Can Network Theory-based Targeting Increase Technology Adoption?

Lori Beaman — Northwestern Univ.
Ariel BenYishay — Coll. of William and Mary
Jeremy Magruder — UC-Berkeley
Ahmed Mushfiq Mobarak — Yale University

January 2021

Abstract

Can targeting information to network-central farmers induce more adoption of a new agricultural technology? By combining social network data and a field experiment in 200 villages in Malawi, we find that targeting central farmers is important to spur the diffusion process. We also provide evidence of one explanation for why centrality matters: a diffusion process governed by complex contagion. Our results are consistent with a model in which many farmers need to learn from multiple people before they adopt themselves. This means that without proper targeting of information, the diffusion process can stall and technology adoption remains perpetually low.

JEL Codes: O16, O13

Keywords: Social Learning, Agricultural Technology Adoption, Complex Contagion, Malawi

* Contact: Beaman: l-beaman@northwestern.edu, BenYishay: abenyishay@wm.edu, Magruder: jmagruder@berkeley.edu, Mobarak: ahmed.mobarak@yale.edu.

We thank the CEGA/JPAL Agricultural Technology Adoption Initiative (ATAI) and 3ie for financial support. Beaman acknowledges support by the National Science Foundation under grant no. 1254380. We gratefully acknowledge the support and cooperation of Paul Fatch, Readwell Musopole and many other staff members of the Malawi Ministry of Agriculture. Thomas Coen, Niall Kelleher, Maria Jones, Ofer Cohen, Allen Baumgardner-Zuzik, and the IPA-Malawi country office provided invaluable support for data collection. Hossein Alidaee and Tetyana Zelenska provided excellent research assistance. We thank, without implicating, the editor and anonymous referees, Arun Chandrasekhar, Matt Jackson, Kaivan Munshi, Chris Udry and numerous seminar audiences for very helpful comments. The study is registered with the AEA registry under AEARCTR-0002017 (see Beaman et al. 2018). IRB approval was provided by Northwestern University (STU00030251) and MIT (COUHES 1005003884). The data repository is openicpsr-130605 (see Beaman et al. 2021). All errors our own.

1 Introduction

Technology diffusion is critical for growth and development (Alvarez et al. 2013, Perla and Tonetti 2014). Information frictions are potential constraints to technology adoption, and social relationships can serve as important vectors through which individuals learn about, and are then convinced to adopt, new technologies.[1] Adoption of apparently productive new technologies has often been frustratingly slow (Ryan and Gross 1943, Munshi 2007, Jack 2011, Qiao 2015). This generates a policy priority: how can policy-makers effectively use social relationships to promote technological diffusion? In this paper, we implement a field experiment in which we choose entry points of information (seeds) into a social network and introduce a productive new agricultural technology via those seeds across 200 villages in Malawi.

A rich empirical literature has documented faster diffusion when technologies were seeded with people who are central in the network (Banerjee et al. 2013 in the context of microfinance in India; Banerjee et al. 2019 in the context of immunization in India; Kim et al. 2015 looking at health behaviors in Honduras). Targeting information to central agents in a network can even work better than broadcasting information widely (Banerjee et al. 2020).

These empirical patterns that establish the importance of centrality may be surprising given recent theoretical discussion by Akbarpour, Malladi and Saberi (2020) (henceforth AMS) which shows that in many canonical diffusion models, adding a few additional seeds leads to more diffusion than targeting central people to serve as seeds. The class of models AMS consider require three conditions: first, agents must adopt a new behavior after a single exposure to someone else who has adopted in the network. This is called 'simple contagion' and is the base for workhorse models like the Susceptible-Infected-Recovered (SIR) model. Second, the time period for adoption is sufficiently long;

[1] Large literatures in economics (Duflo and Saez 2003, Munshi 2008, Magruder 2010, Beaman 2012), finance (Bursztyn et al. 2013), sociology (Rogers 1962), and medicine and public health (Coleman et al. 1957, Flodgren et al. 2007, Oster and Thornton 2012) show that information and behaviors spread through inter-personal ties.

and finally, social interaction within the network is frequent. The intuition for the AMS result is straightforward: whether central or not, people are connected to their local network, so given enough time and enough talk, messages will spread through their connections and quickly reach the well-connected people at the network's center. Adding a few more seeds at random increases the probability that at least one of them will be close to the well-connected center to begin with, making targeting relatively unimportant. However, if any of the three criteria fail, then targeting may be necessary to prevent information frictions from curbing widespread technological diffusion.

Our paper helps bridge the gap between these theoretical and empirical results. We implemented a randomized controlled trial where we used different variants of the threshold model of diffusion (e.g. Granovetter 1978, Centola and Macy 2007, Acemoglu et al. 2011) to choose seeds. This creates a unifying framework which both generates variation in seed centrality across treatment arms and also helps us explore why targeting may matter for technology diffusion.

Our experiment takes place in an important real-world context: agricultural extension services in developing countries. Agricultural productivity growth in Africa has stalled (World Bank 2008), in part because of a slow adoption rate of new technologies. Agricultural extension is the key policy tool governments use to promote technology adoption (Anderson and Feder 2007), and it often relies on social learning.[2] We partnered with the Ministry of Agriculture in Malawi to run an experiment that could enhance the effectiveness of its extension services by partnering with two "seed" farmers in each study village who could induce widespread social learning. The experiment was implemented in 200 villages, with 50 villages in each of the four treatment groups. The specific technology promoted,

[2] A large literature has established that social learning about agricultural practices influences the uptake of new technologies among farmers (Griliches 1957, Foster and Rosenzweig 1995, Munshi 2004, Bandiera and Rasul 2006, Conley and Udry 2010, Burlig and Stephens 2019, Islam et al. 2019).

'pit planting', has the potential to significantly improve maize yields in arid areas of rural Africa.[3] It is a practice that was largely unknown in Malawi, and learning is therefore crucial for the diffusion of this technology.

In the Benchmark treatment, extension agents chose the seeds as they normally would (status quo or picking by experts). In the remaining treatment groups, we strategically chose the seeds using detailed social network data we collected in every village. We ensured that selected seeds in different treatments would inhabit different parts of the network by exploiting variations on the threshold model of diffusion to suggest pairings of seeds that may be more or less effective, given different underlying diffusion processes. In the second treatment group, we selected seeds who would (in theoretical simulations) optimize diffusion over a 4-year period, if the diffusion process is characterized by a complex contagion. Complex contagion is a diffusion process in which technology only diffuses when individuals are connected to at least two knowledgeable farmers. The pair of seeds chosen by this complex contagion treatment are both central in the network. Seed selection in the third treatment is the result of simulations of the simple contagion variant of the threshold model, where farmers only need to know one knowledgeable farmer. In simple diffusion, a single central seed will diffuse to the dense part of the network so that a second seed is best used to diffuse to the more distant periphery. As a result, one of the seeds is network-central while the second person is typically more peripheral. This variant of the model is similar to those considered in the AMS framework. In the final treatment group, we used geography to proxy for social network data, to create a cheaper, "scalable" approach coupled with the complex contagion model. These seeds are typically low centrality, but are close to each other in the network.

[3] It has been shown to increase productivity by 40-100% in tests conducted under controlled conditions (Haggblade and Tembo 2003); in large-sample field tests conducted under realistic "as implemented by government" conditions (BenYishay and Mobarak 2019), and using experimental variation among villagers in the present study.

During the 3-year period of the experiment, pit planting adoption grew from 0% to about 11% in the villages with two highly central seeds. This rate of increase in adoption is comparable to the spread of some very profitable new agricultural technologies (e.g. Munshi 2007). Ryan and Gross (1943) show that it took 10 years for hybrid seed corn to be adopted in Iowa in the 1930s. The adoption rate is 3 percentage points lower in Benchmark villages in years 2 and 3, though only the year 2 differences are statistically significant.

We also test whether the initial advantage of central seeding will likely dissipate over time by examining another important metric: whether any farmers in the village other than the seeds adopt. If there is no diffusion within the first three years, it is unlikely that conversation and experience over longer time horizons will inspire broad technology adoption. We observe a critical failure of expert-based seeding. There is no diffusion of pit planting in 45% of the Benchmark villages after 3 years. In villages with two highly central seeds, there was a 56% greater likelihood ($p<0.01$) that at least one person other than the seeds adopts the technology in the village, relative to the Benchmark. The results clearly indicate that targeting central seeds was necessary to generate adoption of pit planting in Malawi.

We then turn to understanding *why* central targeting was so important in this context. One potential explanation is that the variant of the threshold model that we used to select seeds captures the underlying diffusion process. AMS and Jackson and Storms (2018) demonstrate that targeting on the basis of centrality is more important when there is complex contagion. We show that different thresholds for technology adoption are naturally micro-founded through a naive Bayesian learning model, as we discuss in section 5.1. We anticipate that learning about a new agricultural technology in a developing country is precisely a context in which agents may have a high threshold. This fact would have clear policy relevance: if farmers need to learn from more than one informed connection before they themselves adopt, this would generate a very slow and in many cases permanently stalled adoption pattern, just as we observe in Benchmark villages. Overcoming this problem would necessitate targeting central individuals, as in Banerjee et al (2019).

The diffusion we observe demonstrates several empirical regularities consistent with complex contagion. Though, we note that it is difficult to differentiate complex contagion from other reasons that targeting multiple central farmers may improve technology adoption. We observe three patterns suggested by complex contagion. First, a key insight from the threshold model is that poor targeting could lead to a complete failure of adoption within the village, as we see in our data. Second, consistent with our theory, we show that treatment effects are largest (i) in villages where there is more to learn, because baseline knowledge was lowest, and (ii) among farmers whose land is most suited to pit planting. Third, we use our farmer-level data to provide direct evidence in support of complex contagion. We leverage the random treatment assignment to identify that farmers who are connected to two seeds are more likely to learn about and adopt pit planting than farmers connected to only one seed, holding network position constant.

The targeting method used in this paper is a proof of concept, relying on an expensive method of collecting network data. As such, it is not intended to be practical or directly scalable. The next step is to use cheaper ways to identify highly central individuals. One could use gossips (Banerjee et al. 2019); cell phone data (Bjorkegren 2018, Blumenstock et al 2019) or other administrative data (Bennett and Bergman 2020); or aggregated relational data from a sample of individuals (Breza et al. 2020) to achieve this.[4]

---

[4] In our paper we did one lower cost method, the geography-based targeting strategy. It generated some gains in adoption relative to the benchmark. However, physical proximity does not appear to be a good proxy for social connections in this context. A variety of other papers test the ability of local institutions, such as nominations or focus groups, to identify useful partners: Kremer et al. (2011) identify and recruit 'ambassadors' to promote water chlorination in rural Kenya, Miller and Mobarak (2015) first market improved cookstoves to 'opinion leaders' in Bangladeshi villages before marketing to others, and BenYishay and Mobarak (2018) incentivize 'lead farmers' and 'peer farmers' to partner with agricultural extension officers in Malawi. We also develop an intuitive algorithm to identify central farmers that can be implemented with a small number of interviews, and simulations on our data show that this method would generate large gains in technology adoption.

The rest of the paper is organized as follows. We start with the experimental setting and design, along with details on the implementation of the intervention. Section 3 describes the data. Section 4 presents the average treatment effects on pit planting adoption. In section 5, we propose a theoretical model to explain the results, and provide supplemental evidence of the proposed mechanism. Section 6 discusses cost-effective and policy-relevant alternatives to the data-intensive network-theory based procedures we used in this paper, and discuss other options available in the literature. Section 7 concludes.

2 Field experiment

2.1. Setting

Our experiment on technology diffusion within an agricultural extension system takes place in 200 villages randomly sampled from 3 Malawian districts with largely semi-arid climates (Machinga, Mwanza, and Nkhotakota). Approximately 80% of Malawi's population lives in rural areas (World Bank 2011), and agricultural production in these areas is dominated by maize: 97% of farmers grow maize, and over half of households grow no other crop (Lea and Hanmer 2009). Technology adoption and productivity in maize is thus closely tied to welfare.

The existing agricultural extension system in Malawi relies on Agricultural Extension Development Officers, henceforth extension agents, who are employed by the Ministry of Agriculture and Food Security (MoAFS). Many extension agents are responsible for upwards of 30-50 villages, which implies that direct contact with villagers is rare. According to the 2006/2007 Malawi National Agricultural and Livestock Census, only 18% of farmers participate in any type of extension activity. Extension agents cope with these staff shortages by relying on a small number of lead farmers, who are trained, but not incentivized, to disseminate knowledge via social learning. Against this backdrop of staff shortages, maximizing the reach of social learning in the diffusion process may be a cost-effective way to improve the effectiveness of extension.

2.2. Experimental design

We partner with the Malawi Ministry of Agriculture to select the appropriate technologies to promote and engage extension staff to train exactly two seed farmers in each study village. Our experimental variation only changes how those seed farmers are chosen and holds all other aspects of the training constant.

The experiment has four treatment arms. The Benchmark treatment is the status-quo benchmark, where extension agents were asked to select two seed farmers as they normally would in settings outside the experiment. In the remaining three treatment groups, we strategically chose the seeds to ensure that partner farmers were located in different parts of the network.

