Can Network Theory based Targeting Increase Technology Adoption?

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

November 2014

PRELIMINARY DRAFT. PLEASE DO NOT CIRCULATE OR CITE

Abstract

In order to induce farmers to adopt agricultural technologies in Malawi, we apply diffusion models of simple and complex contagion on rich social network data from 200 villages in Malawi to identify optimal seed farmers to target and train on the new technologies. A randomized controlled trial compares these theory-driven network targeting approaches to simpler, scalable strategies that either rely on a government extension worker or an easily measurable proxy for the social network (geographic distance between households) to identify seed farmers. Adoption rates over three years are greater in villages that received the theory-based data intensive targeting treatments. The data, interpreted through the lens of the theory, yield insights on the nature of diffusion, and are most consistent with a learning environment where farmers need to know more than one person with knowledge of the technology before they adopt themselves.

JEL Codes: O16, O13

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

* Contact: Beaman: l-beaman@northwestern.edu, BenYishay: a.benyishay@unsw.edu.au, Magruder: jmagruder@berkeley.edu, Mobarak: ahmed.mobarak@yale.edu.

We thank the CEGA/JPAL Agricultural Technology Adoption Initiative (ATAI) and 3ie for financial support. 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, Sylvan Herskowitz, 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, Chris Udry for comments.

1. Introduction

The slow adoption of agricultural technologies is a persistent puzzle in development economics. Lack of credible information is one potential constraint to adoption, and social relationships can serve as important vectors through which farmers learn about, and are then convinced to adopt, new agricultural technologies (Griliches 1957, Foster and Rosenzweig 1995, Munshi 2004, Bandiera and Rasul 2006, Conley and Udry 2010).[1] Beyond academic research, social influence often plays a key role in technology dissemination schemes. For example, agricultural extension services often rely on training a few farmers in a new technology, and expect knowledge to diffuse from those farmers to other farmers in the area. The reliance on network-based diffusion is particularly strong in developing countries, where extension resources are scarce. If key individuals within a network are more effective communicators, then agricultural extension will be most effective if it can target these key individuals. In this project, we use a large-scale field experiment in Malawi to evaluate whether network theory-based targeting strategies for disseminating information can be used to increase adoption of a new agricultural technology for farmers in arid regions of Africa.

There is a rich theoretical literature on diffusion processes (see Jackson 2008 Chapter 7 for a review). For tractability, we refine our focus to an important class of diffusion models: threshold models, where individuals adopt if they are connected to at least a threshold number of adopters (e.g. Granovetter 1978; Centola and Macy 2007; Acemoglu *et al* 2011). We test the predictions of the underlying model by experimentally varying the identity of information seeds, that is, the relatively scarce individuals who are trained in the new technology, and from whom information

[1] This is more broadly true in other areas of economics, sociology and public health where scarce information may slow adoption of technologies, behaviors, or opportunities. Large literatures in economics (Munshi 2008, Duflo and Saez 2003, Magruder 2010, Beaman 2012), finance (Beshears et al. 2013, Bursztyn et al. 2013), sociology (Rogers 1962), and medicine and public health (Coleman et al 1957; Doumit et al 2007) show that information and behaviors spread through inter-personal ties.

may spread.[2] Within threshold diffusion models, the importance and identity of optimal seeds depends sharply on the threshold parameter. In the case of diffusions where individuals have a low threshold, the choice of seeds is fairly innocuous: if one connection is suitable to motivate adoption, then adoption diffuses quickly for most choices of seeds. If multiple connections are needed to encourage adoption, however, then the choice of the seed farmers is critical. Many (and often most) potential seed pairings would yield no adoption at all. Centola and Macy (2007) characterize these two threshold models as either a simple contagion (when the threshold is 1) or a complex contagion (when the threshold exceeds 1).

To assess whether diffusion models can improve the effectiveness of public policy, we test whether training theoretically optimal diffusion partners (under different assumptions on the contagion threshold) leads to greater adoption of a new technology. To do so, we select optimal network partners using a full social network census, which we collected in 200 villages in Malawi. On those 200 networks, we simulated the optimal partners under different assumptions about the median threshold, determined who would be the best choices for that diffusion model, and gave their names to the Ministry of Agriculture extension workers for training. We then trace adoption patterns in these villages over the next 2-3 seasons to test which sets of partners are most effective.

We benchmark the adoption in villages with our theoretically optimal seeds against a treatment where agricultural extension agents use local knowledge to select partners to train. Typically, this involves asking village leaders to nominate a pair of extension partners. Interventions that rely on local institutions may use a great deal of information in selecting these influential people, including their eagerness to try the new technology, their persuasiveness as communicators, and the

[2] One challenge in adapting theoretical results for this goal is that many if not most key predictions are derived for the implications of network structure for diffusion (see Jackson 2008 Chapter 7). Yet, existing learning networks are predetermined and not experimentally manipulable. Moreover, it seems natural to expect that heterogeneity in underlying social structures reflects important heterogeneity in local conditions and institutions, particularly those related to the learning environment, raising concerns over validity of estimates which would leverage this variation (e.g. Allcott 2014).

trust other villagers have in their opinions. As such, our benchmark renders a strong test of diffusion theory: our theoretically optimal partners were selected only by their position in the network, without the advantage of these characteristics.

Our experiment focuses on the decision to adopt 'pit planting', a traditional West African technology which is close to unknown in Malawi. Pit planting has the potential to significantly improve maize yields in arid areas of rural Africa.[3] Agricultural productivity has remained especially low and flat in sub-Saharan Africa for the last 40 years, and low adoption of productive technologies bears a significant part of the blame (World Bank 2008). The network targeting experiments are therefore conducted in an important setting that holds large consequences for growth in Africa.

We find that the data-intensive, theory-driven targeting of optimal seed farmers out-performs the simpler approaches to choosing seeds in terms of technology diffusion across the village over two or three years. Network theory based targeting increases adoption by 3-4 percentage points more than relying on the extension worker, during the 3-year period of the experiment when pit planting adoption grew from 0% to about 10%. Complex contagion models suggest that one of the potential consequences of poor targeting is complete failure to adopt within the village, and we find that this possibility is empirically relevant: using theory-based procedures to identify seeds leads to a 50% greater likelihood that at least one other person in the village adopts. The results suggest that simply changing *who* is trained in a village on a technology on the basis of social network theory can increase the adoption of new technologies compared to the Ministry's existing extension strategy. We also evaluate a more policy-relevant alternative to the data-intensive approach by choosing optimal seeds using geographic proximity as a proxy for network connections. The data show that while physical proximity is not always a perfect proxy for social

[3] It has been shown to increase productivity by 50-100% in lab and field 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 2014), and using experimental variation among villagers in this study.

connections, even the low-cost geography based targeting strategy generates some gains in adoption relative to the status quo benchmark. This strategy is much cheaper to implement than the theory-driven approaches, which suggests that developing methods to identify other low-cost proxies for social network structure would be a useful policy-relevant avenue for future research.[8]

After documenting these basic program evaluation results, we return to the theory to generate additional predictions on patterns of diffusion we should observe in our data under models of complex contagion or simple contagion. Theory predicts, for example, that if the learning environment is complex, then connections to two or more seeds should be highly predictive of individual-level adoption decisions, even relative to individuals who have a direct connection to one seed. Clustering agents to focus on only one part of the network at the expense of the rest of the village should increase technology adoption relative to dispersing agents to cover the whole village. We examine these predictions in our data, and find that knowledge diffusion and technology adoption patterns among the 4000 surveyed farmers in 200 Malawian villages are most consistent with a complex learning environment. As complex learning environments are the ones in which theory suggests the selection of partners may have large implications for results, this finding supports the need of a more careful consideration of diffusion patterns. Finally, the analysis yields some policy-relevant insights: (a) to promote a new agricultural practice, tightly clustering dissemination agents in the dense part of the network may be preferable to dispersing those agents, and (b) developing other low-cost proxies for network structure, and using them to identify seeds may be a cost-effective policy tool to speed up the diffusion process.

