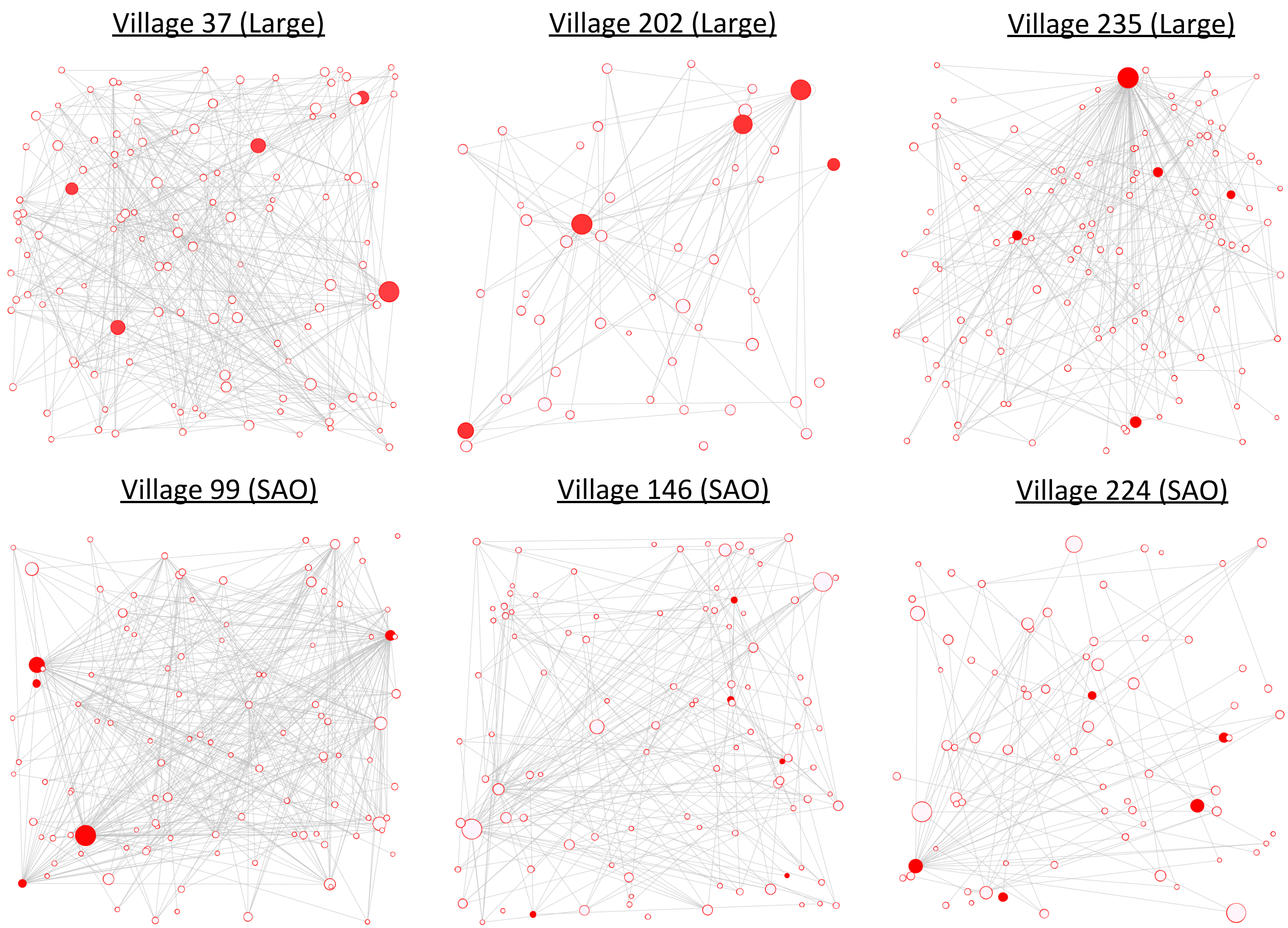


Figure 4: Distribution of centrality measures from social network survey



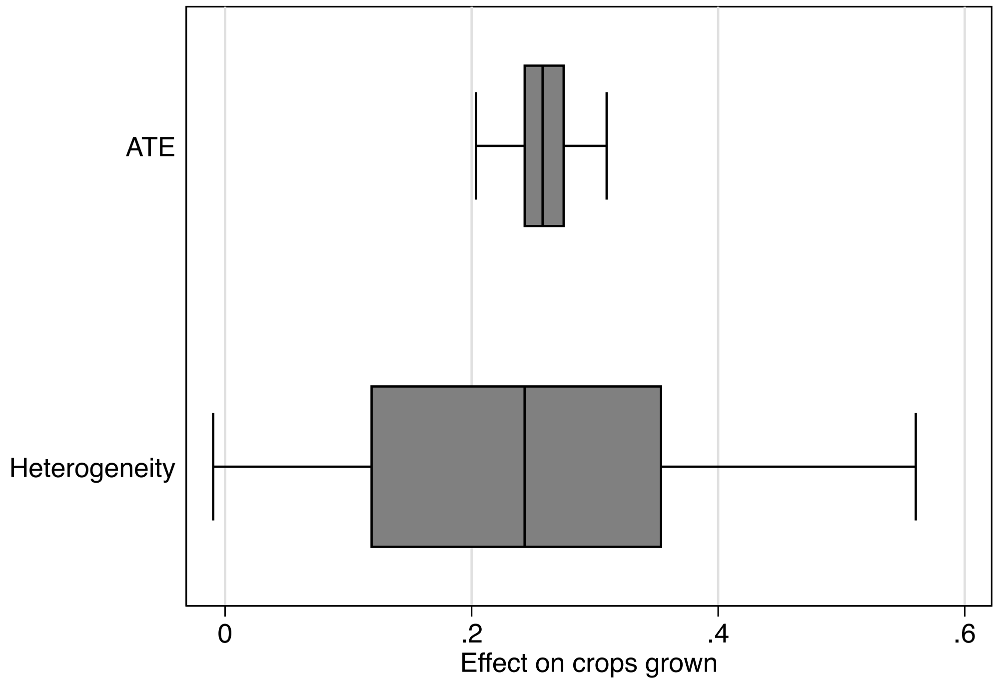
Notes: Figure shows the histograms for the 3 centrality measures from the baseline social network survey with all households (N=21,926).

Figure 5: Example network diagrams for 6 villages in the sample



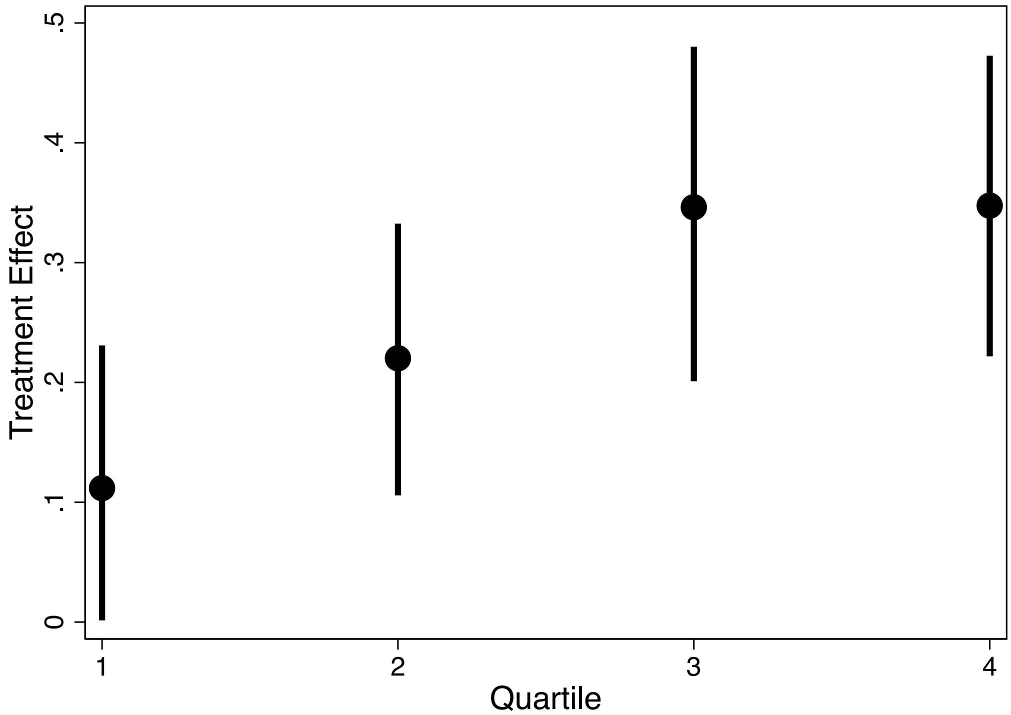
Notes: The figure maps the social network for 6 BD56 villages. The top 3 villages are large-farmer villages and the bottom 3 are villages with SAO selection. The nodes (dots) represent farmers and the size of nodes is proportional to farm size, where larger dots indicate larger farmers. The shaded red dots indicate the 5 farmers chosen as entry points while the hollow red dots denote the remaining farmers.