We identified farmers with different centrality characteristics in each of the study villages by choosing partners who would be the "theoretically optimal" choices as seeds under alternative formulations of the threshold model (e.g. Granovetter 1978, Centola and Macy 2007, Acemoglu et al. 2011). The threshold model of diffusion postulates that individuals adopt a behavior only if they are connected to at least a threshold number of adopters (λ).[5]

The three treatment arms in which we selected the seeds using the threshold model are as follows:[6]

2. Complex Contagion: this treatment identified seeds by maximizing simulated diffusion when $E[\lambda] \approx 2$ using network relationship data. The two selected seeds are usually both very central in the network.

---

[5] In section 5.1, we will present a micro-foundation which demonstrates how a learning model can generate thresholds. In this version of the model, the threshold is based on the number of people informed about the technology, as opposed to the number of adopters directly.

[6] In other words, we randomly assign the "threshold model formulations" to different villages. Randomization was stratified by district, and implemented using a re-randomization procedure which checked balance on the following covariates: percent of village using compost at baseline; percent village using fertilizer at baseline, and percent of village using pit planting at baseline. Randomization was implemented in each district separately.

3. Simple Contagion: this treatment identified seeds by maximizing simulated diffusion when $E[\lambda] \approx 1$ using network relationship data. In most networks, this identifies one seed who is central and one who is not.
4. Geo Treatment: this treatment typically identifies two seeds who are near each other in the network, but are not be central. This resulted from maximizing simulated diffusion when $E[\lambda] \approx 2$ using network data constructed using only geographic proximity.

The intuition for why the different formulations of the threshold model generates these different targeting strategies is as follows. When many farmers have a threshold for adoption above 1, what this literature calls *complex contagion*, targeting becomes essential because one needs to seed information in part of the network that is dense and where the seeds have connections in common. In this model, identifying two seeds who are both central to the network is important for diffusion.[7] In contrast, when the threshold is generally equal to one, what the literature calls *simple contagion*, identifying a single seed in the central part of the network is sufficient to achieve widespread diffusion. In this case, a second seed is optimally located in a more distant part of the network, so that both the center and the periphery can achieve quick take-up. Identifying the optimal seeds in each of these cases requires rich network data, described in section 3. We also implemented a fourth treatment, "Geo", which substitutes household locations for the network graph under the assumption that nearby households are likely to be connected.

In Online Appendix A.1, we discuss in detail the algorithm used to choose the seeds. Note that in all villages, we can construct which farmers would have been chosen as Simple diffusion seeds, Complex diffusion seeds or Geo seeds, irrespective of the village's assigned treatment condition. We

[7] As we will see later, this feature has significant ramifications for targeting: while randomly selected seeds are quite likely to be relatively close to the center of any network, groups of randomly selected seeds remain unlikely to share ties in common.

call the counterfactual seed farmers "shadow" farmers. We also use the term 'partner' to refer to an individual who would be a Simple, Complex or Geo seed irrespective of whether they are trained and therefore become a seed.[8] We do not observe shadow farmers for the Benchmark treatment.[9]

Table 1 demonstrates the centrality of the two selected partner farmers in each treatment arm. The most central of the two partners (Rank 1 Partners) – as measured by eigenvector centrality[10] in column (1) – is similarly central in both the Complex and Simple diffusion, but less central in Geo. However, the second partner highlights the key difference in the treatments. The second partner in the Complex treatment is much more central than the second partner in the Simple diffusion treatment. In Geo, neither partner is very central, but they are similarly central – highlighting that geography in this context was not a good proxy for social connectedness but that the targeting strategy was *ex ante* similar to the Complex diffusion strategy. If we use an alternative measure of social connectedness, degree (the number of contacts a person has) we see a similar pattern. Both Complex partners have many connections. The most connected partner in the Simple diffusion treatment is similar in the number of contacts to the most connected partner in Complex diffusion, but the second partner is much less connected.

The Benchmark seeds – which were chosen by extension officers using their own criteria – show an intermediate level of centrality as measured by both eigenvector centrality and degree. Overall, this arm of the experiment constitutes a meaningful and challenging test for the network-based targeting treatments since the extension agents were able to use valuable information not

[8] As an example, a Simple partner is a seed if the village is randomized to be a Simple village or a shadow farmer if the village is Complex, Geo or Benchmark.

[9] We did not ask extension workers to name the seed farmers they would choose and then ask them to train other seeds, since we thought it would lead to high non-compliance.

[10] Eigenvector Centrality is weighted sum of connections, where each connection's weight is determined by its own eigenvector centrality (like Google page-rank).

available to researchers, such as the individual's motivation to take on the role. The Benchmark treatment is similar to what the Malawi Ministry of Agriculture and other policymakers would normally do, so this is the most relevant counterfactual.[11]

2.3. Agricultural technologies

In this section, we describe the two technologies introduced to seed farmers and in Online Appendix A.2 we analyze data on crop yields to give further insights into the benefits of the technologies.

Pit planting

Maize farmers in Malawi traditionally plant seeds in either flat land or after preparing ridges. Ridging has been shown to deplete soil fertility and decrease agricultural productivity over time (Derpsch 2003, 2004). In contrast, pit planting involves planting seeds in a shallow pit in the ground, in order to retain greater moisture for the plant in an arid environment, while minimizing soil disturbance. In our sample, pit planting was not widely practiced at baseline: 9 out of 4,004 farmers (0.22%) planted with pits the year prior to treatment. The technique is practiced more widely in the Sahel, and has been shown to greatly enhance maize yields both in controlled trials and in field settings in East Africa, with estimated gains of 50-113% in yields (Haggblade and Tembo 2003, BenYishay and Mobarak 2019). In Online Appendix A.2, we present evidence that pit planting increased yields by 44% (a treatment on the treated estimate) for our trained seed farmers. The enhanced productivity is thought to derive from three mechanisms: (1) reduced tillage of topsoil, which allows nutrients to remain fixed in the soil rather than eroding, (2) concentration of water around the plants, which aids in plant growth during poor rainfall conditions, and (3) improved fertilizer retention.

[11] Normally the Ministry only trains one "Lead Farmer" per village, not two. In most villages, the Lead Farmer will already be established, except for villages in which there hasn't been an extension officer assigned to the village for a long time. The extension agents would have had to select a second seed farmer in Benchmark villages due to the experiment.

Practicing pit planting may involve some additional costs. First, hand weeding or herbicide requirements may increase because less land is tilled, though focus groups undertaken by the authors suggest that weeding demands were actually reduced substantially relative to ridging. Second, digging pits is a labor-intensive task with large up-front costs. However, land preparation becomes easier over time, since pits should be excavated in the same places each year, and estimates suggest that land preparation time falls by 50% within 5 years (Haggblade and Tembo 2003). BenYishay and Mobarak (2019) find that in Malawi, labor time decreases while the change in other input costs are negligible in comparison. Labor costs are minimized when pit planting is used on flat land.

Crop residue management

Seed farmers were also trained in crop residue management (CRM), a set of farming practices which largely focus on retention of crop residues in fields for use as mulch. Alternative practices commonly used by farmers include burning the crop residues in the fields and removing them for use as livestock feed and compost. The trainings emphasized the value of retaining crop residues as mulch to protect topsoil, reduce erosion, limit weed growth, and improve soil nutrient content and water retention. There is little experimental evidence on the impacts of CRM on soil fertility, water retention, and yields in similar settings.

2.4. Seed farmers: descriptive statistics, training, and take up

Extension agents chose the seed farmers in the Benchmark villages, and the researchers chose the seeds in the remaining treatment villages. We already discussed in Table 1 how central the seeds are in different treatments. Online Appendix Table A2 provides some summary statistics describing how the chosen seeds differ in terms of farm size and a wealth index.[12] The most striking pattern is

[12] Table 1 is not demonstrating balance in the randomization of villages across treatment arms. Note that there are only 100 Benchmark farmers since we never observe shadow Benchmark farmers.

that the farmers selected as seeds under the geographic treatment are significantly poorer than other seeds. This is because many households live on one of their plots in Malawi. Households who are geographically close to lots of people will mechanically have less land, and these households tend to be poorer overall.

We observe that there are more households connected to both seeds in Complex villages than in other treatment arms. 35% of our random household sample has a connection to a Simple seedpartner, and 6% are connected to both Simple partners. By contrast, 18% of households are connected to two Complex partners. For the Geo-based partner, 10% of households are connected to two Geo partners. Online Appendix Table A3 displays the distribution of how far – in social distance – households are from the partner farmers in the different treatment arms.

In addition to the names of the two seed farmers, we provided extension agents with replacement names in all non-Benchmark villages in case either of the chosen seeds refused to participate in the training.[13] Refusal was uncommon: extension agents trained 93% of the selected seeds or their spouses. We conduct intent-to-treat analysis using the original seed assignment.

The seed farmers received a small in-kind gift (valued at US$8) if they themselves adopted pit planting in the first year. There was no gift or incentive provided on the basis of others' adoption in the village or the seeds' own adoption in subsequent years. Online Appendix Table A4 demonstrates that the training (and incentive) was effective at inducing adoption, but not perfectly. Seed farmers, relative to the shadow farmers, are more likely to know how to do pit planting and more likely to adopt pit planting during the first agricultural year.[14] 30% of seed farmers adopted pit planting during

[13] As the technologies themselves were new, the extension agents were themselves trained by staff from the Ministry's Department of Land Conservation.

[14] Seed farmers are also more likely to adopt crop residue management (CRM) in year 1. However, by year 2 there is no longer a meaningful gap in the CRM adoption rate, and in fact the adoption rate among shadow farmers is declining over time. Given this pattern, and the fact that CRM was not a "new" technology in this area, we focus our analysis on the adoption of pit planting. We include CRM adoption results in Online Appendix Table A6.

year 1, compared to 5% of shadow farmers (p<.01). Moreover, the adoption rate among seed farmers is the same across all treatment arms: Complex, Simple, Geo, and Benchmark.

Knowledge and adoption rates of pit planting increase among the shadow farmers over time. Knowledge of pit planting among the seeds is declining slightly between year 1 and years 2 and 3, but there remains a significant knowledge gap between seed and shadow farmers even in year 3. Adoption remains more or less constant among seed farmers. Online Appendix A.3 and the notes to Online Appendix Table A4 provide the details on the econometric specification used for these results.

3 Data

After training the seed farmers, we collected up to three rounds of household survey data. Online Appendix Figure A1 shows the timeline of these data collection activities. We describe each major data source in turn.

Social Network Census Data

Targeting based on different network characteristics requires relatively complete information on network relationships within the village (Chandrasekhar and Lewis 2016). More than 80% of households in every sample village participated in the census.[15]

The main focus of the social network census was to elicit the names of people each respondent consults when making agricultural decisions. General information on household composition, socioeconomic characteristics of the household, general agriculture information, and work group membership was also collected. Agricultural contacts were solicited through several prompts.[16] These

[15] We interviewed at least one household member from 89.1% of households in Nkhotakota, 81.4% in Mwanza and 88.6% in Machinga. We interviewed both a man and a woman in about 30% of households.

[16] We first asked in general terms about farmers with whom they discuss agriculture. To probe more deeply, we also asked them to recall over the last five years if they had: (i) changed planting practices; (ii) tried a new variety of seed, for any crop; (iii) tried a new way of composting; (iv) changed the amount of fertilizer being used for any crop; (v) tried a new crop, such as paprika, tobacco, soya, cotton, or sugar cane; or (vi) started using any other new agricultural technology. If they responded affirmatively, we asked respondents to name individuals they knew had previously used the technique in the past and whether they had consulted these individuals. Finally, we asked them if they discussed farming with any

responses were matched to the village listing to identify links. Individuals are considered linked if either party named each other (undirected graph), and all individuals within a household are considered linked.

Sample Household Survey Data

We collected survey data on farming techniques, input use, yields, assets, and other characteristics for a sample of approximately 5,600 households in the 200 sample villages. We attempted to survey all seed and shadow farmers in each village, as well as a random sample of 24 other individuals, for a total of about 30 households in each village.[17] In villages with fewer than 30 households, all households were surveyed. Three survey rounds were conducted in Machinga and Mwanza in 2011, 2012 and 2013, and two survey rounds were conducted in Nkhotakota in 2012 and 2013.[18] The first round asked about agricultural production in the preceding year—thus capturing some baseline characteristics—as well as current knowledge of the technologies, which could reflect the effects of training. Since the data was collected at the start of a given agricultural season, but after land preparation was complete, we observe three adoption decisions for pit planting for farmers in Mwanza and Machinga, and two decisions for farmers in Nkhotakota. Since crop residue management (CRM) decisions are made the end of an agricultural season after harvest, we observe CRM decisions for two agricultural seasons in Mwanza and Machinga, and one in Nkhotakota.

relatives, fellow church or mosque members, or farmers whose fields they pass by on a regular basis, or if there are any others with whom they jointly perform farming activities. We also elicited their close friends and contacts with whom they share food, though we did not include these contacts as agricultural connections for the purposes of our network mapping.

[17] In Simple, Complex and Geo villages there were 6 (2x3) seed and shadow farmers to interview, while in Benchmark villages there were 8 (2x4) seeds and shadows. Recall we do not observe Benchmark farmers in Simple, Complex and Geo villages.

[18] Unanticipated delays in project funding required us to start training of extension agents and seed farmers in Nkhotakota in 2012 instead of 2011 as we did in Mwanza and Machinga.