In addition to the broader social learning literature, our approach is grounded in two recent empirical literatures. First, a few recent interventions in development economics have either used

[8] For example, promising results in Banerjee et al (2014) imply that households know who is central in their village, and this type of information may be easily elicited from a random sample of people.

strategies that involve either purposefully partnering with important nodes in social networks to promote new technologies[9], or tested whether adoption changes when individuals at particular network locations were selected as partners (Banerjee et al 2012). These approaches typically identify social network partners through informal methods, such as village focus groups or established village leadership positions[10]. These informal approaches rely on existing village institutions to select dissemination partners, and have the advantage of being easy and inexpensive to implement. However, it becomes difficult to anticipate how similar procedures would perform in other contexts since the selection process for network partners may be idiosyncratic to local institutions or may even change endogenously when the learning process is manipulated[11].

The empirical relevance of this latter concern is presented forcefully by Carrell et al (2013). Carrell et al estimate the distribution of peer effects on air force academy student performance, and create new classrooms which are either formed to maximize positive peer effects based on the empirical trends. Surprisingly, they find worse outcomes in their optimal classrooms, particularly for the low ability students who they anticipated being most positively influenced. They document several patterns that suggest that the structure of peer effects changed when classroom composition was changed.

It is possible that the absence of theory contributed to the negative outcome in Carrell et al 2012: while some varied theories may anticipate particular types of peer effects in particular places, the peer effects literature more generally emphasizes empirical results over theory (and Carrell et al 2012 is no exception). Our approach builds on these two literatures by not only attempting to

[9] Kremer et al (2011) identify and recruit 'ambassadors' to promote water chlorination in rural Kenya, Miller and Mobarak (2014) first markets improved cookstoves to 'opinion leaders' in Bangladeshi villages before marketing to others, Kim et al (forthcoming) promotes multi-vitamins and water chlorination through network nodes in Honduras, and BenYishay and Mobarak (2014) incentivize 'lead farmers' and 'peer farmers' to partner with agricultural extension officers in Malawi.

[10] While our benchmark group is motivated by a comparison to "business-as-usual" practice, in this context they closely resemble the treatment groups from these other interventions, again highlighting how the bar for success is set high.

[11] Banerjee et al (2013) show that heterogeneity in network characteristics across individuals in the same leadership position leads to differential take-up of a microfinance product by their connections.

control the social learning process but by the use of theoretical predictions on learning behavior rather than expressed empirical patterns or local institutions to generate the selection[12]. Since the diffusion theory used to pick the partners is not specific to the context, we may have stronger priors that these theoretically-informed results will be relevant elsewhere.

The paper is organized as follows. We present the theoretical model on which the experimental design is based in Section 2. Section 3 discusses all field activities, including data collection and intervention implementation. Section 4 describes the characteristics and activities of the seed farmers and the performance of the technology in the field. Section 5 presents the basic program evaluation results. Section 6 presents more detailed theoretical predictions and associated empirical results on the nature of contagion and network diffusion. Section 7 concludes.

2. Theoretical model and experimental design

Our experiment takes place in 200 villages randomly sampled from three 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: more than 60% of the population's calorie consumption derives from 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 directly tied to welfare.

The existing agricultural extension system in Malawi relies on Agricultural Extension Development Officers (AEDOs) who are employed by the Ministry of Agriculture and Food Security (MoAFS). Many AEDOs are responsible for upwards of 30-50 villages, which implies that direct contacts are sparse. According to the 2006/2007 Malawi National Agricultural and Livestock

[12] The social learning context may facilitate the use of theory compared to peer effect results, for the same reasons that many of the empirical concerns with peer effect estimates are more muted in adoption contexts. When we are considering a new technology, there is only one direction for information to flow, lessening the reflection problem; moreover, insofar as we are concerned with the spread of information we understand the mechanism that underlies that effect (an informed person tells an uninformed one) somewhat better than we understand why students benefit from high-performing peers, for example.

Census, only 18% of farmers report participating in any type of extension activity. Against this backdrop of staff shortages, incorporating social learning in the diffusion process may be a cost-effective way to improve the effectiveness of extension.

2.1 Diffusion Models and Experimental Design

We develop network-theory based strategies to disseminate information about new agricultural technologies in partnership with the Malawi Ministry of Agriculture. The underlying theoretical basis for these strategies is the linear threshold model (Granovetter 1978; Acemoglu *et al* 2011). This model posits that an agent will adopt a new behavior once adoption behavior among his connections crosses a threshold. The model was originally designed to study a wide array of collective behaviors including riots, voting, migration, and new technology adoption. The underlying rationale for this formulation is either that the net benefits of adoption are a function of neighbors' adoption decisions (e.g. because a farmer expects to continue learning from neighbors' experiences on how to make best use of the technology), or because farmers need to hear about the new technology from multiple sources before they are persuaded to adopt (when the threshold is above 1).

We employ two different versions of the threshold model in different arms of our experiment. The first version, called "simple contagion," postulates that the average individual needs to know only one other household who has adopted the technology in order to be convinced to adopt herself. Centola and Macy (2007) shows that some types of information – such as knowledge of job opportunities - spread through simple contagion. However, other behaviors may require multiple sources of information before they are adopted, and we explore this using a complex contagion model in a second arm of our experiment[13]. Centola (2010) provides experimental

[13] 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.

evidence that health behaviors diffuse more quickly through networks where links are clustered than through those where links are random (holding network size and degree constant), 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. In contrast to Centola, Eguiliz, and Macy (2007), they argue that long links continue to be valuable especially with the number of seeds is small. While this literature has focused on identifying the ideal network structures for maximizing diffusion, we instead apply these models in a field experiment to understand how to target information within a network in order to best exploit the pre-existing social network architecture of villages in Malawi.

The experiment to select two "seed farmers" in these villages based on these two models was implemented as follows. We first collect network relationships data (described in detail in section 3) on the census of households in each village before launching any field intervention activities. The social network structures observed in these data allow us to construct network adjacency matrices for each of the 200 villages in our sample. 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 $N(\lambda, 0.5)$ but truncated to be strictly positive. We conduct simulations with $\lambda=1$ and $\lambda=2$ in all villages to observe optimal seeds under simple and complex contagion respectively. In these simulations, when an individual is connected to at least τ individuals who adopted, he adopts in the next period. Once an individual adopts, we assume that all other household members also adopt, since agricultural plots are held at the household level in Malawi.[14] We run the model for four periods, which corresponds to our data collection activities, in that we surveyed the sample villages at baseline, and for up to three agricultural seasons after the interventions were implemented.

[14] The simulation excludes disadoption, so adoption decisions are permanent.

The final step to prepare for the experimental interventions is to choose the "optimal" partner farmers for each village as prescribed by the theoretical simulation randomly assigned to that village. To accomplish this, we pick a pair of individuals in the village and assign them the role of seed farmers, and predict the village adoption rate after four seasons under the specified contagion theory. Given the randomness built in to the model, we simulate the model 2000 times and create a measure of the average adoption rate induced by these two seeds. We repeat this process for every other possible pair of seed farmers in the village, and ultimately select the pair that yields the highest average adoption rate. We thus obtain the optimal pair of seeds for each village under each potential treatment arm.