Figure 6: ATE and heterogeneous effect on number of crops grown



Notes: The figure shows the average treatment effects and the heterogeneous effect on the number of crops grown across across 100 equal-sized splits into training and validation datasets datasets. For each split, we estimate separate LASSO regressions for treatment (BD56) and control (BD51) farmers in the training dataset. In each case the number of crops grown is regressed on a set of 24 covariates, z_i. Using the selected covariates for each group, we calculate the estimated heterogeneity index for each farmer in the validation dataset as $\hat{s_0}(z_i) = E(y_i|D_i = 1, z_i) - E(y_i|D_i = 0, z_i)$. Using the validation dataset, we then regress the observed number of crops on the treatment, $\hat{s_0}(z_i) - \bar{s_0}$, the interaction between treatment and $\hat{s_0}(z_i) - \bar{s_0}$, and upazila fixed effects. The top bar in the figure shows the distribution of the 100 estimates of the ATE (the coefficients on the treatment indicator). The bottom bar shows the 100 estimates of the heterogeneity effect (the coefficient on the interaction between treatment and $\hat{s_0}(z_i) - \bar{s_0}$). The vertical line represents the average across the 100 splits, the box the interquartile range, and the whiskers give the min and max.

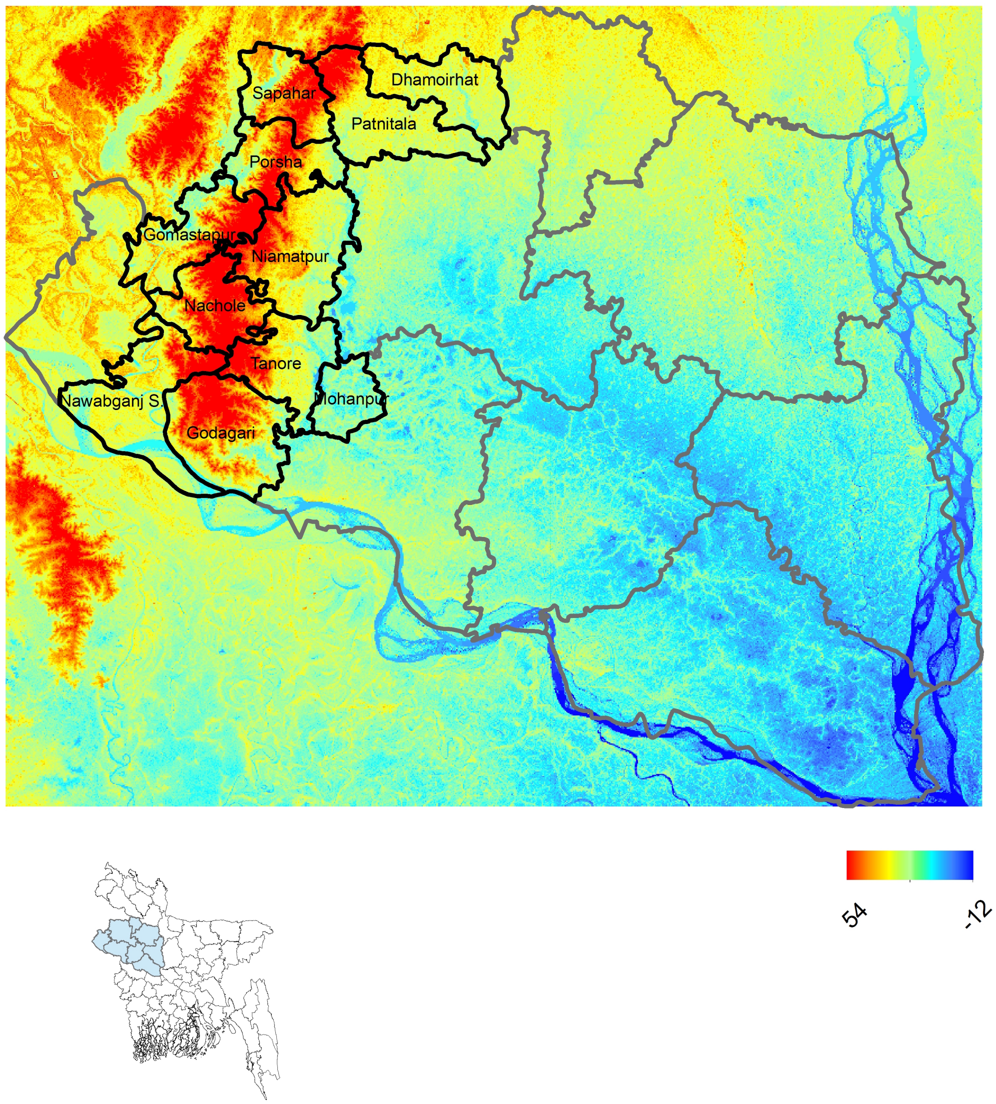
Figure 7: Effects on number of crops grown by quartiles of the predicted effect