Randomization and Balance

Randomization was stratified by district, and implemented using a re-randomization procedure which checked balance on three village-level covariates.[19] Online Appendix Table A5 shows how observable baseline characteristics from the social network census vary with the treatment status of the village. The table also shows p-values from the joint test of all treatment groups. The table notes provide details on the specification used. Few differences across treatment groups are statistically significant. Overall, the joint test reveals no differences for 10 out of 12 variables. Farm size is the most concerning: farmers in the Benchmark villages have larger farm sizes on average than farmers in Simple and Complex villages, and the joint test across the network treatment variables is significant at the 10% level. Additional analysis available from the authors controls for this variable in all specifications and finds that all results are robust to this control.

4 Average treatment effects on diffusion

In this section, we report experimental results on village-level outcomes across the four treatment arms.

4.1 The advent of diffusion

We focus on the advent of diffusion in our sample villages as a key outcome. While the speed of diffusion may matter in some settings, we think that a key policy goal is to have diffusion start in as many villages as possible. If there is no diffusion in a village after 3 years, it is likely that the technology will never be widely adopted.

Therefore, we first focus on 'any adoption' as an indicator for villages which have at least one household (other than the seeds) that adopted pit planting. Our village-level regression is as follows:

$$Y_v = \alpha + \beta_1 Complex_v + \beta_2 Simple_v + \beta_3 Geo_v + X\gamma + \varepsilon_v$$

[19] The three variables include: percent of village using compost at baseline; percent village using fertilizer at baseline, and percent of village using pit planting at baseline. We control for these variables in the analysis.

Where $X\gamma$ are variables used in the re-randomization routine, specified in the table notes, and district fixed effects. The results are reported in Table 2. First note that in year 2, we observe the start of the diffusion process in only 42% of Benchmark villages. This increases in year 3 to a modest 54%. This is evidence that this is an environment where igniting diffusion is challenging. The first two columns of Table 2 show that the propensity for 'any adoption' in year 2 is statistically significantly larger in villages where both seeds were highly central (Complex diffusion treatment) relative to Benchmark villages. The 25 percentage point gap is large relative to the 'any adoption' rate of 42% in our Benchmark villages. The 'any adoption' rate in Complex villages is also 15 percentage points larger than in Geo villages ($p = 0.10$) and 10 percentage points larger compared to villages assigned to the simple diffusion treatment ($p = 0.30$). In year 3, Simple, Complex and Geo villages all attain a statistically higher rate of 'any adoption' than Benchmark villages. 85% of Complex villages had at least one non-seed adopter, compared to 73% of Simple and Geo villages and 54% of Benchmark villages.

4.2 Adoption rates across treatment arms

We also look at the speed of diffusion, captured by the adoption rate. Columns (3) and (4) in Table 2 document treatment effects on the adoption rate, which is defined as the proportion of non-seed farmers who adopted pit planting in each agricultural season. Both Simple and Complex diffusion villages have higher adoption rates relative to the Benchmark in year 2. Compared to the Benchmark rate of 3.8%, Complex and Simple villages both experience a 3.6 percentage point higher adoption rate. We cannot reject that the adoption rates are the same in Simple, Complex and Geo villages. The adoption rate increases across all four types of villages in year 3. The adoption rate increases in the Benchmark villages, the reference category, from 3.8% to 7.5% from years 2 to 3. With the smaller sample size of 141 villages in year 3, we cannot reject that the adoption rate is the same across all treatment types, though the point estimate on Complex remains the largest, and is equal in magnitude to the effect size observed in year 2. The adoption rate in Complex villages in year 3 is 11%.

4.3 Discussion: Why did targeting central seeds matter?

Targeting central seeds as in the Complex treatment led to higher adoption and was particularly important for avoiding the scenario in which no farmers adopted at all. In many diffusion models, this total failure of adoption would be quite surprising: generically, nearly everyone is connected to the network, and so some diffusion should have taken place in Benchmark villages, too. Akbarpour, Malladi and Saberi (2020) [AMS] describe characteristics of diffusion processes where targeting has an advantage. First, in early stages of diffusion, targeting will speed up the adoption process. But with time, the diffusion process in Benchmark villages could catch up to Complex diffusion villages. However, the results in section 4.1 suggest that for many villages, a longer time horizon will not lead to substantially more adoption. With virtually no adoption after 3 years, it is unlikely those villages will ever have widespread adoption of pit planting.

Second, when information sharing is sufficiently infrequent, targeting may matter. We use data on conversations about pit planting that respondents had with others in the village to look directly at this explanation. Each respondent was asked questions about their relationship and conversations with the two seed farmers, randomly selected shadow farmers, and a random sample of other village residents.

Approximately 18% of farmers report talking about pit planting with trained seeds each year. This is a reasonably high rate of information passing, such that we would anticipate that the AMS dynamics of information eventually reaching the central farmers would be at play in a SIR-type model. In fact we also observe that many (13-14% of respondents) are also having conversations with shadow partners about pit planting, likely because of those very dynamics. We can provide a lower bound on how much the experiment induced additional conversations about pit planting using the random variation in the experiment itself. For example, we compare the frequency of conversations with the Complex seed farmers in Complex diffusion villages, to the frequency of conversations with Complex shadow farmers in other villages. This is a conservative, downwardly-based estimate as many (and perhaps most, given how unusual pit planting was at baseline) of the conversations with Complex shadow farmers will also have occurred because of the experiment. However, this conservative estimate is sufficient to argue that it is unlikely that farmers are not talking enough to generate an adoption cascade.

Table 3 shows that the experiment indeed induced seed farmers to discuss pit planting with fellow villagers using the following econometric specification.

$$Y_{ij} = \alpha + \beta_1 Trained_j + \delta_1 ComplexPartner_j + \delta_2 SimplePartner_j + \delta_3 GeoPartner_j + X\gamma + \varepsilon_v$$

Y_{ij} is an indicator for whether respondent *i* discussed pit planting with partner (either seed or shadow) farmer *j*. $Trained_j$ is 1 if a partner was trained in pit planting[20] and 0 otherwise. $ComplexPartner_j$ is an indicator for whether the partner *j* is a complex partner (either seed or shadow) and $SimplePartner_j$ and $GeoPartner_j$ are defined analogously. $X\gamma$ are variables used in the re-randomization routine, specified in the table notes, and district fixed effects. β_1 is our coefficient of interest. Since we only consider conversations with treated partners and shadow partners, whether a potential conversation partner was actually trained is random and we can interpret the effect of training on conversations as exogenous.

We find that about 5% (ranging from 3.7% in year 1 to 6.4% in year 3) more respondents report a conversation about pit planting with trained seeds than with untrained seeds. In Online Appendix A.4, we suppose that only these 5% of conversations are attributable to the training, and find that this

[20] This arises for complex partners in Complex diffusion villages, Simple partners in Simple diffusion villages, and Geo partners in Geo villages.

lower bound exceeds the conversation threshold AMS establish in which random seeding should generate an adoption cascade in simple diffusion models.

An additional possibility that AMS highlight is that targeting may be more important in a range of diffusion models outside of the class of "simple diffusion" models they consider; in the next section, we consider an important model outside of this class: the threshold diffusion model.

5 Complex contagion

In this section, we propose that the threshold model we used to select seeds offers a potential explanation for why targeting central seeds matters for diffusion. As AMS make clear and Jackson and Storms (2018) formalize, targeting will be advantageous relative to random seeding when diffusion is governed by a threshold model. The intuition for the importance of targeting in the threshold model is illustrated with the example network shown in Figure 1. In this thought experiment, we train two seed farmers in period 0 such that they are fully informed about a new technology. Diffusion occurs as farmers become informed in subsequent periods.

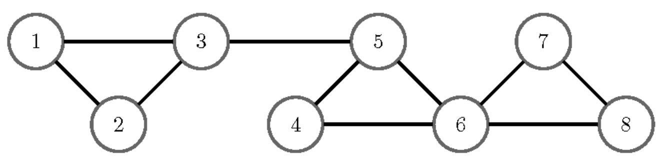

Figure 1: An example network

Suppose that farmers in this network become fully informed of a new technology if anyone they are connected to has been fully informed. This is what we call simple contagion. In this network, the ideal seed farmers will be farmer 6 and then either farmer 1, 2 or 3. With any of these configurations, all farmers are informed in period 1. In general, quickly diffusing information about the new technology will be easy: in 70% of all possible seed pairings, all farmers will be fully informed by the end of the second period. Targeted seeding is not necessary in this model.

However, if farmers need to know two other farmers before they have sufficient information to be fully informed, the diffusion process looks very different. Consider seeding farmers 5 and 8. During the first period, farmer 6 will become informed. In the second round, farmers 4 and 7 are informed. The diffusion process then stops with 3 out of a possible 6 non-seed farmers informed. There are 4 seed pairings which can achieve this 50% adoption rate, but it is not possible to get any higher.

Crucially, without a focus on targeting, there is a good probability that there is no diffusion: in 40% of seed pairings, there is no diffusion whatsoever. Threshold models therefore generate the empirical result we observed: when non-central farmers are trained, there may be no diffusion at all.

In the next subsections, we will provide a micro-foundation of the threshold model based on social learning. We then provide three pieces of empirical evidence that are consistent with the idea that complex contagion is a reason why targeting central seeds was effective in this setting.

5.1 A micro-foundation for the threshold model of diffusion

Social learning is known to be important in technology adoption decisions (e.g. Griliches 1957, Conley and Udry 2010). This section demonstrates how social learning naturally micro-founds the threshold model. Our theoretical framework considers a learning environment with three characteristics. First, we suggest that adoption of a new technology takes place only when farmer beliefs about the profitability of the technology pass a critical threshold. Second, there are limited inherent benefits to learning about a technology if farmers are not ultimately persuaded to adopt it. Third, learning is costly: farmers must invest time to learn about and master a new productive technology, and revealing ignorance may subject them to social costs (e.g. Banerjee et al. 2020; Chandrasekhar, Golub and Yang 2019).

These facts together mean that technology diffusion will be characterized by rational ignorance: farmers will be unwilling to pay learning costs in environments where they are unlikely to update their beliefs enough to adopt the new technology. Moreover, if farmers aggregate multiple signals to update their beliefs via Bayes' rule, technology adoption will be characterized by multiple equilibria: when few are informed about the technology, few will be willing to pay learning costs and few will adopt; when many are informed, more farmers will pay learning costs and ultimately adopt.

In Online Appendix A.5, we adapt the naive learning model in Banerjee et al. (2016) to include small costs of learning. We model technology diffusion as a learning process with three key phases: (1) the farmer has to decide whether to acquire information, (2) she combines the new information with her priors via Bayes' rule , and (3) based on her revised information set, she then decides whether to adopt the new technology. We demonstrate that farmers who learn in this way follow a threshold model (Granovetter 1978, Acemoglu et al. 2011): a farmer will become informed about a new technology once at least λ of her connections become informed. Since uninformed farmers do not adopt, this means that farmers without sufficient informed connections will not adopt.

Taking the model to the data, the micro-foundation is useful for a few purposes. First, it demonstrates that agricultural learning can lead to diffusion with thresholds. In our micro-foundation, thresholds arise because farmers rationally choose not to learn when there is insufficient information in the network to change their behavior. This suggests that the learning problem could generate the results in section 4.1: because farmers need to be exposed to multiple informed agents to make an informed adoption decision. As a result, targeting central farmers is critical for diffusion: poor targeting may lead to no diffusion at all. Second, we can characterize the learning problems which lead to higher thresholds: thresholds are higher when the expected benefits are lower, or when signals are noisier. We therefore learn that seeding matters more in contexts where (in expectation) the benefits of adoption are relatively low; or in cases where a given signal is quite likely to be noisy. We return to this prediction in section 5.3.

5.2 Complex contagion model simulations compared to empirical results

There are three main pieces of evidence that suggest that complex contagion may have led to higher diffusion in the villages in which both seeds had high centrality. First, a key consequence of not targeting the right seeds in an environment where a sizeable fraction of agents have a threshold above 1 is that the diffusion process can be completely stalled. We will discuss this evidence in this subsection. Second, we show heterogeneous treatment effects to argue that the complex diffusion treatment was particularly effective in exactly the contexts in which we would anticipate the treatment to be effective. And finally, we analyze individual-level data to show that farmers who were directly connected to two seeds – as opposed to just one seed – are most likely to adopt pit planting.

Table 2 already demonstrated that that complex diffusion led to a higher rate of "any non-seed adoption." Figure 2 presents the same evidence but side by side with what our simulations predicted. The left part of Figure 2 shows the *predicted* fraction of villages with 'any adoption' from simulating the model for all sample villages when $\lambda=1$ (Simple contagion) and $\lambda=2$ (Complex contagion).[21] Since the goal is to compare these simulations to the actual data, we design the simulations to reflect the fact that we only observe a random sample of households in these villages.[22] The right part of Figure 2 shows the empirical counterpart: 'any adoption' rates in the data in years 2 and 3.

---

[21] These simulations exclude 12 villages where at least one of the extension worker chosen seeds (Benchmark) was not observed in our social network census. This occurred because the spatial boundaries of villages are not always clearly delineated in Nkhotakota.