2.2 Interventions

The two seed farmers in each village are trained in the targeted technologies by the Malawi Ministry of Agriculture extension staff. Our experimental variation only changes the process by which the seed farmers are selected in each village, and holds all other aspects of the training constant. Within each district, we randomly assign villages to one of the following four treatment arms (or seed farmer selection process)[15]:

1. <u>Simple Contagion:</u> Simple diffusion (λ=1) model applied to the network relationship data
2. <u>Complex Contagion:</u> Complex diffusion (λ=2) model applied to network relationship data
3. <u>Geo Treatment:</u> Complex diffusion (λ=2) model applied to an adjacency matrix where geographic proximity proxies for a network connection
4. <u>Status Quo Benchmark:</u> Extension worker selects the seed farmers

Treatment arms 1 and 2 were described above. In treatment arm 3, the simulation steps are the same as in the Complex Contagion case, except that we apply the procedure to a different

[15] Randomization was 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.

adjacency matrix that is generated 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.

The fourth group is the status-quo benchmark, where AEDOs were asked to select two seed farmers as they normally would in settings outside the experiment. Comparing the adoption performance of network theory-based targeting against this benchmark constitutes a meaningful and challenging test for the simple and complex contagion treatments. In principle, the AEDOs could use valuable information not available to researchers, such as the individual's motivation to take on the role, to select highly effective seed farmers. It is not clear that the theory-driven diffusion strategies would out-perform this benchmark. Another option would have been to randomly select seed farmers from the population, but that would have constituted a weaker test, and one with little real-world relevance as extension programs rarely randomly choose their partners. Allowing extension staff to select the seeds is what the Malawi Ministry of Agriculture and other policymakers would normally do, so this is the most relevant counterfactual.

Note that the Simple, Complex, and Geo seed farmer selection strategies were simulated in all 200 villages, so we know – for example – who the optimal simple contagion seed farmers would have been in a village randomly assigned to the complex contagion or the geo treatment. We label the counterfactual optimal farmers as "shadow seeds." This is very useful for analysis, because in any regression where we examine decisions made by the actual seed farmers to understand who they are and the attributes they possess, the shadow seeds form the relevant comparison group as the comparison of seeds and shadows utilize the random variation created by the experiment. When we report effects on the broader village population, we exclude both the actual and the shadow seeds from the analysis.

Finally, we note that in approximately 50% of villages, there was at least one seed who was optimal under both the simple and complex models. This happens when there is a very obvious individual in the network who is essential for diffusion[16]. In these cases, the two treatment arms are naturally less distinguishable.

3. Field Activities: Implementation of Interventions and Data Collection

3.1. Training of Seed Farmers

After we produced the lists of seed farmers for each village using the procedures described above, the AEDO assigned to the village trained the two seed farmers. As the technologies themselves were new, the AEDOs were themselves by staff from the Ministry's Department of Land Conservation (details on the technologies are discussed below). We provided AEDOs with two seed farmer names for each village in experimental arms 1-3, and then replacement names if either of the first two refused to participate. Refusal was uncommon, and we conduct intent-to-treat analysis using the original seed assignment. The trainings took place in April-May of 2011 for Machinga and Mwanza districts, and March-April of 2012 for Nkhotakota. Following the training of seed farmers by AEDOs, all seed farmers were also informed that they would receive a small in-kind gift (valued at US$8) if they themselves adopted pit planting in the first year (and that the gift would be given only in the first year). The gift was given at the time of follow up data collection and verified on the farm by the enumerator[17].

3.2 Technologies

In this section we describe the two technologies introduced to seed farmers and in section 4.3 we analyze data on crop yields to give further insights into the benefits of the technologies.

[16] Consider a network that is configured as a star. In both threshold models, the middle individual will be selected as a seed.
[17] There was no gift or incentive offered or provided on the basis of others' adoption in the village.

Pit Planting

Maize farmers in Malawi traditionally plant in either flat land or after preparing ridges. Ridging has been shown to deplete soil fertility and decrease agricultural productivity over time (Derpsch 2001, 2004). In contrast, pit planting, which is the main technology we train the seed farmers on, 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. The technique is practiced elsewhere in Africa, and has been shown to greatly enhance maize yields both in controlled trials and in field settings (BenYishay and Mobarak 2014). In the next section we offer further evidence on yield impacts in our sample of villages. The enhanced productivity is thought to derive from two mechanisms: (1) reduced tillage of topsoil, which allows nutrients to remain fixed in the soil rather than eroding, and (2) concentration of water around the plants, which aids in plant growth during poor rainfall conditions. The gains from the first mechanism over a counterfactual of continued ridging are thought to accumulate over time, while the gains from the second are expected to accrue even in the very short run. Studies of pit planting in southern Africa have found returns of 50-100 percent for maize production (Haggblade and Tembo 2003) within the first year of production.

Practicing pit planting may involve some additional costs. First, only a small portion of the surface is tilled with pit planting, and hand weeding or herbicide requirements may therefore increase. Second, digging pits is a labor-intensive task with potentially 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 (2014) show that the yield effects of pit planting are large in four other districts of Malawi, while the change in costs is negligible in comparison.

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. The trainings also addressed potential concerns about modifications in semi-arid areas (where there are fewer residues available), pest infestation, fire prevention, and alternative sources of livestock feed. There is little experimental evidence on the impacts of CRM on soil fertility, water retention, and yields in similar settings.

3.3 Data

The interventions were designed on the basis of social network census data collected from all sample villages at baseline. After training the seed farmers, we collected up to three rounds of longer household survey data for sub-samples of the village populations. Figure 1 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—including relational statistics of these networks—requires relatively complete information on network relationships within the village (Chandrasekhar and Lewis 2011). To collect this data, our field teams listed all adults in each of our sample villages and created a database with all adult names and household structures for each village. For each household, a roster of all household member names, nicknames, maiden names, genders, relationships, and ages was completed. Netbook computers were used by the field teams to identify links in real-time. The field teams completed a census within each village, attempting to interview one man and one woman in each household. In practice, we reached more than 80% of households participating in the census in every sample village.

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. The individual questionnaires asked about agricultural contacts several ways: first by asking 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 some 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 relatives, fellow church or mosque members, or farmers whose fields they pass by on a regular basis. We also elicited contacts with whom they share food and close friends. These 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 30 households in each village.[18] In villages with fewer than 30 households, all households were surveyed. Three survey rounds were conducted in Machinga and Mwanza in October-December of 2011, 2012 and 2013. In Nkhotakota, two survey rounds were conducted in October-December of 2012 and 2013.[19] The initial rounds referenced 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, we observe 3 adoption decisions for pit planting for farmers in Mwanza and Machinga, and 2 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.

Rainfall Data

Because the effects of the technologies vary across rainfall conditions, we obtain daily precipitation data over 9km grid cells from aWhere (2014). aWhere's weather data are assembled from ground meteorological stations and orbiting weather satellites, with daily precipitation data derived from Colorado State University's near-real time implementation of a high resolution, global, satellite precipitation product. The data product is a multi-sensor combination of several satellite passive microwave precipitation algorithms available in near-real time from NOAA, which is then processed using a 3-D spline interpolation. Using these data, we construct seasonal total precipitation at each village location.

Randomization and Balance

[18] 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.
[19] Unanticipated delays in receiving project funding required us to start training of AEDOs and seed farmers in Nkhotakota in 2012 instead of 2011 as in Mwanza and Machinga.

Table A1 shows how observable characteristics from the social network census vary with the treatment status of the village. The table shows the results of a regression of the dependent variable listed in the column heading on indicators for the respondent residing in a benchmark, simple, complex, or geo treatment villages. District fixed effects are included in the regression, and standard errors clustered at the village level. P values from statistical tests comparing across the different treatment groups as well as a joint test of all treatment groups are displayed. Given the large number of comparisons made in Table A1, few differences across treatment groups are statistically significant. Farm size, in column (9), is the most concerning: Farmers in the Benchmark villages have larger farm sizes on average than farmers in Complex villages in particular, and the joint test across the treatment variable is significant at the 5% level.