Notes: The figure shows the estimated treatment effects by quartile of the heterogeneity index for 100 equal-sized splits into training and validation datasets. For each split, we estimate separate LASSO regressions for treatment (BD56) and control (BD51) farmers in the training dataset. In each case the number of crops grown is regressed on a set of 24 covariates, z_i. Using the selected covariates for each group, we calculate the estimated heterogeneity index for each farmer in the validation dataset as $\hat{s_0}(z_i) = E(y_i|D_i = 1, z_i) - E(y_i|D_i = 0, z_i)$. Using the validation dataset, we then regress the observed number of crops on the treatment and upazila fixed effects *separately for the four quartiles of* $\hat{s_0}(z_i)$. The heavy dots show the averages across the 100 sample divisions while the bands display the range from the 5th to 95th percentiles.

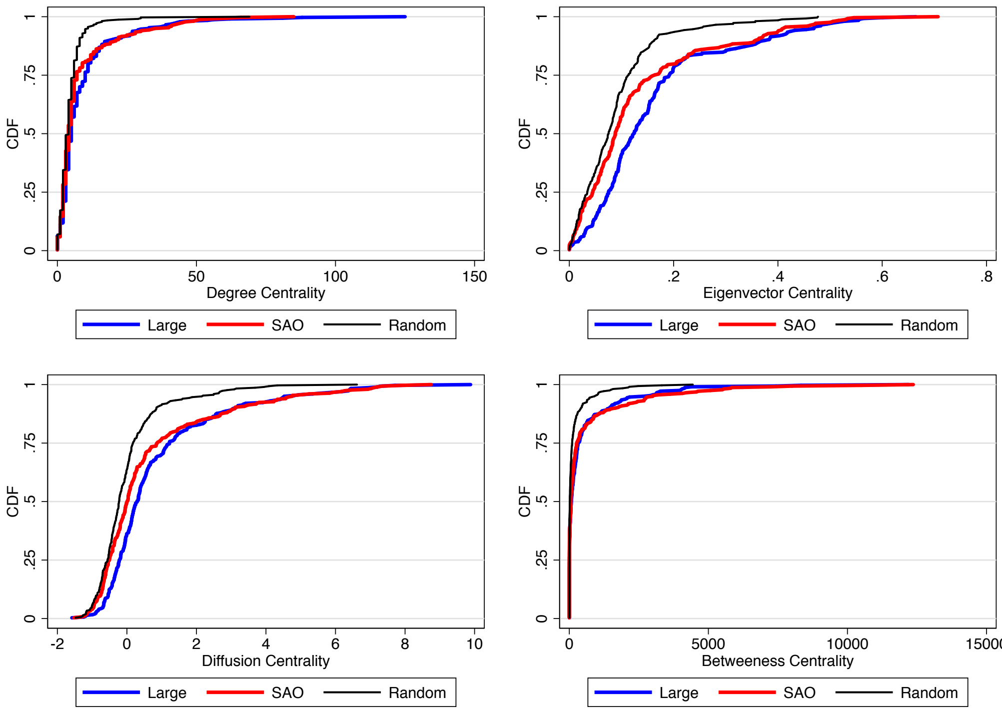
Appendix: For Online Publication

Figure A1: Location of study area