[22] The simulations use the full social network to predict becoming informed, measured here through adoption. We then sample from the full network to better mimic our data. In the model, the rate of any adoption is identical in years 2 and years 3. If there was no adoption by year 2, there is no way there will be any additional adoption taking place in year 3. The sampling process, however, generates the increase over time observed in the figure. If the rate of adoption is low, as is empirically the case, then a random sample may miss all adopters. As the number of adopters increases over time, the random sample is more likely to pick up an adopter and hence the rate of any adoption increases over time in the figure.

When the threshold is set to λ=1, diffusion is predicted to be widespread. In year 2, 85% of villages where Geo and Benchmark partners were trained are predicted to have some sampled diffusion, and that rate goes up to 94% with Simple and Complex partners. The predicted rates of 'any diffusion' are even higher in year 3.

The risk of no diffusion increases if the diffusion process is characterized by complex contagion. In that case, the model predicts that more than half of the villages assigned Simple, Geo or Benchmark partners will not see any sampled diffusion at all in year 2. In contrast, when Complex seeds are trained, 70% of villages are predicted to experience some diffusion in year 2.

Comparing the theoretical simulations to the data on the right side of Figure 2 shows that the data are more consistent with the patterns generated by a complex (rather than simple) learning environment in three distinct ways. First, simple contagion simulations suggest that we should observe a much higher fraction of villages with some adoption than is true in the data. Second, simple contagion predicts that the 'any adoption' outcome should not be very sensitive to the identity of the seed farmer who is initially trained. In contrast, the identity of the seed farmer dramatically alters this outcome in the data. Finally, the complex contagion simulations predict that the Complex partners will maximize the fraction of villages with some adoption, which we observe in the data.

5.3 Heterogeneity analysis

The micro-foundation of the threshold model suggests that targeting Complex diffusion seeds will be particularly effective in contexts in which the information about pit planting will be most valuable. We use two different approaches to identify groups of such farmers. First, the Ministry of Agriculture recommends pit planting only for flat land, and labor costs of pit planting are lower on flat land.[23] Focus group discussions in our sample villages confirmed that villagers thought pit planting

[23] Pit planting is possible on land with some slope, but in those cases, the pits need to be constructed differently, and our extension workers were not trained on that technique.

was more suitable for flat rather than sloped land. We therefore expect farmers who own flat land will be most interested in information about pit planting. The second heterogeneity test we do exploits variation in knowledge about pit planting at baseline. While pit planting is in general a new technology in Malawi, there is heterogeneity across villages in how novel it is. In the median village, 4.3% of farmers reported having ever tried pit planting at baseline while 0.2% were currently practicing pit planting across all villages.

Table 4 explores the heterogeneity in treatment effects across these two dimensions, by interacting the randomized treatments with an indicator for "Farmer likely to receive a Good Signal." This "Good Signal" variable is first defined as the farmer having flat land in columns (1) and (2), and then re-defined as "Village with lower-than-median familiarity with the technology at baseline" in columns (3) and (4). "Bad signal" refers to the converse of these characteristics. The equation estimated:

$$y_{ivt} = \beta_0 + \beta_1 Simple_v * Bad\ Signal + \beta_2 Complex_v * Bad\ Signal + \beta_3 Geo_v * Bad\ Signal + \beta_4 Good\ Signal + \beta_5 Simple_v * Good\ Signal + \beta_6 Complex_v * Good\ Signal + \beta_7 Geo_v * Good\ Signal + \delta X_v + \epsilon_{ivt}$$

The reference group comprises of farmers who are likely to receive a bad signal in Benchmark villages. Our hypothesis is that among those who receive a positive signal, we will observe more diffusion in Complex villages if the true model is Complex.

Columns (1) and (2) show that adoption in year 2 is higher for farmers who have flat land in Simple, Complex and Geo villages compared to farmers with flat land in Benchmark villages. In year 3, we see that Complex villages continue to have a larger adoption rate than Benchmark villages for farmers with flat land. Columns (3) and (4) show that the Complex treatment performs best in villages where the technology was relatively novel. In this sub-sample, the adoption rate is statistically significantly higher in Complex diffusion treatment villages compared to both the Simple and the Benchmark treatments in year 3.

To summarize, these heterogeneity tests indicate that targeting central seeds is most effective precisely in the types of villages and for the types of farmers where information was most valuable, as the theoretical model helped us predict.

5.4 Knowledge and adoption of farmers by social distance to seeds

In this subsection, we provide more direct evidence in line with the Complex Contagion model. We look at knowledge of pit planting and adoption decisions by individuals, as a function of how many seeds they are connected to. If thresholds are larger than one, those with connections to 2 seeds should be the most likely to adopt pit planting. We estimate the following equation:

$$Y_{iv} = \alpha + \beta_1 1TSeeds_{iv} + \beta_2 2TSeeds_{iv} + \beta_3 1Simple_{iv} + \beta_4 2Simple_{iv} + \beta_5 1Complex_{iv} + \beta_6 2Complex_{iv} + \beta_7 1Geo_{iv} + \beta_8 2Geo_{iv} + \theta_v + \varepsilon_{iv}$$

$1TSeeds$ is an indicator for the respondent being directly connected to exactly one seed farmer, and $2TSeeds$ indicates the respondent was directly connected to two seed farmers. Seeds and shadows are removed from the analysis. Since network position is endogenous, we also control for whether an individual is connected to one or two Simple, Complex or Geo (actual or shadow) partners, but these coefficients are not displayed in the table. Identification therefore comes from variation in the experiment. As an example, we can compare two farmers who are both connected to two Simple partners, but where one farmer is in a village randomly assigned to the Simple treatment and his friend is trained as the seed, while the other farmer's friend was not trained.

In the theoretical model, individuals have to become informed prior to adopting. As an empirical matter, it is unclear what level of knowledge is associated with "being informed" as used in the model. In Table 5 we therefore consider three variables which represent increasing levels of information: whether the respondent has heard of pit planting; whether the respondent knows how to implement pit planting; and whether the respondent adopted pit planting (which implies not only knowledge but also that the signals that the respondent received were sufficiently positive). In year 1, the training led to more information transmission to those directly connected to seeds. In particular, those who have a direct connection to both seed farmers had the most knowledge. This is true for both measures of "knowledge": whether the respondent had heard of pit planting and whether they reported being capable of implementing it. Respondents with two connections are 8.4 percentage points more likely to have heard of pit planting than those with no connection to a seed. This represents a 33% increase in knowledge relative to the mean familiarity among unconnected individuals. This effect is also statistically significantly different from the effect of being connected to one seed (p = 0.02). They are also 6.2 percentage points more likely to report knowing how to pit plant, a 108% increase over unconnected individuals and again significantly different from the effect of being connected to one seed (p = 0.072). These knowledge effects are suggestive – but not conclusive – of a complex contagion process (λ=2) rather than simple contagion. The increased awareness of pit planting and knowledge of pit planting among households connected to two seeds persists into year 2 (columns 2 and 5), and two connections is again significantly more advantageous than one connection (p = 0.04 and 0.095, respectively).

We see no effect on adoption in the first year (column 7) among individuals directly connected to either one or two seeds. However, we do observe an adoption effect in year 2. This temporal pattern of results is consistent with the set-up of our theoretical model: individuals become informed in year 1 and then some choose to adopt in year 2. Column (8) shows that households with two connections to trained seeds are 3.9 percentage points more likely to adopt in the second year than those with no connections, which represents a 90% increase in adoption propensity. Though the point estimate of the effect of 2 connections is considerably larger than the effect of a connection to one seed (3.9 pp compared to 1.2 pp), we cannot statistically reject that households with a connection to only one treated seed adopt less frequently (p = 0.16). We also observe that individuals who are within path length 2 of at least one seed (that is, a friend of a friend) are 2.2 percentage points more likely to adopt.

The predictions of the model for which individuals learn about pit planting are weakened as time passes and knowledge diffuses through the network. In all three of our dependent variables, this diffusion can be observed through large increases in knowledge and adoption over time in our reference category: individuals with no direct connections to a seed. Among this group awareness increases from 22% to 39% from year one to three, while "knowing how" to pit plant increases from 6% to 15% and adoption increases from 1% to 4%. In principle, this diffusion should reduce power on our exogenous variation, as the number of connections to informed individuals becomes less correlated with the number of signals available to farmers. In practice, by year 3 we still see significance on the effects of two direct connections on one of our two knowledge variables ("knowing how" to pit plant, column 6), but we no longer see significant differences from direct connections in adoption or awareness of pit planting. Consistent with the hypothesis that this loss in precision is due to diffusion in the network, we see that adoption increases among those at moderate distance to the seeds in year 3: column (9) shows that households within path length 2 are more likely (3.7 pp) to have adopted over those who are socially more distant.[24]

In summary, analysis using individual-level data demonstrates that individuals who are initially close to the trained seeds are more likely to adopt than individuals with no direct connections – as one would expect if the experiment is inducing social network-based diffusion. The data also suggest that having two direct connections – and not just one – is important for diffusion. This is further evidence consistent with the complex contagion model: farmers may need to know multiple informed connections before becoming informed, and then subsequently adopting, themselves.

[24] This is a lower power test of the model than the direct connections test as it is imperfectly correlated with the number of informed, indirect connections to seeds (which is unobserved). We do not see a significant effect of this variable on knowledge outcomes, though coefficients are positive.

6 Cost-effective, policy-relevant alternatives to data-intensive targeting methods

Our experiment was designed to be a proof of concept. We showed that targeting multiple highly central farmers improves technology diffusion, but eliciting the social network map to achieve these gains is expensive. Our geography-based treatment arm was an attempt to assess how much of the diffusion benefit derived from applying network theory could be achieved without having to resort to expensive data collection methods (since each household's physical location is much easier to observe than network relationships). This specific approach was not an unqualified success. Online Appendix Table A2 showed that Geo seeds tended to have less land and were therefore poorer. Therefore, while the idea of using geography as a proxy for one's network may be intuitive, the implications of geographic centrality may be context-specific, and inappropriate as a network-based targeting proxy in some cases.

Combining our experimental results with research on other inexpensive procedures to identify the optimal seeds under complex contagion theory would make network-based targeting more policy relevant and scalable. A few recent papers have suggested promising, less expensive methods for inferring network characteristics. Banerjee et al. (2019) suggests that despite the implicit challenges in learning about network structure, the simple question of "if we want to spread information about a new loan product to everyone in your village, to whom do you suggest we speak?" is successful in identifying individuals with high eigenvector centrality and diffusion centrality, who ultimately improve the diffusion process. Breza et al. (2020) suggest that aggregate relational data collected from a smaller sample combined with a census can yield accurate estimates of network characteristics. Mobile phones may also be a way to inexpensively identify highly central individuals (Bjorkegren 2008, Blumenstock et al. 2019).

While we cannot test the viability of these approaches with our data, we can explore via simulations some alternate strategies that extension officers could use to identify useful partners. We suppose that an extension agent enters a village and randomly selects a small number of farmers to interview, and only asks one question from our social network census: "Do you discuss agriculture frequently with anyone in the village? What is the name of the person you speak with about agriculture frequently?" The response to this question generates a small list of names. The extension agent can then use the responses to select any follow-up interviews. Using simulations, we predict that strategies which leverage the highest degree respondent from the random sample can approach the performance of the optimal targeting. More specifically, we can achieve 73% of the optimal adoption rate with just 2 total interviews and 84-90% of the targeting gains with around 7 interviews.[25]

7 Concluding remarks

This paper provides evidence that diffusion of a new technology is accelerated by targeting information to central nodes within a social network. In a field experiment conducted in collaboration with the Ministry of Agriculture in Malawi, we selected farmers at different positions in the village network, leveraging threshold theory to suggest useful partners under different diffusion mechanisms. We found that farmers were most likely to adopt pit planting in villages where the two trained seed farmers were centrally located within their villages' social network. These partners were chosen to optimize diffusion under complex contagion: when thresholds for diffusion were larger than one.

Because two central partners may be optimal under several diffusion models, we also explore whether the underlying diffusion process is well-characterized by complex contagion. We present multiple pieces of evidence consistent with this mechanism. In particular, we demonstrate that a total failure of diffusion occurs frequently in villages where experts selected the seed farmers. High thresholds can generate this risk. Moreover, farmers who are connected to two seed farmers are also

[25] See Online Appendix A.6 for more details and alternative targeting strategies.

most likely to adopt pit planting in the second year of the experiment. This is consistent with the fact that under complex contagion, multiple connections to seeds are needed before farmers adopt.

The methodological approach in this paper is not directly scalable for policy because of the high costs of collecting network data. But there is very promising work in the literature on ways to cost-effectively identify central individuals within social networks (Banerjee et al. 2019). Our simulations also suggest that with only about 7 interviews per village, it is possible to identify individuals who can trigger the diffusion process. There are also additional options available to identify central nodes within a network depending on the context, including new approaches such as cell phone data.