Attrition

4 Characteristics of the Seed Farmers and the Technology

4.1 Characteristics of Each Type of Seed Farmer

The simulations of the simple and complex contagion models generated different optimal seeds in most but not all cases. In 50% of villages, there was at least 1 seed who was judged as optimal in more than one (simple, complex or geo) models. The experimental design also allowed extension workers to choose any seed farmer they wanted in the benchmark treatment, and this may have sometimes coincided with the network theory-targeted seeds. However, the treatment arms generated different types of seed farmers in general as discussed below. They also generated different clustering patterns. For example, 35% of our random household sample has a connection to a simple seed, and 6% are connected to both simple seeds. However, 18% of households are connected to two complex seeds and 28% are connected to one complex.[20]

Table 1 compares the seed farmers chosen in the four different experimental arms in terms of observable characteristics such as wealth and land size from our survey data, and in terms of centrality measures computed from our social network census data. The most striking pattern in Table 1 is that the seeds selected under the geographic treatment are much poorer than other seeds. This is because many households live on their farm land in Malawi. Therefore households who are geographically closer to other people also have less land, and these households tend to be poorer overall. Therefore while the idea of using geography as a proxy for one's network may be intuitive, the implications of geographic centrality may be highly context-specific.

Seed farmers selected through the complex contagion simulations are the most "central" across all measures of network centrality we compute. Seed farmers in the complex contagion villages have three (20%) more direct connections to others in the village than the seed farmers chosen by the extension workers. Seeds in complex contagion villages also possess the highest between-ness and eigenvector centrality measures, which imply that they are important nodes in these villages.[21] Simple seeds have similar betweenness centrality as complex seeds, but lower eigenvector centrality and closeness.

Figure 2 shows five example villages from our data with network links mapped and the locations of the simple, complex and geo seeds within the village social networks. One feature common across these villages is that the simple seeds tend to be more distant from one another than do the complex or geo seeds. In village 45, for example, one central household was chosen as a seed

---

[20] For the geo-based seeds, 20% of households are connected to one, 10% connected to two and a larger fraction than in either network theory-based treatments – 70% – are connected to no seed.

[21] Eigenvector Centrality is weighted sum of connections, where each connection's weight is determined by its own eigenvector centrality (like google pagerank). Betweenness centrality captures that a person is important if one has to go through him to connect to other people. Therefore it is calculated as the fraction of shortest paths between individuals in the network that passes through that individual. See Jackson (2008) for more details.

in both Simple and Complex models, but the selection of the second farmer reveals the main difference between these models. In complex contagion, the second seed farmer is directly connected to the first seed and is also quite central in the network. The second simple seed, however, is far more removed from the giant component in the network. Under simple contagion, training the first seed is sufficient to induce the diffusion process to occur within the main cluster in the village, and the second - more removed farmer - was otherwise unlikely to adopt without being directly targeted. More broadly, the difference in network locations between simple and complex seeds occurs because the simple contagion diffusion pathways from each seed farmer need not overlap, while it is crucial that at least one individual be linked to both seeds under complex contagion. Accordingly, targeting is less important for eventual adoption outcomes under simple contagion compared to complex contagion, an idea we return to in section 6.2.

All the example villages in Figure 2 show that the geo seeds are generally close to one another. This is because the underlying diffusion model is complex contagion. However, they are located in more peripheral locations within the network, as anticipated given the summary statistics in Table 1. Figure 3 shows four example villages which also include Benchmark seeds. As Table 1 suggests, Benchmark farmers are more central in the network than Geo farmers, but less central than Complex farmers. Most importantly, they are rarely sufficiently clustered in the network to spark the diffusion process if decisions are governed by the complex contagion model.

4.2 Do Seed Farmers Adopt the Technology Themselves?

In order for us to learn about the diffusion process in this experiment, it must be the case that training seed farmers first led to some initial level of technology adoption. Table 2 shows that indeed the interventions increased the likelihood that the seed farmers themselves adopted the technologies. The sample is restricted to seed and shadow farmers only, so this specification captures the causal effect of the intervention, and not differences in adoption across farmers at different positions within the network.[22] Panel A focuses on pit planting and Panel B on crop residue management. Seed farmers who are trained on the technology are 17-25 percentage points more likely to adopt than comparable shadows across all three years. Adoption rate among shadows was 5-14% across years, so this represents a large increase. We provided an in-kind incentive for the seed to adopt pit planting in the first year but not thereafter. The persistent adoption difference is suggestive that the seeds who tried out pit planting found the technology to be advantageous. We never provided the seeds any incentive to adopt CRM, but the trained farmers were also 13 percentage points more likely to use CRM in the first year. CRM was a much better-known technology to begin with, with 33% of shadows practicing it in the first year. CRM adoption dropped, however, in the second year among both actual seeds and the shadows.

These results are consistent with the observation that pit planting is a newer and unknown technology for which information constraints were probably more relevant. Pit planting adoption among those trained was also persistent, which suggests that the seed farmers found the method useful. In contrast, CRM take up did not persist, which could mean that the technology was not well suited for these farmers. This makes analysis of the diffusion of CRM more complicated, because it is possible that the message "do not adopt" was passed within the network, and adoption propensity among others in the village may not be the right outcome variable for our experimental design.

Table 3 restricts the sample to only seed farmers who were trained (and drops all shadows) to examine whether adoption behavior varies across the four types of seeds in the four experimental arms. In the first year, there are no differences in adoption propensities (or in the likelihood of recalling the existence of the technology) across the four types of seeds. Columns (2) and (3) show that seed farmers in simple contagion villages become relatively more likely over time to adopt the

[22] Benchmark villages are also excluded since we do no observe counterfactual Benchmark seeds in Simple, Complex and Geo treatment villages.

technology. Their adoption propensity is significantly higher than the AEDO-selected seeds in years 2 and 3, which is striking because AEDOs could have screened partner farmers based on their interest in using the new technology. On the other hand, differences in years 2 and 3 could also be an outcome of the experiment, as seed farmers receive more feedback from other members of their network who try out pit planting, which in turn affects their own decision to continue. Columns (1)-(2) show that there are no significant differences in adoption in seasons 1 or 2 for crop residue management.

4.3 Effect of Technology Adoption on Crop Yields

We collected data on maize yields in our follow-up surveys, and we use this to show in Appendix Table A2 that the technologies we promoted led to an increase in output. We further use rainfall variation to study heterogeneity in the yield gains, because pit planting is more productive under arid conditions, when soil moisture retention in the pit is most important. This allows us to establish that the information about pit planting that diffused through the networks was likely positive on average. That in turn would allow us to interpret more adoption of pit planting as a signal of greater information diffusion.

We compare seed farmers to shadow farmers to study yield effects, exploiting the randomization in the experimental design.[23] In an intent-to-treat specification, maize yields among seed farmers (who were both trained on the technologies and promised a small reward to adopt) are 11% greater than the yields experienced by the comparable shadows. The second column of Table A2 examines the heterogeneity in this yield effect across rainfall states. This specification allows a linear interaction with rainfall, and indicates that the productivity on the seed farmers' plots is 32% greater in the bottom quintile of rainfall in our sample, and we estimate a zero effect in the top quintile of rain. To put the effect size in perspective, the returns to pit planting are as large as the

[23] Benchmark villages are again excluded.

yield increase from moving from the bottom quintile of rain to the fourth quintile. The heterogeneity results strongly suggest that the yield increases for seed farmers comes from adoption of pit planting.

We report the local average treatment effect using an IV regression in the third column in which we instrument pit planting adoption with an indicator for being randomly assigned the role of actual seed farmer who was trained and incentivized to adopt (rather than a shadow). In this specification, pit planting adoption is associated with a 45% increase in maize yield. However, we cannot rule out that CRM adoption also increased yields, potentially violating the exclusion restriction in the IV estimation.

4.4 Seeds Farmers' interactions with other villagers

Thus far, we have documented that the seed farmers trained on the technologies are more likely to adopt the technology themselves, realize some productivity gains from pit planting and persist with adoption, and that some types of seeds are more network-central than others. Next, we investigate whether these seed farmers exert any effort to disseminate information about pit planting to their neighbors in the village.