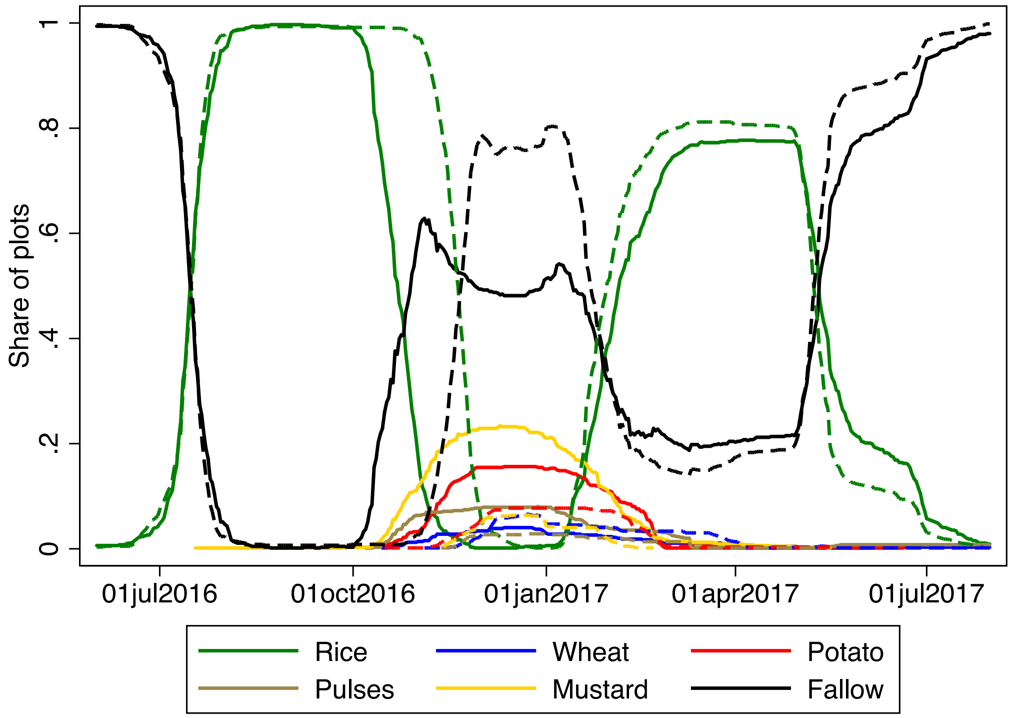
Notes: The figure shows the location of the 11 study Upazilas within Rajshahi district of Bangladesh. The shading corresponds to elevation, measured in meters.

Figure A2: Cumulative distributions of network statistics for different types of entry points



Notes: Each graph shows the cumulative distribution function of the relevant network statistics, separately for the three different types of entry points. The network centrality measures are calculated for each entry point using the baseline social network survey.

Figure A3: Annual land allocation for plots grown with either BD56 or BD51



Notes: The data are for the plots where either BD56 or BD51 was planted by the entry points. The vertical axis gives the share of plots that were allocated to the crop on the date corresponding to the horizontal axis. The sold lines are for the treatment (BD56) farmers and the dashed lines are for the control (BD51) farmers.