Our paper also suggests a direction for future research. We provide evidence that agricultural technologies need to be seeded with multiple, central individuals to encourage adoption; this and other evidence in this paper is inconsistent with "simple" diffusion models. In contrast, the evidence in this paper is consistent with models where diffusion requires a concentration of information, such as complex contagion. Further research is needed to understand if farmers often face high thresholds to adoption. Our micro-founded diffusion model suggests a key dimension to consider when assessing if contagion is likely to be simple or complex: the noise of the signal. Rosenzweig and Udry (2020) highlight the importance of aggregate stochastic shocks in distinguishing the returns to agricultural investment, microenterprise investment, and human capital from large-scale survey data. Farmers, entrepreneurs, and parents likely have access to far fewer data points than these large-scale surveys when they attempt to infer the returns to investments and schooling, which – together with our model – may suggest that high thresholds bind for a number of problems of interest to economists. However, in contexts in which agents are learning about concepts that are less noisy than returns – say the availability of microfinance, how to enroll in welfare, or whether a firm is hiring – simple contagion may be the right model. Characterizing which productive investments should diffuse easily through social networks – and which need extensive and targeted diffusion – is crucial but beyond the scope of this paper.

References

Acemoglu, Daron, Ozdaglar, Asuman Ozdaglar, and Ercan Yildiz. 2011. "Diffusion of Innovations in Social Networks." IEEE Conference on Decision and Control (CDC). doi: 10.1109/CDC.2011.6160999.

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi. 2020. "Just a Few Seeds More: Value of Targeting for Diffusion in Networks." *Mimeo*, Stanford Unviersity.

Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas, Jr. 2013. "Idea Flows, Economic Growth, and Trade." National Bureau of Economic Research. doi: 10.3386/w19667.

Anderson, Jock R., and Gershon Feder. 2007. "Agricultural Extension." In *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. doi: 10.1111/j.1468-0297.2006.01115.x.

Banerjee, Abhijit, Emily Breza, Arun G. Chandrasekhar, and Benjamin Golub. 2020. "When Less is More: Experimental Evidence on Information Delivery during India's Demonetization." NBER Working Paper No. 24679.

Banerjee, Abhijit, Emily Breza, Arun G. Chandrasekhar, and Markus Mobius. 2016. "Naive Learning with Uniformed Agents." *Mimeo*, Stanford Unviersity. doi: 10.3386/w25497.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2013. "The Diffusion of Microfinance." *Science,* 341 (6144): 1236498. doi: 10.1126/science.1236498.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson. 2019. "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials." *Review of Economic Studies,* 86 (6): 2453-2490. doi: 10.1093/restud/rdz008.

Beaman, Lori. 2012. "Social Networks and the Dynamics of Labor Market Outcomes: Evidence from Refugees Resettled in the U.S." *Review of Economic Studies,* 79 (1): 128-161.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak. 2018. "Can Network Theory-based Targeting Increase Technology Adoption?" AEA RCT Registry. August 06. doi: 10.1257/rct.2017-1.0.

Beaman, Lori, Ariel BenYishay, Jeremy Magruder, and Ahmed Mushfiq Mobarak. 2021. "Replication data for: Can Network Theory-based Targeting Increase Technology Adoption?" American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. openicpsr-130605.

BenYishay, Ariel, and Ahmed Mushfiq Mobarak. 2019. "Social Learning and Incentives for Experimentation and Communication." *Review of Economic Studies*, 86 (3): 976-1009. doi: 10.1093/restud/rdy039.

BenYishay, Ariel, Maria Jones, Florence Kondylis, and Ahmed Mushfiq Mobarak. 2020. "Gender Gaps in Technology Diffusion." *Journal of Development Economics*, 143 (2020): 102380, 1-27. doi: 10.1016/j.jdeveco.2019.102380.

Bennett, Magdalena, and Peter Bergman. "Better Together? Social Networks in Truancy and the Targeting of Treatment." *Journal of Labor Economics* 39 (1): 1-36. doi: 10.1086/708672

Bjorkegren, Daniel. 2018. "The Adoption of Newtork Goods: Evidence from the Spread of Mobile Phones in Rwanda" *The Review of Economic Studies*, 86 (3): 1033-60. doi: 10.1093/restud/rdy024

Blumenstock, Joshua, Guanghua Chi, and Xu Tan. 2019. "Migration and the Value of Social Networks." *Mimeo,* UC-Berkeley.

Breza, Emily, Arun G. Chandrasekhar, Tyler H. McCormick, and Mengjie Pan. 2020. "Using Aggregated Relational Data to Feasibly Identify Network Structure Without Network Data." *American Economic Review,* 110 (8): 2454-2484.

Burlig, Fiona, and Andrew Stephens. 2019. "Reap What Your Friends Sow: Social Networks and Technology Adoption." *Mimeo*, U. of Chicago.

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. doi: 10.3982/ECTA11991.

Carrel, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogeous Peer Group Formation. *Econometrica*, 81 (3): 855-882. doi: 10.3982/ECTA10168.

Centola, Damon. 2010. "The Spread of Behavior in an Online Social Network Experiment." *Science,* 329 (5996): 1194-1197. doi: 10.1126/science.1185231

Centola, Damon, and Michael Macy. 2007. "Complex Contagions and the Weakness of Long Ties." *American Journal of Sociology*, 113: 702-34.

Chandrasekhar, Arun G., Benjamin Golub, and He Yang. 2019. "Signaling, Shame, and Silence in Social Learning." *Mimeo,* Stanford University.

Chandrasekhar, Arun G., Horacio Larreguy, and Juan Pablo Xandri. 2020. "Testing Models of Learning on Social Networks: Evidence from a Framed Field Experiment." *Econometrica*, 88 (1): 1-32. doi: 10.3982/ECTA14407.

Chandrasekhar, Arun G., and Randall Lewis. 2016. "Econometrics of Sampled Networks." *Mimeo,* Stanford University.

Christakis, Nicholas A., and James H. Fowler. 2010. "Social Network Sensors for Early Detection of Contagious Outbreaks." *PloS One,* 5 (9): e12948. doi: 10.1371/journal.pone.0012948.

Coleman, James, Elihu Katz, and Herbert Menzel. 1957. "The Diffusion of an Innovation Among Physicians." *Sociometry,* 20 (4): 253-270. https://www.jstor.org/stable/2785979.

Conley, Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology." *American Economic Review,* 100 (1): 35-69. doi: 10.1257/aer.100.1.35.

DeMarzo, Peter, Dimitri Vayanos, and Jeffrey Zwiebel. 2003. "Persuasion Bias, Social Influence, and Unidimensional Opinions." *Quarterly Journal of Economics*, 118 (3): 909-968.

Derpsch, Rolf. 2003. "Conservation Tillage, No-tillage and Related Technologies." In: García-Torres L., Benites J., Martínez-Vilela A., Holgado-Cabrera A. (eds) *Conservation Agriculture*. Springer, Dordrecht. doi: 10.1007/978-94-017-1143-2_23.

Derpsch, Rolf. 2004. History of Crop Production, With and Without Tillage. *Leading Edge,* 150-154.

Flodgren, Gerd, Elena Parmelli, Gaby Doumit, Melina Gattellari, Jeremy Grimshaw, and Mary Ann O'Brien. 2007. "Local Opinion Leaders: Effects on Professional Practice and Health Care Outcomes." *Cochrane Database Systematic Review,* 24 (1), CD000125. doi: 10.1002/14651858.cd000125.pub3.

Duflo, Esther, and Emmanuel Saez. 2003. "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics,* 118 (3): 815-842. doi: 10.1162/00335530360698432.

Feld, Scott L. 1991. "Why Your Friends Have More Friends Than You Do." *American Journal of Sociology,* 96 (6): 1464-1477. https://www.jstor.org/stable/2781907.

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.

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology*, 6: 1420-1443.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technical Change." *Econometrica,* 25 (4): 501-522. https://www.jstor.org/stable/1905380.

Haggblade, Steven, and Gelson Tembo. 2003. "Conservation Farming in Zambia." EPTD Discussion Paper No. 108, International Food Policy Research Institute.

Islam, Asad, Philip Ushchev, Yves Zenou, and Xin Zhang. 2019. "The Value of Information in Technology Adoption: Theory and Evidence from Bangladesh." *Mimeo*, Monash University.

Jack, B. Kelsey. 2011. "Market Inefficiencies and the Adoption of Agricultural Technologies in Developing Countries." *Mimeo*, University of California, Santa Barbara.

Jackson, Matthew, ed. 2008. *Social and Economic Networks.* Princeton, New Jersey: Princeton University Press.

Jackson, Matthew, and Evan C. Storms. 2019. "Behavioral Communities and the Atomic Structure of Networks." *Mimeo*. https://ssrn.com/abstract=3049748.

Kim, David A., Allison 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. doi: 10.1016/S0140-6736(15)60095-2.

Kremer, Michael, Edward Miguel, Sendhil Mullainathan, Clair Null, and Alix P. Zwane. 2011. "Social Engineering: Evidence from a Suite of Take-up Experiments in Kenya." *Mimeo*, UC Berkeley.

Magruder, Jeremy. 2010. "Intergenerational Networks, Unemployment, and Inequality in South Africa." *American Economic Journal: Applied Economics,* 2 (1): 62-85.

Miller, Grant, and Ahmen Mushfiq Mobarak. 2015. "Learning About New Technologies Through Social Networks: Experimental Evidence on Nontradtional Stoves in Bangladesh." *Marketing Science,* 34 (4): 480-499.

Monsted, Bjarke, Piotr Sapiezynski, Emilio Ferrara, and Sune Lehmann. 2017. "Evidence of Complex Contagion of Information in Social Media: An Experiment Using Twitter Bots." *PloS ONE,* 12 (9): e0184148. doi: 10.1371/journal.pone.0184148.

Munshi, Kaivan. 2004. "Social Learning in a Heterogeneous Population: Technology Diffusion in the Indian Green Revolution." *Journal of Development Economics,* 73 (1), 185-215. doi: 10.1016/j.jdeveco.2003.03.003.

Munshi, Kaivan. 2007. "Information Networks in Dynamic Agrarian Economies." *Handbook of Development Economics,* 4 (48), 3085-3113. doi: 10.1016/S1573-4471(07)04048-X.

Munshi, Kaivan. 2008. "Social Learning and Development." In L. E. Blume, and S. N. Durlauf, *New Palgrave Dictionary of Economics.* Palgrave Macmillan.

Oster, Emily, and Rebecca Thornton. 2012. "Determinants of Technology Adoption: Private Value and Peer Effects in Menstrual Cup Take-Up." *Journal of the European Economic Association,* 10 (6), 1263-1293.

Perla, Jesse, and Christopher Tonetti. 2014. "Equilibrium Imitation and Growth." *Journal of Political Economy,* 122 (1), 52-76.

Qiao, Fangbin. 2015. "Fifteen Years of Bt Cotton in China: The Economic Impact and Its Dynamics." *World Development*, 70: 177-185. doi: 10.1016/j.worlddev.2015.01.011.

Rogers, Everett M. 1962. *Diffusion of Innovations.* New York, NY: The Free Press.

Rosenzweig, Mark R. and Christopher Udry. 2020. "External Validity in a Stochastic World: Evidence from Low-Income Countries." The Review of Economic Studies 87(1): 343-81. https://doi.org/10.1093/restud/rdz021

Ryan, Bryce, and Neal C. Gross. 1943. "The Diffusion of Hybrid Seed in Two Rural Iowa Communities." *Rural Sociology,* 8 (1), 15-24.

Udry, Christopher. 2010. The Economics of Agriculture in Africa: Notes Toward a Research Program. *African Journal of Agricultural and Resource Economics, 5*(1): 284-299. doi: 10.22004/ag.econ.156665

World Bank. 2008. World Development Report 2008: Agriculture for Development. Washington, DC: The World Bank.