Table 4 uses data collected in the first follow-up data collection on conversations about pit planting that all respondents had with others in the village. Each respondent was asked questions about seven other individuals in their village, whether they knew them, and what they had discussed. The seven individuals comprised of the two seed farmers, some randomly selected shadow farmers, and a random sample of other village residents.[24] The empirical challenge with documenting more conversations with the seeds trained on the technologies is that these seeds were chosen to be network central, and such individuals would have more conversations with others regardless of our

[24] In Nkhotakota, the definition of the boundaries of the village is not uniformly agreed upon. In some cases, the extension workers selected seed farmers from outside of the geographic area that our listing exercise defined as a village. We have limited information about connections between individuals in the village and the seed farmers in such cases.

experimental treatments. We instead exploit the random variation in the experiment, and compare conversations with the (say) complex farmers who were assigned the role of seed farmer by our intervention to communication with the complex shadows in other villages who are observably similar, and who *would have been* the seed had those comparison villages been assigned to the complex contagion treatment.[25] In other words, we test whether a potential seed *being trained* on pit planting increases the likelihood that he talks to others about pit planting.

Table 4 shows that the experiment did induce the seed farmers to discuss pit planting with fellow villagers. Column (1) shows that there are more discussions with the "simple seed" in both Simple and Complex treatment villages compared to the benchmark villages. As expected, the effect is significantly larger in Simple treatment villages (4.6 percentage points) than in Complex treatment villages (1.9 percentage points), and these represent large increases over the mean value (2) in the benchmark villages. We observe a treatment effect even in Complex villages because, as mentioned above, there is considerable overlap in the optimal seeds chosen through the complex contagion and the simple contagion simulations. Recall that approximately 50% of villages have at least one farmer who is optimal under both Simple and Complex models.[26] Columns (2) and (3) show, analogously, increases in conversations about pit planting with the complex farmer in Complex treatment villages (a 3.6 percentage point increase compared to benchmark) and with the geo farmer in Geo villages (3.1 percentage points). In summary, the seed farmers trained in the pit planting method discussed the technology with others in their villages as a result of our experiment.

[25] While all sample respondents in Simple treatment villages were asked about simple farmers, not all respondents in the remaining villages were, since we chose a random subset of shadow farmers. This is analogously true for complex and geo villages. We therefore flexibly control for the number of simple (complex, geo) farmers we asked about in the regression where the dependent variable is talking about pit planting with the simple (complex, geo) farmer.

[26] If we exclude from this regression villages where there is overlap in the optimal farmers, we observe an increase in conversations with Simple farmers only in the Simple contagion treatment villages.

5. Program Evaluation Results

5.1 Does social network-based targeting increase adoption?

The Granovetter (1978) and Acemoglu et al (2011) threshold model of network diffusion suggests that to maximize technology adoption, information or other inducements to adopt should be targeted to key individuals within a network. The first step in our program evaluation therefore examines whether threshold model-based targeting improves the adoption rate of a seemingly productive, welfare-enhancing technologies. Table 5 focuses on pit planting adoption, because our analysis of seed behavior and yield effects indicate that the experiment induced persistent adoption of only pit planting among seeds, and there is suggestive evidence that seeds experienced higher yields by practicing pit planting.[27] We compare the pit planting adoption rates in all three seasons between villages where social network-based targeting was implemented, against the benchmark villages where AEDOs chose the seeds, and villages in which geographic proximity was used as a proxy for network connections.

The dependent variables measure adoption propensities in the village computed using only farmers who are neither seeds nor shadow farmers). We capture adoption in three ways: the adoption rate in the village (columns (1)-(3)), the total number of adopters (columns (4)-(6)), and an indicator for whether there was any adoption (columns (7)-(9)). The latter serves as an indicator for longer-term adoption: if there is no new adoption by season 3 (as happens in 46% of benchmark villages[28]), there is little prospect for continued adoption of pit planting.

We see no differences in the village-level adoption rate of pit planting in the first season: this likely reflects a time lag between information acquisition and adoption information among non-seeds. In season 2, however, villages where information was targeted to farmers based on the network-based simulations achieve a higher level of adoption of pit planting, by 3.1 percentage

[27] Adoption patterns for crop residue management are presented in the appendix.
[28] We only observe year 3 adoption in in Mwanza and Machinga.

points, than in benchmark villages. Since the adoption rate is only 4% in the benchmark case, this constitutes an 70% increase in adoption rates over villages where AEDOs selected seeds. These benchmark villages experience a significant increase in the adoption rate between seasons 2 and 3 (7.7 percent compared to 4.4 percent).[29] The adoption rate remains 2.2 percentage points higher in villages where social network-based targeting was applied, but this gap is not statistically different.

We see a very similar pattern when we use the number of adopters rather than the adoption rate as our dependent variable (with the estimation weighted by village size). There is no difference in season 1, a significant increase in season 2 (1.7 additional adopters, increasing the 1.9 adopters on average in benchmark villages by almost 90%), and a qualitatively similar magnitude but imprecisely estimated difference in season 3 (1.35 additional adopters over the 4 adopters in benchmark villages).

In columns (7)-(9), we again see that there is no difference across treatments during season 1 in an indicator for any non-seed adoption. In season 2, only 46% of benchmark villages have at least one adopter (among our randomly selected sample). This measure rises to 65% in the network-based villages, a difference which is significant with at the 5% level. In season 3, 54% of benchmark villages have some adoption, while Network-based targeting achieves at least some adoption in 79% of all villages. A key difference between the benchmark and the use of network-based targeting is thus on the extensive margin, i.e., whether there is any diffusion at all. The theoretical simulations had also suggested that differences in diffusion rates would become apparent in the second or third periods. This was a key feature of the complex contagion simulations in particular, which we will discuss in more detail in section 7.

[29] Appendix table A3 shows the analogous results for crop residue management. Overall, there is less evidence of gains from social network targeting for CRM than pit planting.

These large relative gains in diffusion are not at all obvious ex-ante, because the extension worker may have chosen seeds to optimize on useful personality traits such as diligence, stature, credibility or interest in participation, all of which are either unobservable to the researcher, or not used as inputs in the simulation of the threshold models.

5.2 Is geographic proximity a good proxy for social connectedness?

While we find that network theory-based targeting statistically increases pit planting adoption, the small absolute value of that increase was not cost effective: the procedure is data intensive and eliciting social network connections in each village is expensive. We anticipated this drawback of the theory-driven approach, and thus included the geography-based treatment arm, which is more feasible for government extension agencies to replicate and scale. Table 5 provides some suggestive evidence that the geography proxy may be able to provide some of the gains in adoption observed under targeting using social network theory, particularly in the medium run. For the adoption rate, the Geo effects are similar in size to the Network-theory treatment effects, but less precisely estimated. We cannot reject that the Geo treatment is the same as the benchmark villages, nor statistically different than the Network villages. The Geo treatment does not perform as well as the network-based treatments in generating a larger *number* of non-seed adopters. The point estimate in season 3 is smaller and statistically different from the network-theory villages (p=.06). In terms of the extensive margin of any adoption in the village, the Geo treatment villages exhibit a statistically significant 18 percentage point increase by season 3 relative to benchmark villages, and this gain is statistically similar to using the data intensive, theory-driven procedures to target.

Table 1 provides some insights into the underlying reasons for these differences. The Geo seed farmers are on average much poorer, and are often in more remote locations in the network of social connections (as indicated by lower eigenvector centrality values in Table 1). The Geo seeds are generally clustered together (since their selection process employed a simulation based on the complex contagion model), and there is some diffusion to their geographic neighbors. This leads to the observed increase in the extensive margin. However, since these seeds are less connected and in a less dense part of the network than are the simple and complex seeds, overall there is a slower pace of diffusion (e.g. to their secondary connections) than in the Network partners treatment.