Table A1: Balance of household characteristics across treatment arms

	Treatment Arm:							
	Control	Large	SAO	Random	Large + Demo	SAO + Demo	Random + Demo	Joint p-value
Education	4.235	3.951	4.663	4.604	4.216	4.368	4.768	0.381
	(4.314)	(4.202)	(4.521)	(4.293)	(4.125)	(4.466)	(4.235)	
Age	41.356	41.908	41.660	41.650	41.820	40.938	41.261	0.829
	(12.407)	(11.926)	(12.348)	(12.300)	(12.143)	(12.023)	(12.039)	
Owns Shallow Tubewell	0.103	0.149	0.160	0.085	0.072	0.086	0.115	0.356
	(0.304)	(0.356)	(0.367)	(0.278)	(0.259)	(0.280)	(0.319)	
Aman Rice Area (Bigah)	4.071	4.293	5.029	4.221	4.678	4.775	4.265	0.770
	(5.656)	(5.550)	(12.429)	(6.027)	(5.530)	(5.983)	(5.152)	
Aman Other Crop Area (Bigah)	0.348	0.375	0.319	0.462	0.300	0.236	0.395	0.682
	(1.567)	(1.893)	(1.478)	(9.638)	(1.676)	(0.794)	(1.412)	
Boro Rice Area (Bigah)	3.328	3.002	3.812	3.344	2.515	3.289	2.889	0.383
	(4.264)	(4.296)	(5.847)	(5.312)	(4.171)	(4.699)	(4.060)	
Boro Other Crop Area (Bigah)	1.125	1.252	1.332	1.140	1.478	1.100	1.346	0.840
	(2.454)	(2.513)	(3.325)	(2.362)	(3.043)	(2.219)	(2.358)	
Aman Urea Fertilizer (KG per Bigah)	21.427	21.953	21.260	22.161	20.644	21.313	20.588	0.843
	(15.109)	(15.459)	(21.673)	(21.759)	(17.417)	(25.641)	(14.648)	
Aman DAP Fertilizer (KG per Bigah)	15.834	16.110	15.519	16.435	14.889	15.813	15.157	0.326
	(11.660)	(15.169)	(20.073)	(13.662)	(6.739)	(15.346)	(10.332)	
Aman Rice Yield (KG per Bigah)	17.756	17.499	18.063	17.989	17.477	17.570	17.837	0.927
	(3.814)	(4.180)	(3.263)	(3.089)	(3.280)	(3.851)	(3.483)	
Grows Short-Duration Rice	0.011	0.035	0.007	0.004	0.007	0.018	0.008	0.831
	(0.102)	(0.185)	(0.085)	(0.066)	(0.081)	(0.135)	(0.088)	
Grows Wheat	0.236	0.260	0.243	0.190	0.372**	0.231	0.286	0.353
	(0.425)	(0.439)	(0.429)	(0.393)	(0.483)	(0.422)	(0.452)	
Grows Mango	0.086	0.063	0.071	0.093	0.076	0.076	0.063	0.958
	(0.280)	(0.242)	(0.256)	(0.290)	(0.265)	(0.266)	(0.244)	
Grows Potato	0.083	0.052	0.086	0.082	0.074	0.061	0.077	0.872
	(0.275)	(0.222)	(0.281)	(0.274)	(0.261)	(0.239)	(0.266)	
Grows Pulses	0.047	0.103	0.095	0.076	0.077	0.075	0.048	0.518
	(0.212)	(0.305)	(0.293)	(0.265)	(0.266)	(0.264)	(0.215)	
Grows Onion	0.049	0.037	0.050	0.021*	0.047	0.039	0.057	0.275
	(0.215)	(0.188)	(0.219)	(0.144)	(0.211)	(0.194)	(0.232)	
Grows Garlic	0.017	0.009	0.014	0.013	0.009	0.017	0.006	0.429
	(0.128)	(0.096)	(0.118)	(0.113)	(0.095)	(0.130)	(0.078)	

The summary statistics are calculated using the door-to-door census with 21,926 households. Each column shows mean values of each variable for either the control group or one of the six treatment groups. Standard deviations are reported in parentheses below each mean value. Asterisks indicate a statistically significant difference (1% ***, 5% **, and 10% *) between that arm and the control arm, where p-values are calculated by regressing each variable on a constant and indicators for each of the six treatment groups (standard errors adjusted for clustering at the village level). The final column shows the joint p-value of each of these regressions. Aman refers to the wet season prior to the door-to-door baseline (2015) and Boro refers similarly to the most recent dry season (2015-2016). 1 Bigah = 0.33 Acres.