Figure 2: % of villages where at least some non-seeds adopted in data and simulations

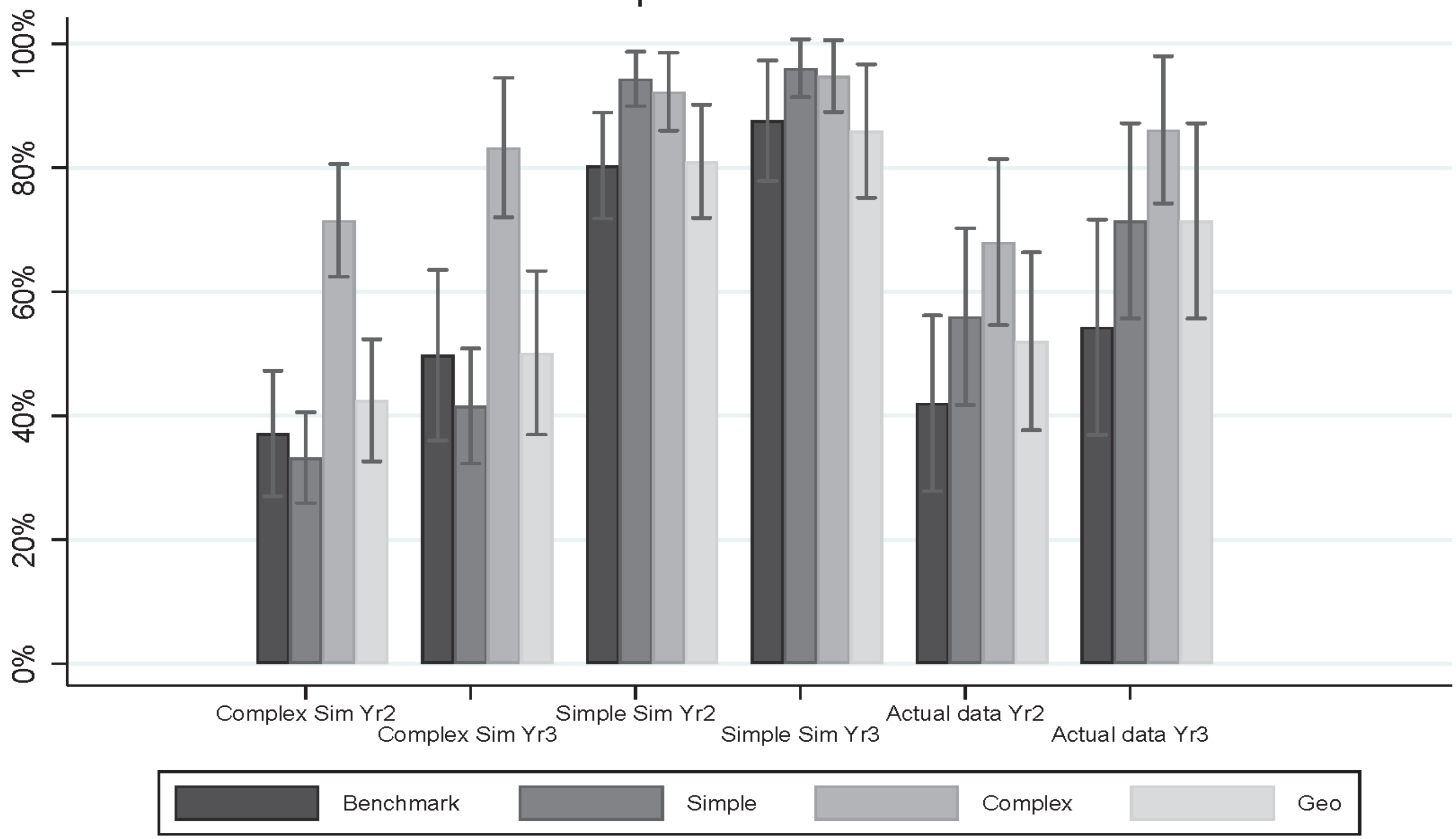

Table 1: Centrality of partner farmers across treatments

	Eigenvector Centrality		Degree	
	Rank 1 Partner	Rank 2 Partner	Rank 1 Partner	Rank 2 Partner
	(1)	(2)	(3)	(4)
Treatment arm:				
Complex diffusion	0.28	0.19	17.49	13.39
Simple diffusion	0.27	0.07	16.59	6.70
Geo	0.15	0.10	9.48	6.34
Benchmark	0.21	0.13	13.29	9.80

1 The sample includes all partner farmers, including seeds and shadows. However, benchmark partners are restricted to only seed farmers (and hence the sample size is smaller) because Benchmark shadow farmers are not observed in Complex, Simple or Geo villages.

Table 2: Village-Level Regressions of Adoption Outcomes Across Treatment Arms

	Any Non-Seed Adopters		Adoption Rate	
	(1)	(2)	(3)	(4)
Complex Diffusion Treatment	0.252	0.304	0.036	0.036
	(0.093)	(0.101)	(0.016)	(0.026)
Simple Diffusion Treatment	0.155	0.189	0.036	0.006
	(0.100)	(0.111)	(0.017)	(0.022)
Geographic treatment	0.107	0.188	0.038	0.013
	(0.096)	(0.110)	(0.027)	(0.034)
Year	2	3	2	3
N	200	141	200	141
Mean of Benchmark Treatment (omitted category)	0.420	0.543	0.038	0.075
SD of Benchmark	0.499	0.505	0.073	0.109
p-values for equality in coefficients:				
Simple = Complex	0.300	0.240	0.981	0.173
Complex = Geo	0.102	0.220	0.937	0.491
Simple = Geo	0.623	0.990	0.950	0.783

Notes

1 The reference group is the Benchmark treatment.

2 The "Any non-seed adopters" indicator in columns (1)-(2) excludes seed farmers. The adoption rate in columns (3)-(4) include all randomly sampled farmers, excluding seed and shadow farmers.

3 Sample for year 3 (columns 2 and 4) excludes Nkhotakota district.

4 All columns include controls used in the re-randomization routine (percent of village using compost at baseline; percent village using fertilizer at baseline; percent of village using pit planting at baseline); village size and its square; and district fixed effects. Standard errors are clustered at the village level.

Table 3: Conversations Farmers Report Having about Pit Planting with Seed and Shadow Partners

	Conversation about pit planting		
	(1)	(2)	(3)
Trained	0.037	0.050	0.064
	(0.008)	(0.008)	(0.009)
% convo with trained seed	0.179	0.181	0.190
% convo with shadow partner	0.141	0.130	0.127
N	15,115	16,704	11,607
Year	1	2	3

Notes

1 Sample excludes seeds and counterfactual / shadow farmers.

2 In our survey, we asked respondents about conversations they had with the seed farmers and randomly selected counterfactual/shadow farmers. In this table we refer to farmers who would be seeds under the different treatments as partners, whether they are trained seeds or are shadow farmers. An observation is a respondent-partner-year pair.

3 The following indicator variables are also included: whether the contact that the respondent was asked about was a simple partner, complex partner or geo partner, irrespective of whether they were trained. All columns include controls used in the re-randomization routine (percent of village using compost at baseline; percent village using fertilizer at baseline; percent of village using pit planting at baseline); village size and its square; and district fixed effects.

4 Standard errors are clustered at the village level.

Table 4: Heterogeneity in Farmer-Level Adoption Decisions Across Treatment Arms

	(1)	(2)	(3)	(4)
Bad Signa l* Complex	0.006	-0.027	0.013	-0.045
	(0.024)	(0.036)	(0.015)	(0.033)
Bad Signal * Simple	-0.008	-0.036	0.019	-0.008
	(0.024)	(0.037)	(0.017)	(0.034)
Bad Signal * Geo	0.002	-0.068	0.031	-0.054
	(0.031)	(0.031)	(0.035)	(0.032)
Good Signal	-0.037	-0.062	-0.007	-0.064
	(0.017)	(0.024)	(0.022)	(0.038)
Good Signal * Complex	0.059	0.067	0.054	0.083
	(0.018)	(0.025)	(0.024)	(0.030)
Good Signal * Simple	0.064	0.029	0.054	0.021
	(0.021)	(0.020)	(0.029)	(0.020)
Good Signal * Geo	0.042	0.022	0.026	0.031
	(0.020)	(0.023)	(0.022)	(0.029)
Good Signal Type	Flat Land	Flat Land	Unfamilliar Tech	Unfamilliar Tech
Year	2	3	2	3
N	3546	2645	3954	3023
Mean of Bad Signal in Benchmark Treatment (omitted	0.066	0.123	0.046	0.104
SD	0.248	0.33	0.21	0.305
p-values for equality in coefficients:				
Simple, Good = Complex, Good	0.828	0.113	0.986	0.032
Complex, Good = Geo, Good	0.482	0.103	0.297	0.138
Simple, Good = Geo, Good	0.364	0.755	0.351	0.680

Notes

1 The reference group is Bad Signal recipients in the Benchmark treatment.

2 In columns (1)-(2), households with any flat land are those who have Good Signal=1 and those with all sloped land have Good Signal=0. In columns (3)-(4), households in villages where less than 4.32% (the median) of households ever tried pit planting at baseline are those who have Good Signal=1.

3 Sample for year 3 (columns 2 and 4) excludes Nkhotakota district.

4 All columns include controls used in the re-randomization routine (percent of village using compost at baseline; percent village using fertilizer at baseline; percent of village using pit planting at baseline); village size and its square; and district fixed effects. Standard errors are clustered at the village level.

Table 5: Diffusion within the Village

	Heard of PP			Knows how to PP			Adopts PP		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Connected to 1 seed	0.002	0.030	0.016	0.017	0.021	-0.031	0.008	0.012	0.004
	(0.024)	(0.022)	(0.029)	(0.016)	(0.017)	(0.023)	(0.011)	(0.015)	(0.017)
Connected to 2 seeds	0.084	0.124	0.064	0.062	0.068	0.110	0.016	0.039	0.014
	(0.038)	(0.040)	(0.064)	(0.028)	(0.029)	(0.051)	(0.014)	(0.019)	(0.035)
Within path length 2 of at least one seed	-0.018	0.016	0.067	0.005	0.022	0.028	0.013	0.022	0.037
	(0.028)	(0.027)	(0.042)	(0.018)	(0.021)	(0.028)	(0.008)	(0.013)	(0.021)
Year	1	2	3	1	2	3	1	2	3
N	4155	4532	3103	4155	4532	3103	4203	3931	2998
Mean of Reference Group (No connection to any seed)	0.223	0.286	0.391	0.057	0.095	0.147	0.013	0.044	0.043
SD of Reference Group	0.416	0.452	0.488	0.232	0.293	0.355	0.113	0.206	0.203
p-value for 2 connections = 1 connection	0.018	0.013	0.442	0.072	0.091	0.004	0.522	0.164	0.760

Notes

1 Sample excludes seed and shadow farmers. Only connections to simple, complex and geo seed farmers are considered (no connections to benchmark farmers included).

2 The dependent variable in columns (1)-(3) is an indicator for whether the respondent reported being aware of a plot preparation method other than ridging and then subsequently indicated awareness of pit planting in particular. In columns (4)-(6), the dependent variable is an indicator for whether the farmer reported knowing how to implement pit planting. The dependent variable in (7)-(9) is an indicator for the household having adopted pit planting in that year.

3 In all columns, additional controls include indicators for the respondent being connected to: one Simple partner, two Simple partners, one Complex partner, two Complex partners, one Geo partner, two Geo partners, within 2 path length of a Simple partner, within 2 path length of a Complex Partner, and within 2 path length of the geo partner. Also included are village fixed effects. Standard errors are clustered at the village level.

4 The reference group is comprised of individuals with no direct or 2-path-length connections to a seed farmer.

Online Appendix

Can Network Theory-based Targeting Increase Technology Adoption?

Lori Beaman, Ariel BenYishay, Jeremy Magruder, Ahmed Mushfiq Mobarak

A.1. Technical Details of Simulation Methods

To identify our seed partners, we used the social network census of households in all study villages. The social network structures observed in these data allow us to construct network adjacency matrices for each of the 200 villages. Next, we conduct technology diffusion simulations for all villages using these matrices, where each individual in the village draws an adoption threshold τ from the data, which is normally distributed[26] $N(\lambda, 0.5)$ but truncated to be strictly positive. We conduct simulations with $\lambda=1$ and $\lambda=2$ in all villages to evaluate simple and complex contagion respectively.

In the simulations, when an individual is connected to at least τ individuals who are informed, she becomes informed in the next period. Once an individual is informed, we assume that all other household members are immediately also informed. We also assume that becoming informed is an absorbing state. As seed farmers are trained by extension agents, we assume all assigned seed farmers become informed.

We run the model for four periods.[27] Given the randomness built into the model, we simulate the model 2000 times for each potential pair of seeds in the village, and create a measure of the average information rate induced by each pair. We designate the pair that yields the highest average three-period information rate in our simulations as the two "*optimal seeds*" for each village for that particular model (simple contagion, $\lambda=1$ or complex contagion, $\lambda=2$). Armed with the identities of the optimal

[26] Heterogeneity in the model comes from variation across individuals in the net benefits realized by adopting pit planting. This affects the threshold number of connections an individual would need to have in order to get enough signals to be induced to adopt.

[27] We collected data for up to three agricultural seasons ("years") after the interventions were implemented, so our theoretical set-up largely matches our empirical research design. With knowledge of the value of λ, a policymaker could use the model to maximize adoption over any timeframe they cared about, either more short-term or more long-term.

seeds under each model, we then randomly assign different villages in the sample to the treatment arms. Optimal seeds identified through the complex contagion (λ=2) simulation are trained in villages that were randomly assigned to treatment 1. Optimal seeds identified through the simple contagion (λ=1) simulation are trained in the randomly chosen villages assigned to treatment 2.

To determine seeds for villages in the Geo treatment arm, the simulation steps are the same as in the complex contagion case, except that we apply the procedure to a different adjacency matrix. To capture the idea that geography may be an easy way to capture key features of a social network, we generate an alternative adjacency matrix by making the assumption that two individuals are connected if their plots are located within 0.05 miles of each other in our geo-coded location data. We chose a radius of 0.05 miles because this characterization produces similar values for network degree measures in our villages as using the actual network connections measures.

A.2. Effect of technology adoption on crop yields

In order to estimate the returns of adopting the new technologies on yields, we compare seed farmers to shadow farmers. Online Appendix Table A4 demonstrates that there were large differences in adoption rates between seeds and shadow farmers. To estimate the impact of adoption on yields, we estimate an ITT specification exploiting that random difference in take-up:

$$y_{ivt} = \beta Seed_{ivt} + \gamma X_v + \delta_t + \epsilon_{ivt} \qquad (1)$$

where y_{ivt} is log maize yields for farmer *i* in village *v* at time *t*, $Seed_{ivt}$ is an indicator for being the selected seed farmer, X_v are control variables used during the re-randomization routine (see notes in Table 2), village size, village size squared, district fixed effects plus baseline land size. δ_t are year dummies. We use data from years 2 and 3. In the intent-to-treat specification in Online Appendix Table A1, column (1), maize yields among seed farmers are 13% greater than the yields experienced by the shadow seeds. The fact that the technologies we promoted led to an increase in output strongly suggests that the information about pit planting that diffused through the networks was likely positive on average.