Overall, the program evaluation results suggest that in future work we will need to develop other simple and inexpensive procedures that can identify individuals who our social network data (combined with theory) chose as seed farmers in order to make network-based targeting more policy relevant and scalable. However, the social network theory-based strategies we employ show promise that they can increase adoption and experimentation with new productive technologies. Moreover, recent evidence indicates that less expensive approaches may well be feasible: Banerjee et al (2014) have shown that in India a simple question like "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. It is also striking that this does not appear to be the process that government extension workers in Malawi follow, even when they are given complete freedom to select seeds. The AEDO-selected seed farmers exhibit lower eigenvector centrality than the seeds selected through our simple or complex contagion based simulations.

6. The Threshold Model and the Nature of Contagion

Having documented the main program evaluation effects, we now return to the theory to generate additional predictions on the specific structure of diffusion under simple versus complex contagion. We then test these implications using both (a) individual level data on network connections and adoption decisions, and (b) experimental variation between the simple and complex contagion models that were employed in different subsets of villages. In the process, we generate evidence on the specific structure of diffusion through social networks, and whether the learning environment for new agricultural technologies in developing countries more closely reflects complex contagion or simple contagion.

6.1 Individual-level analysis

The diffusion process we observe in Table 5 should start out among individuals close to the seeds and then percolate through the rest of the network. We therefore assess whether individuals who are directly connected to trained seed farmers have higher knowledge of pit planting and higher adoption rates. Panel A of Table 6 compares individuals who are connected one or two trained seeds to those who are not connected to any. However, since network position is clearly endogenous, we control for whether an individual is connected to one or two simple, complex or geo (actual or shadow) seeds irrespective of whether those connections were trained on the new technologies. We are therefore controlling for the respondent's network position, and only using variation generated by the experiment. To illustrate, we compare, say, two farmers who are both connected to exactly two 'simple seeds', but where one farmer is in a village randomly assigned to the "simple contagion" treatment (so that his connections were actually trained on the technology), while the other was not. This analysis is conducted using only connections to simple, complex and geo seeds, since we do not observe shadow control seed farmers. The equation we estimate is:

$$Y_{ij} = \alpha + \beta_1 1TSeeds + \beta_2 2TSeeds + \beta_3 1Simple + \beta_4 2Simple + \beta_5 1Complex + \beta_6 2Complex + \beta_7 1Geo + \beta_8 2Geo + \theta_j + \varepsilon_{ijt}$$

Where β_1 and β_2 are the only two variables of interest, $1TSeeds$ is an indicator for the respondent being directly connected to a trained seed farmer and $2TSeeds$ indicates the respondent was directly connected to two trained seed farmers. We interpret the effects of variables associated with β_3 through β_8 as those of control variables that capture the respondent's overall network position with respect to the (actual and shadow) seed farmer links, and these coefficients are omitted from the table.[30] This specification constrains the effect of being connected to trained seeds to be the same across targeting treatments.

Table 6 Panel A shows the above specification by agricultural season for each of two outcomes: adopted pit planting and the heard of pit planting. In season 1, we see no effect of the information targeting on adoption among individuals directly connected to either one or two seeds, relative to those with no connections.[31] However, column (2) shows that in season 1, the training does lead to more information transmission to those directly connected to seeds, and in particular, those who have a direct connection to both the seed farmers who were trained on the technologies. Respondents with two connections are 7.3 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=.02) and even more strikingly, statistically larger than two times the effect of a single connection (p=.057).

This comparison is interesting, because the complex contagion model clusters the seed farmers in one part of the network, while the simple contagion model disperses the dissemination agents. The simple and complex contagion models thus differ sharply in what sort of pattern they suggest across these two parameters: If everyone behaves as though adoption is a simple contagion, we should expect having two connections to seeds to be no more effective than having a single connection. In contrast, the complex contagion model suggests that in the first period, only people

[30] For example, $1Simple$ indicates that the respondent is directly connected to one simple seed while $2Simple$ says that the respondent has connections to two simple seeds. $1Complex$, $2Complex$, $1Geo$ and $2Geo$ are analogously defined for complex seeds and geo seeds respectively.

[31] The control variables show that individuals with certain positions – such as those with one direct connection to a simple seed – are more like to adopt and hear of pit planting even when that seed is not trained on the new technologies. This highlights the importance of using the variation induced by the experiment since unobserved factors, related to one's position in the network or characteristics correlated with it, also affect adoption.

with multiple connections to seeds would be encouraged to adopt. Thus, our empirical results provide support for the hypothesis that learning has complex contagion attributes[32].

The information effect in year 1 translates into an adoption effect in year 2. Column (3) shows that households with two connections to trained seeds are 4.3 percentage points more likely to adopt in the second season than those with no connections, which represents a 76% increase in adoption propensity. Though the point estimate of the effect of 2 connections is more than twice as large as the effect of a connection to one seed (4.3 pp compared to 1.5 pp), though we cannot statistically reject that households with a connection to only one treated seed adopt less frequently (p=.167). We continue to observe (10.3 percentage point) higher awareness of pit planting in season 2 among those with two connections, and can reject that a single connection is sufficient (p = 0.064).

By season 3, however, we no longer see differences in either adoption or knowledge. This may be because the diffusion process has progressed to individuals further from the seeds by the third year. Looking at the means in Panel A, we observe that both the adoption rate and awareness of pit planting has increased among individuals with no direct contacts (to 6.3% and 39% respectively), thus eroding the difference between direct and indirect contacts as information spreads further out from the seeds over time. It is also possible that the smaller sample for the season 3 outcomes limits the precision of these estimates. To distinguish between these alternatives, we further test whether we observe significant differences in connections of path length 2 (i.e., friends-of-friends). The results, shown in Panel B of Table 6, reflect the same specification but use one indicator for whether the household is within 2 path lengths (i.e., friends-of-friends) from a treated

[32] There are some challenges in directly comparing the stark theoretical predictions of the data, most notably that it is hard to compare the timing directly: one agricultural season may represent one or several learning iterations through a network, depending on whether direct observation or simple discussion is sufficient. One which is robust is that if contagion is simple, there should be no additional effect of having a second seed connection, which is why we focus on that dimension here.

seed, controlling for being of path length 2 or less to other shadow seed farmers. The effects on both adoption and knowledge steadily increase over time, and we see that awareness of pit planting among these indirect contacts is significantly higher in season 3 compared to those who are further away from seed farmers.

In summary, analysis using individual-level data demonstrates that the increases in village adoption that we observed in table 5 are driven by individuals who are initially close to the trained seeds. Moreover, individuals who are connected to two seeds are the most likely to benefit from network-based diffusion. While we lack statistical power to definitively determine this, the results are suggestive that having two direct connections – and not just one – is important for diffusion, as would be the case in the complex contagion version of the threshold model. In the next section, we delve deeper into the comparison between simple and complex contagion using our experimental variation.

6.2 The Advent of Diffusion Simple and Complex Contagion

A key feature that distinguishes the complex contagion model from simple contagion is that for many potential pairs of partners, the diffusion process may never start if the learning environment is complex. This is because complex contagion requires agents to cross a higher threshold (we model it as $\lambda=2$ rather than $\lambda=1$ as in simple contagion) before they are convinced to adopt. While we use an average threshold of 2 in the complex contagion model, more generally thresholds above 1 will lead to optimal targeting of partner farmers in the same part of the network (and would also be forms of complex contagion). Broadly speaking, the advent of diffusion (the initial adoption by at least one non-trained farmer) would never happen for many villages and many possible pairs of partnerships, if learning were complex. Perhaps more than anything else, this possibility highlights the importance of understanding diffusion processes.