Table A2: Balance of household characteristics for entry points

	Control (BD51)	BD56 Treatment	p-value
Education	5.361	5.392	0.832
	(4.620)	(4.643)	
Age	43.326	43.690	0.577
	(12.585)	(12.225)	
Owns Shallow Tubewell	0.178	0.191	0.528
	(0.383)	(0.393)	
Aman Rice Area (Bigah)	8.553	8.977	0.523
	(11.097)	(10.494)	
Aman Other Crop Area (Bigah)	0.514	0.616	0.415
	(1.648)	(2.037)	
Boro Rice Area (Bigah)	6.483	6.247	0.795
	(7.781)	(8.606)	
Boro Other Crop Area (Bigah)	2.063	2.299	0.454
	(3.709)	(3.951)	
Aman Urea Fertilizer (KG per Bigah)	21.879	21.316	0.541
	(24.672)	(15.565)	
Aman DAP Fertilizer (KG per Bigah)	16.363	15.978	0.735
	(17.281)	(18.909)	
Aman Rice Yield (KG per Bigah)	17.888	17.566	0.207
	(3.797)	(3.881)	
Grows only rice	0.389	0.337	0.132
	(0.488)	(0.473)	
Grows Short-Duration Rice	0.017	0.024	0.488
	(0.129)	(0.153)	
Grows Wheat	0.304	0.312	0.946
	(0.460)	(0.464)	
Grows Mango	0.126	0.117	0.792
	(0.332)	(0.322)	
Grows Potato	0.121	0.098	0.359
	(0.327)	(0.297)	
Grows Pulses	0.078	0.111	0.127
	(0.268)	(0.314)	
Grows Onion	0.048	0.068	0.323
	(0.215)	(0.252)	
Grows Garlic	0.013	0.028	0.065
	(0.115)	(0.166)	

The analysis uses the door-to-door census conducted at the beginning of the experiment. Data are limited to the 1,747 entry points that consented to participate. Each column shows mean values and standard deviations are reported in parentheses below. The final column shows the p-value for the comparison of means, based on a regression of each characteristic on the treatment indicator and Upazila (strata) fixed effects. Standard errors are clustered at the village level.

Table A3: Differences between SAO selected and random farmers, adjusting for farm size

	Degree		Eigenvector		Betweenness	
	(1)	(2)	(3)	(4)	(5)	(6)
SAO-based selection	3.582***	1.917**	0.042***	0.025**	394.084***	281.922***
	(1.044)	(0.838)	(0.012)	(0.011)	(103.649)	(89.698)
Farm Size		0.284***		0.003***		19.124***
		(0.053)		(0.000)		(4.634)
Mean in random group	4.56	4.56	0.09	0.09	164.19	164.19
Number of Observations	639	639	511	511	639	639
R squared	0.037	0.221	0.036	0.175	0.033	0.094

The data are limited to the 640 selected entry points in the random and SAO villages. The dependent variables are degree centrality (columns 1-2), eigenvector centrality (columns 3-4), and betweenness centrality (columns 5-6). Farm size is the total sum of cultivated area (across all three agricultural seasons). The omitted group in each regression is the villages where demonstrators were selected randomly. The standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A4: Analysis of take up by entry points

	(1)	(2)	(3)
Treatment village	-0.025		
	(0.035)		
Demo Village		0.060	
		(0.039)	
Random + Demo			0.085
			(0.062)
SAO			0.076
			(0.066)
SAO + Demo			0.067
			(0.067)
Large			0.053
			(0.067)
Large + Demo			0.157**
			(0.068)
Strata (Upazila) Fixed Effects	Yes	Yes	Yes
Mean in omitted group	0.71	0.65	0.69
Number of Observations	1795	953	953
R squared	0.046	0.059	0.064

The data are from the first midline with 1,795 entry points. Column 1 uses all observations and columns 2 and 3 use only observations from treatment (BD56) villages. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A5: Cultivation practices by treatment

	(1) Harvest Date	(2) 2nd Crop	(3) Boro Crop	(4) N Crops
Treated Village	-25.350***	0.278***	-0.035	0.243***
	(1.384)	(0.035)	(0.039)	(0.046)
Strata fixed effects	Yes	Yes	Yes	Yes
Mean in Control	8.64	0.24	0.82	2.06
Number of Observations	1242	1242	1242	1242
R squared	0.381	0.284	0.257	0.278

The data are limited to the plots where either BD56 or BD51 was planted by the entry points. The dependent variable in column 1 is the date of the harvest, measured in days after November 10, 2016. The dependent variables in columns 2 and 3 are indicators for whether the plot was sown with the Rabi (in-between) crop and the Boro (dry-season) crop. The dependent variable in column 4 is the total number of crops grown across all seasons. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A6: Cultivation practices by treatment and type of entry point