Since only about 30% of seeds adopted pit planting, we also report the local average treatment effect using an IV regression in column (2) in which we instrument pit planting adoption with an indicator for being randomly assigned as the seed (rather than a shadow). In this specification, pit planting adoption is associated with a 44% increase in maize yield. However, we cannot rule out that CRM adoption also increased yields, potentially violating the exclusion restriction in the IV estimation.[28]

A.3. Adoption rates among seeds (compared to shadow farmers)

Online Appendix Table A4 compares the technology adoption behavior of seed farmers to shadow farmers. We focus on this sub-sample because shadow farmers act as the correct experimental counter-factual for the seed farmers to capture the causal effect of the intervention, removing any bias due to the seeds' position within their networks. We estimate the following equation, and Panel A displays the results:

$$y_{ivt} = \beta Seed_{ivt} + \delta_v + \epsilon_{ivt} \qquad (1)$$

where the dependent variable is an indicator for adoption, and δ_v are village fixed effects. Column (1) shows that trained seeds are 52% more likely in year 1 to know how to pit plant than shadow farmers. Columns (4)-(6) show that seed farmers who are trained on pit planting adopt at a rate of 31-32% in all three years, compared to the low 5% adoption rate of shadow farmers in year 1.

[28] We also cannot rule out any labor or other input use response to training which may have positively contributed to yields. Changes in other inputs makes it impossible for us to say that the yields increases map directly into increases in profits.

Panel B of Online Appendix Table A4 restricts the sample to only seed farmers (and drops all shadow farmers) and compares knowledge and adoption among seeds across the four experimental arms as follows:

$$y_{ivt} = \beta_0 + \beta_1 Simple_v + \beta_2 Complex_v + \beta_3 Geo_v + \delta X_v + \epsilon_{ivt} \qquad (2)$$

where X_v include the re-randomization controls (listed in table notes), village size, the square of village size, and district fixed effects. Standard errors are clustered at the village level. Column (1) shows that in the first year, Benchmark seeds are most likely to say they know how to pit plant, while all other seeds are similar. The extension agents evidently chose seed farmers carefully to ensure that their chosen extension partners receive the initial training from them. However, in years 2 and 3, familiarity between Benchmark, Simple and Complex seeds converge and have similar levels of familiarity with pit planting, though knowledge is declining over time. Geo seeds continue to display lower familiarity in subsequent years.

Column (4) shows that there are no differences in adoption propensities across the four types of seeds in the first year. This implies that it is unlikely that any observed differences in village-wide adoption patterns across the four treatment arms, that we will examine later, are driven by initial adoption differences inside the sub-sample of seed farmers. Columns (5) and (6) show that seed farmers in simple contagion villages become relatively more likely over time to adopt the technology. This could be due to the technology diffusion process, or in other words, a consequence of the experiment. Columns (7)-(8) show that there are no significant differences in adoption in years 1 or 2 for crop residue management.

A.4. Conversation frequency and adoption cascades

AMS establish that random seeding is sufficient to generate an adoption cascade when $CD > 1$, where C refers to the probability that a conversation takes place on a given link and D represents the mean degree in the network. Our experimental evidence found that at least 5% of randomly selected respondents were having conversations with seeds due to the training in each year. To map this number to the CD framework, we first suppose that mean degree (the average number of contacts that a person has) is stable over time, so that the mean degree of trained seed partners is the same at the follow-up as in our listing (indeed, in results available from the authors, we demonstrate that whether a respondent reports knowing a seed or shadow farmer at follow-up is the same regardless of whether the seed was actually trained or not).

In our data, the mean village has 77 respondents (households), 2 of whom are seeds. Thus, when we document that training induced at least 5% of respondents to have conversations about pit planting with seeds, we establish that at least 3.75 households per village had a conversation with a seed farmer (3.75 = 0.05*75). Based on Table 1, the mean degree of seeds is 11.63; thus, we expect that seeds have a conversation with 32% of their connections. In other words, the 5% lower bound on conversations about pit planting suggests that $C \approx 0.32$.

Mean degree among farmer households in our study villages is about 7. Thus, in our data $CD > 7 * 0.32 = 2.24$, where the greater than inequality is due to the fact that the 5% of experimentally exogenous conversations is a very restrictive lower bound. In other words, using this bounding exercise, we are confident that $CD > 1$ and so adoption cascades should take place with random seeding.

A.5. Micro-foundation of threshold model

We develop this micro-foundation by extending a framework presented in Banerjee et al. (2016) (hence: BBCM). One key insight in BBCM is that the majority of members of a social network may not have access to *any* useful signal when they are confronted with an entirely new technology. Thus, there are two parts to the learning problem for new technologies: acquiring a signal in the first place (becoming informed) which may be costly, and forming a revised belief on the profitability of the new technology based on the signals received from informed connections. Optimizing farmers adopt a new technology only if their beliefs change, and they are convinced by others that this would be more profitable than alternatives.[29]

There are three key phases of decision-making in our model: (1) the farmer has to decide whether to acquire information[30], (2) she has to combine the new information with her priors, and (3) she then decides whether to adopt the new technology. We will present and solve the model backwards, starting with the third phase.

The farmer will choose to adopt the new technology in phase 3 if she believes that adoption will be profitable. Suppose farmer *j* knows the technology will cost her c_j to adopt and believes the new technology has either profit $\bar{\pi}$ or $\underline{\pi}$ $(\underline{\pi} < c_j < \bar{\pi})$.[31] Since the technology is new and farmer *j* is initially uninformed, she has a uniform prior as to whether the technology is profitable or not. She can aggregate signals given by her connections to update her prior and make an informed adoption decision.

We adopt the same learning environment modeled in BBCM: first, informed farmer *i* disseminates a binary signal, $x_i \in \{\underline{\pi}, \bar{\pi}\}$, which is accurate with probability $\alpha > \frac{1}{2}$. Uninformed farmers do not disseminate a signal. Second, farmers follow DeGroot learning (DeMarzo et al. 2003).

---

[29] A very different micro-foundation for a similar model is explored in Jackson and Storms (2019). In that model, thresholds become relevant as individuals face greater payoffs from conforming to the behavior of their connections. Since coordination incentives for smallholder adoption of new agricultural technologies seem likely to be low, we pursue instead a model based on learning and individual optimization.

[30] There is a growing literature on how agents decide whether to seek out information. Banerjee et al. (2019a) – which builds on theoretical work by Chandrasekhar, Golub and Yang (2019) – demonstrate in the context of India's demonetization that some agents choose to remain uninformed in order to avoid shame. BenYishay et al. (2020) show that agents may choose not to receive agricultural information if the sender is a woman.

[31] Here for simplicity we follow BBCM in assuming that the distribution of profits is binary and known. In practice, there will be uncertainty over a wider range of profits due to the potential performance of the technology under different agroclimatic conditions and different weather realizations. While posterior distributions will be much more complicated under more realistic depictions of uncertainty, the key intuition driving the threshold model will be unchanged.

DeGroot learning can be interpreted as a boundedly rational version of Bayes learning, and suggests that farmers aggregate signals from their connections without attempting to calculate the inherent correlation structure between those signals. That is, if farmer *j* sees a signal of $\bar{\pi}$ from both farmers *i* and *k*, she interprets that as two positive signals without decomposing the likelihood that farmer *i* and *k* are disseminating information obtained from the same source.[32] Once farmers have observed signals from their informed connections, they aggregate those signals via Bayes' rule.

This framework suggests the following for the second phase of the farmer's learning problem: suppose farmer *j* has D_j informed contacts. If farmer *j* decides to learn about the new technology from her informed contacts, and if *H* of those contacts provide the signal $x = \bar{\pi}$, then the farmer's posterior probability that $\pi = \bar{\pi}$ is given by[33]

$$E_j[\pi = \bar{\pi}] = \frac{\alpha^{2H-D_j}}{\alpha^{2H-D_j} + (1-\alpha)^{2H-D_j}}$$

Denote $\tilde{\bar{\pi}} = \bar{\pi} - \underline{\pi}$ and $\tilde{c_j} = c_j - \underline{\pi}$. With that posterior, the farmer would adopt the technology if

$$\frac{\tilde{c}_j}{\tilde{\bar{\pi}}} \leq \frac{\alpha^{2H-D_j}}{\alpha^{2H-D_j}+(1-\alpha)^{2H-D_j}} \leq \frac{\alpha^{D_j}}{\alpha^{D_j}+(1-\alpha)^{D_j}} \tag{1}$$

This model highlights a potential challenge to diffusing new technologies: when few other farmers are informed, then there is a ceiling on how much a new farmer's priors would move even if they receive unanimously positive signals from the informed. At early stages in the diffusion process, D_j may be small for most farmers.

Last, we consider the first phase of the farmer's learning problem, which is her decision to acquire signals and become informed. Here, we depart from BBCM to suggest that there may be a

[32] Chandrasekhar, Larreguy and Xandri (2020) provide laboratory evidence in support of DeGroot learning over Bayes learning in India. Additional citations in favor of this boundedly-rational approximation can be found in BBCM.

[33] A simple proof is given in BBCM.

small cost to receiving a signal η. This cost could be interpreted as "shoe leather" costs of acquiring information (which are not necessarily trivial in villages in rural Malawi as households may be fairly far apart), or as stigma from seeking information (e.g. Banerjee et al. 2019a).

Thus, the farmer *j* with informed degree D_j has an objective given by

$$\max_{d\le D_j}\sum_{h\le d}\frac{1}{2}\left(\left[\alpha^h(1-\alpha)^{d-h}\left(\overline{\pi}-c_j\right)+(1-\alpha)^h\alpha^{d-h}\left(\underline{\pi}-c_j\right)\right]\left(I\left(\frac{\alpha^{2h-d}}{\alpha^{2h-d}+\left(1-\alpha^{2h-d}\right)}>\frac{\tilde{c}}{\tilde{\tilde{\pi}}}\right)\right)\right)-\eta d$$

When $\eta = 0$, the dynamics of learning are explored by BBCM. However, when $\eta > 0$ the dynamics are slightly different. In that case (for small η), farmers will only become informed if

$$\frac{\alpha^{D_j}}{\alpha^{D_j}+(1-\alpha)^{D_j}} > \frac{\tilde{c}_j}{\tilde{\tilde{\pi}}} \tag{2}$$

In other words, farmers only choose to seek information if they have a large enough number of informed connections, such that it is possible that an informed decision would lead them to adopt. In this case (and for small η), farmers will choose to seek information when they have only one informed connection if

$$\frac{\alpha}{\alpha+(1-\alpha)} > \frac{\tilde{c}_j}{\tilde{\tilde{\pi}}} \tag{3}$$

In general, they will choose to become informed with λ informed connections if

$$\frac{\alpha^{\lambda}}{\alpha^{\lambda}+(1-\alpha)^{\lambda}} > \frac{\tilde{c}_j}{\tilde{\tilde{\pi}}} \tag{4}$$

This implies that farmers choose to become informed about new technologies if expectations about the net benefits of technology are high (i.e., low costs and high potential gains), or if signals from individual other farmers are highly accurate. Under certain parameter values, just a single informed contact may be sufficient to induce farmers to seek information. That is the diffusion process that Centola and Macy (2007) refer to as a "simple contagion." They demonstrate that some types of information – for example, job opportunities – spread in this way. On the other hand, if the expected upside of the technology is more modest relative to costs, or if signals from other farmers have low accuracy, then farmers may only be persuaded to seek information when there is sufficient information to be gained from their network.[34] In that case, for many farmers the lowest λ satisfying equation (4) may be larger than 1, and information diffusion follows a process termed "complex contagion" in the literature.[35]

Our interpretation of the microeconomics of the threshold theory is that the thresholds result from an underlying process of farmers deciding *whether* to learn, given their information environment. This motivates an experimental design in which we seed new information in a network to improve the overall information environment, which can change incentives to learn and jump-start the technology diffusion process.

Given that the econometrician is unlikely to observe signal accuracy (α), the threshold required for adoption of a specific new technology is an empirical question. As a numerical example, consider a technology with 30% potential returns (so that $\tilde{\tilde{\pi}} = 1.3\ \tilde{c_J}$). If signals are more than 77% accurate, farmers will choose to become informed if they have a single informed connection, and diffusion will follow a simple contagion. If signal accuracy falls in the range of 65-77% accurate, then farmers will only become informed if they have 2 informed connections, and learning will follow a complex contagion. If signals are less than 65% accurate, then farmers will need at least 3 informed connections to make an adoption decision. In general, agents will face higher thresholds in contexts

34 Though not explicitly considered here, minimal thresholds for learning will also be higher if η (the cost of information acquisition) is larger.

35 Several theory papers have explored the implications of this model. In contrast to the "strength of weak ties" in labor markets proposed by Granovetter (1978), strong ties may be important for the diffusion of behaviors that require reinforcement from multiple peers. Centola (2010) provides experimental evidence that health behaviors diffuse more quickly through networks where links are clustered, consistent with complex contagion. Acemoglu et al. (2011) highlights that when contagion is complex, highly clustered communities will need a seed placed in the community in order to induce adoption. Finally, Monsted et al. (2017) provide experimental evidence generated by twitter-bots that twitter hashtag retweets follow a process which more closely resembles complex than simple contagion.

where signals are noisier, a point with implications for external validity which we return to in the concluding remarks.