We generate testable empirical predictions from this theoretical difference between simple and complex contagion by conducting adoption simulations for each of the two models. We then compare our village-level estimation results to these simulations. As our measure of the advent of diffusion, we use "any adoption", an indicator for villages which have at least one household (other than the seeds and shadows) adopting pit planting.

The left part of Figure 4 shows the *predicted* fraction of villages with "Any Adoption" from simulating the model when $\lambda=1$ (Simple contagion) and $\lambda=2$ (Complex contagion) for years 2 and 3.[33] In each case, we separately simulate this by type of type of seed trained. We further adjust our simulated any adoption measure for the random sample nature of our dataset[34]. The right part of Figure 4 shows the empirical counterpart: the actual (observed) values for this variable in the data in years 2 and 3. When the threshold is set to 1 on average (i.e. assuming simple contagion), diffusion is widespread: in year 2, 85% of villages where Geo and Benchmark partners were trained are predicted to have some measured adoption, and that rate goes up to 94% with Simple and Complex partners. The predicted rates of 'any adoption' are even higher in year 3[35].

When we switch to the complex contagion simulation and increase the (median) threshold from one to two, the risk of no adoption increases. Under complex learning, the model predicts that if Simple, Geo or Benchmark partners are trained, then less than half the villages will see any adoption at all in year 2. When complex seeds are trained, 70% of villages experience some

[33] These simulations exclude 12 villages where at least one of the extension worker chosen seeds was not observed in our social network census. This occurred because the spatial boundaries of villages are not always clearly delineated in Nkhotakota. The simulations use the full social network (that we observe) to predict 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 as well, then a random sample may miss all adopters. As the number of adopters increases over time (only in villages which are experiencing diffusion, so holding the extensive margin constant), the random sample is more likely to pick up an adopter and hence the rate of any adoption increases over time in the figure.

[34] Since we only observe a fraction of villagers, our measure of whether anyone in the village is adopting is biased downward. We simulate the sampling procedure in our data to correct for this bias.

[35] In the simulated data, this difference is attributable to a reduction in the sampling bias as adoption becomes more widespread.

adoption. In year 3, training Complex seeds is predicted to increase the fraction of villages with some adoption dramatically relative to training the alternative seeds. 83.5% of villages are predicted to have some adoption in the complex treatments, compared to 47% to 55% if Simple and Geo partners were trained, respectively.

The right side of Figure 4 shows the actual fractions of villages with "any adoption" during years 2 and 3 in our data under the four different experimental arms. The data appears to match the shares of villages with any adoption simulated under complex contagion (i.e. higher threshold) much more closely than those generated under simple contagion in three distinct dimensions. First, the simple contagion simulations suggest that we should observe a much higher fraction of villages with any adoption than is true in the data. Second, simple contagion predicts that the any adoption outcome should not be 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 simulation predicts that the complex partners will maximize the fraction of villages with some adoption, which is exactly what we observe in the data.

6.3 The Performance of Simple versus Complex Contagion Treatments in the Experiment

Although the advent of diffusion is a major difference in the predictions of simple and complex contagion models, the models also generate a variety of other differences. We therefore test whether a broader set of empirical results are consistent with the simple or contagion models. To do so, we use the simulations to predict what the outcome of the experiment would be under a simple learning environment, and also under a complex learning environment. Table 7 presents these simulation results for three different measures of technology adoption: the adoption rate, the total number of adopters, and an indicator for villages with any non-seed adopters. We predict these outcomes for all four experimental arms that were implemented in the field. Table 7 presents regression results using the simulated data, to mirror the regressions that we run with our actual data in the subsequent table.[36] Panel A shows what we should expect to observe across treatments based on simulations of the model with $\lambda=1$ (Simple contagion), and Panel B reports predictions under $\lambda=2$ (Complex contagion).

Columns (1)-(2) show the results for adoption rate outcomes. Complex partners initially maximize adoption in year 2 even if the learning environment is simple, but in year 3 adoption rate is highest when the simple seeds are trained. However, the effects of training simple and complex seeds are not statistically distinguishable (p=.8) for these outcomes simulated under simple contagion. Under simple contagion, villages where the Geo seeds are trained exhibit the lowest adoption rates. Columns (3)-(4) show a very similar set of results for the number of adopters under simple contagion. Taken together, these results indicate that the simple treatment is not expected to dominate alternative targeting strategies even if the contagion process is simple. This reinforces the intuition that if farmers truly have a low threshold for adoption, the diffusion process is not likely to be sensitive to who is initially targeted with information.

In contrast, when we conduct simulations assuming the complex contagion model is correct, the complex treatment is predicted to increase adoption significantly more than all other treatments (Panel B of Table 7). The Complex treatment out-performs the simple, Geo and Benchmark treatments in terms of all adoption outcomes during both years (with statistical tests for differential effects producing p-values below 0.001 for every comparison).

Table 8 displays the corresponding regressions based on actual data from our experiments. Column (1) shows that both simple and complex contagion villages have higher adoption rates as of season 2. Compared to the benchmark rate of 4.4%, complex villages experience 3.5 percentage point higher adoption rate and simple villages experience 2.7 percentage points. We cannot reject that the adoption rates are the same in Simple and Complex villages. The adoption rates in Geo

[36] The table differs from Figure 1 in two key dimensions: (1) this uses the realized randomization and not all villages as in figure 1, and (2) includes additional stratification control variables as in the empirical analysis.

treatment villages are statistically the same as all other villages, though the point estimate is very similar to Complex villages. Column (2) looks at the adoption rate in season 3. We cannot reject that the adoption rate is the same across all treatment types, though the point estimate on Complex remains similar to year 2. The Benchmark villages experienced an increase in the adoption rate from 4.4 percentage points to 7.7 percentage points in year 3. The difference between Benchmark and Simple villages is essentially zero (point estimate of .006, se=.02) and the point estimate of Geo, while very noisy, decreases in magnitude in year 3 (from 3.6pp (se=.026) to .015 (se=.03)).

Columns (3)-(4) look at the number of adopters in the village, where estimated adopters account for sampling weights since we sampled the same number of respondents irrespective of population size. In both season 2 and 3, there are on average an additional 2 adopters in Complex villages, compared to 1.9 adopters in season 2 and 4 adopters in season 3 in Benchmark villages. This represents a doubling in the number of adopters in season 2, when the point estimate is significantly different from zero at conventional levels. Neither Simple nor Geo villages are statistically different from the Benchmark villages in either season, but qualitatively we observe the point estimates in both treatment groups becoming smaller (relative to Benchmark) from season 2 to season 3. In season 3, the number of adopters in Complex villages is statistically higher than in Geo villages. Finally, columns (5)-(6) look at the extensive margin – whether anyone in the village sample adopted pit planting – and finds that in season 2, this rate is significantly higher in Complex villages compared to Benchmark villages, though not significant across the remaining treatment groups. The point estimate on the Simple indicator is 0.158 compared to 0.210 for Complex, suggesting a similar rate of any adoption. In season 3, Simple, Complex and Geo villages all attain a higher rate of any adoption than Control villages, though qualitatively Complex has the highest adoption rate (85% in Complex compared to 73% in Simple and Geo).

Taken together, the data are most consistent with the predictions generated from the complex contagion simulations. First, the individual-level analysis in Table 6 suggests that connections with two seeds, and not just one seed, are important for farmers to adopt pit planting. Second, the identities of the seeds clearly matter, and there are significant differences in adoption effects across the different treatment arms. This is not consistent with the simulations under simple contagion. Finally, the complex treatment leads to more diffusion (in terms of point estimates) than all other treatments across both years. The complex treatment typically results in significantly more adoption than the benchmark treatment, across all three measures of adoption. The higher adoption rates in the complex treatment are also sometimes significantly different from the Geo or Simple Contagion treatments.