	(1) Harvest Date	(2) 2nd Crop	(3) Boro Crop	(4) N Crops
Treated Village	-25.315***	0.175***	-0.073	0.102
	(1.603)	(0.048)	(0.051)	(0.072)
Treatment Village *	-0.219	0.110*	0.100*	0.210***
SAO	(2.518)	(0.061)	(0.056)	(0.079)
Treatment Village *	0.533	0.173***	0.002	0.174**
Large	(2.302)	(0.065)	(0.062)	(0.085)
SAO	1.535*	-0.040	-0.047	-0.087**
	(0.796)	(0.033)	(0.029)	(0.035)
Large	0.784	-0.051	-0.014	-0.065*
	(0.801)	(0.031)	(0.029)	(0.034)
Mean in Control	8.64	0.24	0.82	2.06
Number of Observations	1242	1242	1242	1242
R squared	0.382	0.291	0.262	0.286

The data are limited to the plots where either BD56 or BD51 was planted by the entry points. The dependent variable in column 1 is the date of the harvest, measured in days after November 10, 2016. The dependent variables in columns 2 and 3 are indicators for whether the plot was sown with the Rabi (in-between) crop and the Boro (dry-season) crop. The dependent variable in column 4 is the total number of crops grown across all season. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A7: Profitability of BD56 and BD51 plots

	(1) Aman	(2) Rabi	(3) Boro	(4) Total
Treated Village	-4576.411***	1436.059***	-850.544	-4199.517***
	(248.751)	(436.314)	(565.122)	(731.727)
Strata fixed effects	Yes	Yes	Yes	Yes
Mean in Control	10309.28	2374.59	10578.11	23459.49
Number of Observations	1200	1205	1228	1156
R squared	0.396	0.442	0.319	0.368

The data are limited to the plots where either BD56 or BD51 was planted by the entry points. The dependent variables are profits per bigah, measured in Bangladeshi Taka (BDT). Approximately 80 BDT=1USD and 3 bigah = 1 acre. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A8: Peer effects on the number of conversations with entry points, separate for villages with and without demonstration plots

	(1)	(2)	(3)	(4)
Peer connections w/ entry points	0.050	0.179*		
	(0.064)	(0.105)		
Peer connections w/ entry points * Demonstration Village		-0.261*		
		(0.137)		
Connected to at least 1 entry point			0.093	0.284*
			(0.098)	(0.147)
Connected to at least 1 entry point * Demonstration Village				-0.383*
				(0.215)
Number of connections	0.001	-0.001	0.001	0.005
	(0.005)	(0.011)	(0.004)	(0.009)
Number of connections * Demonstration Village		0.003		-0.005
		(0.011)		(0.009)
Demonstration Village		0.176**		0.220**
		(0.087)		(0.087)
Strata fixed effects	Yes	Yes	Yes	Yes
Mean of Dep Variable	0.78	0.78	0.78	0.78
Number of Observations	636	636	636	636
R squared	0.301	0.310	0.301	0.312

The dependent variable in all regressions is the number of entry points that the respondent spoke to about BD56. The data are limited to the 64 villages where entry points where chosen randomly and peer effects can therefore be causally identified. The variable *Peer connections w/ entry points* is the number of entry points (from 0 to 5) that the farmer is connected with while *Connected to at least 1 entry point* is an indicator variable for being connected to at least one of the entry points. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.

Table A9: Effects of hitting a large-farmer entry point

	(1) All Villages	(2) Random Villages
At least 1 large entry point	0.079**	0.123*
	(0.032)	(0.071)
At least 1 large entry point * Demonstration Village	-0.101**	-0.131
	(0.048)	(0.104)
Demonstration Village	0.078**	0.119**
	(0.038)	(0.047)
Strata fixed effects	Yes	Yes
Mean of Dep Variable	0.67	0.64
Number of Observations	1919	639
R squared	0.171	0.189

The data are for the 10 random farmers per village that were selected for the information survey. Column 1 is for all 192 BD56 villages, column 2 is for the 64 villages where entry points were selected randomly. *At least 1 large entry point* is an indicator for villages where one of the five largest farmers was selected as an entry point. Standard errors are clustered at the village level. Asterisks indicate statistical significance at the 1% ***, 5% **, and 10% * levels.