Model predictions and implications for the experiment

The micro-foundation of the threshold model suggests that the model would need to be tested using the diffusion of a truly new technology, where would-be adopters are *ex ante* uninformed about the technology and face an important adoption decision. A corollary is that the threshold model should fit the data better in locations where the technology is more novel. A good empirical setting to test the model is also one in which agents are receiving noisy signals from the network.

If thresholds exist and are above one, then seeding the network with multiple sources of information who are clustered in the same part of the network will achieve very different diffusion patterns than seeding the network with the same number of information sources spread more diffusely. Our experimental design will take advantage of this insight. When thresholds are above one, the information environment only induces learning when initial nodes share some connections, which we test using micro data on technology diffusion patterns.

The model highlights that farmers will become informed when they have sufficiently many informed contacts. However, conditional on being informed, they will only adopt the technology if the realization of signals from their connections are sufficiently positive. These two facts suggest two different tests of the model.

PREDICTION 1: If most farmers in a village have a threshold $\bar{\lambda}$, then people who are connected to at least $\bar{\lambda}$ informed farmers should become informed themselves.

PREDICTION 2: Adoption should increase most strongly among farmers who have high net benefits of adoption, who would adopt with a broader range of received signals.[36]

[36] For clarity, the model assumed that the potential net benefits of production were known to the farmer before deciding whether to become informed about the technology. In practice, farmers may or may not be aware that their private net

A.6. Simulation of cost-effective targeting strategies

For our simulations, we suppose that our extension agent starts with a random sample of candidate respondents, and is able to screen out individuals with less than 2 connections. We suppose the extension agent starts with a list of 2-10 randomly selected farmers.

Starting from that random sample of farmers, we solicit each farmer's connections and calculate each random farmer's degree. We then focus on 6 candidate targeting strategies:

A. Trains two randomly selected people from that list (used as a benchmark)
B. Trains the two highest degree people from that list
C. Select two random friends of the highest degree person from that list
D. Trains the two highest degree connections of the highest degree farmer from the random sample (requires interviewing all connections of the highest degree respondent to determine their degree)
E. Selects two farmers from that list at random; interviews two of their connections (selected at random) and trains two of the connections' connections[37]
F. Trains the highest degree respondent and one of his connections (at random).

For each of these five candidate strategies, we simulate adoption rates after 4 rounds of simulations against the seeds chosen by our Complex treatment. We find that Strategy A, selecting two farmers at random, achieves 57% of the adoption produced by the Complex treatment. We can then view the other targeting strategies in terms of their performance above the random benchmark. Strategy B is identical to random selection with only 2 initial interviews, and so similarly generates

benefits to adoption are high before becoming informed. Only when a farmer is *ex ante* aware that she has relatively high net benefits will we see greater adoption associated with a greater propensity to become informed.

[37] This "friends of friends" approach to identifying central people was inspired by Feld (1991), Christakis and Fowler (2010), and Kim et al. (2015), who note that randomly selected connections tend to be more central than randomly selected nodes in a network. We again assume that the extension agent is able to screen out potential trainees with less than two total connections.

57% adoption; however, as the extension agent interviews more people to identify these high degree individuals it performs somewhat better, achieving 70% of the complex contagion adoption with 10 total interviews. Strategies C and D both leverage the highest degree respondent from the initial random sample. These perform the best out of the strategies we consider. Strategy C achieves 73% of the optimized adoption with just two total interviews, which increases modestly to 76% of the optimized adoption as the number of interviews grows to 10 to better identify a high degree individual. Strategy D, our best performing strategy, achieves 84% of the optimized adoption with 2 initial interviews (necessitating 8 total interviews as the connections are interviewed), and up to 90% of the optimized adoption with 8 initial interviews (and 13 total interviews). Strategy E requires a total of 4 interviews, and achieves 69% of the optimized adoption. Strategy F achieves 60% of the optimized adoption with 2 interviews, and up to 67% of optimized adoption with up to 10 interviews.

Clearly the most effective strategies are those that identify a high degree farmer and train her connections. Given the nature of the complex contagion learning process, the intuition is clear: training two high degree friends of someone who is high degree means that three people with many connections in the same part of the network will become informed. With clustered networks, it is likely that others will as well.

Table A1: Agricultural Yields of Seeds Relative to Shadow (Counterfactual) Farmers

	Log of Agricultural Yields	
	(1)	(2)
Seed	0.126	
	(0.061)	
Adopted Pit Planting		0.443
		(0.210)
N	959	959
Mean of Shadows		
Year	2,3	2,3

Notes

1 Sample includes only seed and shadow farmers. Benchmark villages are excluded.

2 Agricultural yields were winsorized. The specification also controls for total farm size; controls used in the re-randomization routine (percent of village using compost at baseline; percent village using fertilizer at baseline; percent of village using pit planting at baseline); village size and its square; and district and year fixed effects. Standard errors are clustered at the village level.

Table A2: Characteristics of the Seeds Chosen by Each Treatment Arm

	Farm Size	Wealth Index (PCA)
	(1)	(2)
Treatment arm:		
Complex Contagion	-0.037	0.380
	(0.19)	(0.23)
Simple Contagion	-0.152	0.113
	(0.19)	(0.23)
Geographic	-0.614	-0.740
	(0.19)	(0.23)
P-values for Tests of Equality in Seed Characteristics		
Simple = Complex	0.335	0.067
Complex = Geographic	0.000	0.000
Simple = Complex = Geographic	0.000	0.000
N	1248	1248
Mean Value for Seeds in Benchmark Treatment (omitted category)	2.06	0.626
SD for Seeds in Benchmark Treatment	2.97	1.7

Notes

1 The sample includes all seeds and shadows. The sample frame includes 100 Benchmark farmers (2 partners in 50 villages), as we only observe Benchmark farmers in Benchmark treatment villages, and up to 6 additional partner farmers (2 Simple partners, 2 Complex partners, and 2 Geo partners) in all 200 villages.

2 Benchmark treatment seeds are the reference category.

Table A3: Distribution of Distance to Partner Farmers

	(1)	(2)	(3)	(4)
Path Distance to Closest Partner	Simple Partner	Complex Partner	Geo Partner	Benchmark Seed
1	38%	42%	24%	33%
2	50%	41%	46%	44%
3	9%	10%	20%	14%
4 +	4%	6%	10%	9%
N	4856	4856	4856	922

Notes

1 The data in this analysis includes respondents in our household surveys, linked to the social network census to capture their connections - direct and indirect - to the partner (or seed) farmers. Seed and shadow farmers are themsevles removed, as well as the 6.5% of households in our sample (419) with zero measured connections.

2 In columns (1)-(3), connections to both seeds and shadow farmers are analyzied, while in column (4) we only look at connections to the Benchmark seed in Benchmark villages.

Table A4: Seed Knowledge and Adoption

	Knows How to Pit Plant			Adopts Pit Planting			Adopts CRM	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Seeds	0.518	0.367	0.245	0.258	0.230	0.182	0.137	0.047
	(0.04)	(0.04)	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)	(0.04)
Years	1	2	3	1	2	3	1	2
N	659	735	503	686	672	489	686	467
Mean of Shadows	0.165	0.187	0.291	0.0541	0.0929	0.139	0.32	0.207
SD of Shadows	0.371	0.39	0.455	0.227	0.291	0.347	0.467	0.406
Panel B								
Simple diffusion	-0.133	-0.067	0.108	-0.006	0.129	0.176	0.078	-0.097
	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.09)	(0.08)	(0.09)
Complex diffusion	-0.120	-0.058	0.007	-0.020	0.002	0.037	-0.001	-0.077
	(0.07)	(0.07)	(0.08)	(0.08)	(0.07)	(0.08)	(0.08)	(0.09)
Geographic	-0.193	-0.255	-0.150	-0.095	-0.064	-0.003	-0.011	-0.075
	(0.07)	(0.07)	(0.09)	(0.08)	(0.07)	(0.08)	(0.08)	(0.10)
Years	1	2	3	1	2	3	1	2
N	343	383	264	353	352	259	353	243
Mean of Benchmark	0.824	0.653	0.547	0.337	0.276	0.238	0.442	0.339
SD of Benchmark	0.383	0.479	0.502	0.476	0.45	0.429	0.5	0.478
p-value for tests of equality in adoption rates across treatment cells:								
Simple = Complex	0.872	0.904	0.242	0.862	0.077	0.108	0.311	0.808
Complex = Geographic	0.377	0.016	0.111	0.36	0.358	0.625	0.886	0.977
Joint test of 3 treatments	0.472	0.021	0.011	0.252	0.008	0.049	0.235	0.795

Notes

1 In Panel A, all columns compare seed farmers to shadow farmers. Village fixed effects are included, and standard errors are clustered at the village level.

2 In Panel B, the sample includes only seed farmers, and the reference group is Benchmark seed farmers. The specification also includes controls which were used in the re-randomization routine (percent of village using compost at baseline; percent village using fertilizer at baseline; percent of village using pit planting at baseline); village size and its square; and district fixed effects. Standard errors are clustered at the village level.

Table A5: Test of Balance across Randomized Treatment Arms

	Simple	Complex	Geo	Benchmark	N	p-value of joint test
	(1)	(2)	(3)	(4)	(5)	(6)
Housing (pca)	-0.159	-0.036	0.023	0.106	14089	0.052
	(0.05)	(0.09)	(0.21)	(0.08)		
Assets (pca)	-0.059	-0.034	-0.040	0.005	14346	0.855
	(0.07)	(0.05)	(0.06)	(0.08)		
Livestock (pca)	0.012	0.025	-0.087	0.014	14346	0.210
	(0.06)	(0.06)	(0.04)	(0.06)		
Basal fertiliser (kg)	51.98	53.11	50.92	50.94	10427	0.970
	(4.78)	(3.14)	(3.17)	(2.23)		
Top dressing fertiliser (kg)	49.82	49.49	50.28	52.11	10526	0.787
	(3.33)	(2.05)	(2.53)	(1.99)		
# of Adults	2.305	2.316	2.299	2.306	14103	0.987
	(0.02)	(0.02)	(0.03)	(0.02)		
# of Children	2.617	2.650	2.619	2.599	14346	0.847
	(0.04)	(0.05)	(0.05)	(0.04)		
Farm size (acres)	1.624	1.676	1.764	1.808	14083	0.064
	(0.08)	(0.06)	(0.09)	(0.08)		
Own land	0.904	0.907	0.903	0.913	14346	0.922
	(0.01)	(0.01)	(0.02)	(0.01)		
Yields	304.20	290.46	303.54	300.77	13500	0.842
	(18.63)	(21.65)	(20.71)	(25.43)		
Provided Ganyu	0.254	0.250	0.242	0.233	14078	0.599
	(0.01)	(0.02)	(0.02)	(0.02)		
Used Ganyu	0.123	0.134	0.150	0.142	14078	0.115
	(0.01)	(0.01)	(0.01)	(0.01)		

Notes

1 Housing, assets and livestock in the first three set of rows are pca scores. Housing includes information on: materials walls are made of, roof materials, floor materials and whether the household has a toilet. Assets includes the number of bicycles, radios and cell phones the household owns. Livestock is an index including the number of sheep, goats, chickens, cows, pigs, guinea fowl, and doves.

2 Columns (1)-(4) give the means and standard errors of the variable listed in the title column in each of the treatment arms. The seeds and the shadow seeds are excluded from the sample. The data is from the social network census.

3 Column (6) shows the p-value of a joint test of significance of all treatment arms. Also included in the specification used for the test are controls used in the re-randomization routine (percent of village using compost at baseline; percent village using fertilizer at baseline; percent of village using pit planting at baseline) and district fixed effects. Standard errors are clustered at the village level.

4 Ganyu is the term used in Malawi for hired wage labor on the farm.

Table A6: Village Level Adoption Outcomes for Crop Residue Management (CRM)

	Any Non-Seed Adopters	Adoption Rate
	(1)	(2)
Complex Diffusion Treatment	-0.083	-0.037
	(0.062)	(0.027)
Simple Diffusion Treatment	-0.064	-0.026
	(0.060)	(0.027)
Geographic treatment	-0.152	-0.054
	(0.070)	(0.029)
Year	2	2
N	141	141
Mean of Benchmark Treatment (omitted category)	0.971	0.204
SD of Benchmark	0.169	0.109
p -values for tests of equality of coefficients…		
Test: Simple = Complex	0.794	0.680
Test: Complex = Geo	0.258	0.366
Test: Simple = Geo	0.336	0.583

Notes

1 The "Any non-seed adopters" indicator in columns (1) excludes seed farmers. The adoption rate in column (2) include all randomly sampled farmers, excluding seed and shadow farmers.

2 Analysis restricted to data from Mwanza and Machinga.

3 All columns include controls used in the re-randomization routine (percent of village using compost at baseline; percent village using fertilizer at baseline; percent of village using pit planting at baseline); village size and its square; and district fixed effects. Standard errors are clustered at the village level.