Note, however, that our simulations do not predict an unambiguous pattern for these outcomes: Table 7 shows that the number of adopters would be higher in Complex villages than in Simple villages in season 2 under both Simple and Complex contagion. The time trajectory, however, provides suggestive evidence in favor of the Complex contagion model. In the simulations, the gap between Complex and Simple villages becomes larger over time in the Complex contagion simulations while the gap narrows or reverses (for number of adopters and the adoption rate, respectively) over time under Simple contagion simulations. Table 8 shows that the gap between Complex and Simple villages widens in season 3 compared to season 2 for both the adoption rate and the number of adopters. These differences, however, are not statistically significant. The simulations also predict a larger increase in the adoption rate over time in both Complex and Benchmark villages compared to what we observe empirically in season 3. One possible reason for this is additional constraints, other than just information, are binding for farmers.

The final piece of evidence that points to Complex contagion is the 'any adoption' rate. The empirical patterns are at odds with the Simple Contagion simulations: in those simulations, Simple, Complex and Benchmark villages would have similar figures and the rate of 'any adoption' in the Benchmark villages is at least 85%. This is inconsistent with the empirical results. The Complex contagion simulations predict that Complex treatment would have the highest rate of 'any adoption', as we see at least qualitatively in the data. However, the 'any adoption rate' in Benchmark villages is lower in year 3 than we would anticipate from the simulations. On net, the evidence points towards Complex contagion, though the study is lacking some statistical power to provide definitive evidence.

Table 9 shows the results of re-estimating these regressions for the subset of villages that were less familiar with the new technologies at baseline. These are the villages where information failures are more likely to be a deterrent to adoption, and thus the locations where our models are most applicable. The complex treatment exhibits the highest rates of adoption in this sub-sample of villages. Moreover, it is often statistically differentiable from the simple contagion treatment as well as benchmark, and retains statistical significance in all three adoption variables in year 3. For these villages in particular, the results align closely with the pattern of estimates from the simulations, which suggests that complex contagion models may be particularly relevant when technologies are truly novel.

7. **Concluding Remarks**

This paper seeks to understand whether social network theory-based targeting of information to farmers within Malawian villages can improve the diffusion of new agricultural technologies. We develop a methodology to select seed farmers who would maximize village-level adoption in theory on the basis of the linear threshold model of diffusion. By partnering with the Ministry of Agriculture and Food Security, we implemented an empirical counterpart to our model simulations as a randomized controlled trial, in order to test whether theory-driven targeting using detailed social network data can increase technology adoption. We find that adoption rates over three agricultural seasons were greater in villages in which seed farmers were selected using model simulations. We also find promising evidence that an inexpensive proxy of the social network, using geographical proximity rather than elicited network connections, can generate gains in adoption rates over the status quo approach of relying on government extension workers. Finally, our results also strongly suggest that farmers are convinced to adopt a new technology only if they receive information about it from multiple sources. This implies that diffusion follows a Complex Contagion pattern. Future work should explore inexpensive proxies for the theory and data-intensive methodology developed in this paper, in order to make these insights more directly policy-relevant and cost-effective.

References

Acemoglu, Ozdaglar and Yildiz (2011). "Diffusion of Innovations in Social Networks," In IEEE Conference on Decision and Control (CDC), 2011.

Anderson, J, and G Feder (2007). "Agricultural Extension." In *Handbook of Agricultural Economics* 3, pp. 2343-2378.

Apicella, C. L., F. Marlowe, J. Fowler and N. Christakis (2012). "Social Networks and Cooperation in hunter-gatherers," *Nature* 481: 497-502.

aWhere, Inc. 2014. WeatherTerrain® Daily Surfaced Weather Data. aWhere, Inc. 4891 Independence St. Wheat Ridge, CO 80033 USA.

Bandiera, O., & Rasul, I. (2006). "Social networks and technology adoption in northern Mozambique." *The Economic Journal*, 116(514), 869-902.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). "The Diffusion of Microfinance." *Science*, 341(6144). doi: 10.1126/science.1236498

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2014). "Gossip and Identifying Central Individuals in a Social Network." *Working paper*.

Bardhan, P. and C. Udry (1999). *Development Microeconomics*. Oxford University Press.

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

BenYishay, A. and A. M. Mobarak (2014). "Social Learning and Communication," Working Paper, Yale University.

Bursztyn, L., F. Ederer, B. Ferman and N. Yuchtman (2013) "Understanding Mechanisms Underlying Peer Effects: Evidence from a Field Experiment on Financial Decisions," *Econometrica*, Forthcoming.

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

Centola, Damon. 2010. "The Spread of Behavior in an Online Social Network Experiment," Science 329:1194-7.

Chandrasekhar, A. and R. Lewis (2011) "Econometrics of Sampled Networks," Working Paper, Stanford University.

Coleman, James, Katz, Elihu, & Menzel, Herbert. (1957). The Diffusion of an Innovation Among Physicians. Sociometry, 20(4), 253-270.

Conley, T. and C. Udry (2010). "Learning about a New Technology." *American Economic Review,* 100(1), pp 35-69.

Derpsch, R. (2001). "Conservation tillage, no-tillage and related technologies." *Conservation Agriculture, A Worldwide Challenge* **1**:161-170.

--------------. (2004). "History of Crop Production, With & Without Tillage." *Leading Edge*. March: 150–154.

Doumit, G., M. Gattellari, J. Grimshaw, M. A. O'Brien (2007). "Local opinion leaders: Effects on professional practice and health care outcomes." *Cochrane Database Syst. Rev.* 1 CD000125.

Duflo, E., M. Kremer, and J. Robinson (2011). "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review*, 101(6), pp. 2350-90.

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

Foster, Andrew and Mark Rosenzweig (1995). "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture". *Journal of Political Economy* 103(6), pp. 1176-1209.

Grannovetter (1978). "Threshold Models of Collective Behavior," *American Journal of Sociology*, 86 (6), 1420-1443.

Grilliches, Zvi (1957). "Hybrid Corn: An Exploration in the Economics of Technical Change." *Econometrica* 25(4), pp. 501-522.

Haggblade, S. and G. Tembo (2003). "Conservation Farming in Zambia." International Food Policy Research Institute EPTD Discussion Paper #108.

Iyengar, R., Van den Bulte, C., & Valente, T. W. (2011). Opinion leadership and social contagion in new product diffusion. *Marketing Science*, 30(2), 195-212.

Kremer, M., E. Miguel, S. Mullainathan, C. Null and A. Zwane (2011). "Social Engineering: Evidence from a Suite of Take-up Experiments in Kenya" Mimeo, UC Berkeley.

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

Manski, C. (1993). "Identification of Endogenous Social Effects: The Reflection Problem." *The Review of Economic Studies* 60(3), pp. 531-542.

Miller, Grant and A. M. Mobarak (2014). "Learning about New Technologies through Opinion Leaders and Social Networks: Experimental Evidence on Non-Traditional Stoves in Rural Bangladesh." *Marketing Science* forthcoming.

Munshi, Kaivan (2004). "Social learning in a heterogeneous population: Technology diffusion in the Indian Green Revolution." *Journal of Development Economics*, 73(1), pp 185-215.

Munshi, Kaivan (2008). "Social Learning and Development," in Lawrence E. Blume and Steven N. Durlauf, eds. New Palgrave Dictionary of Economics, 2nd Edition. Palgrave Macmillan.

Mwato, I.L., Mkandawire, A.B.C. and Mughogho, S.K. (1999). "Combined Inputs of Crop Residues and Fertilizer for Smallholder Maize Production in Southern Malawi." *African Crop Science Journal* 7 (4): 365-373.

Nkhuzenje, H. (2003). "Contribution of promiscuous and specific soybean variety residues to soil fertility improvement and maize yield under smallholder farms in Zomba District, Southern Malawi." Thesis submitted in partial fulfilment of the requirements of Master of Science degree in agronomy (Soil science).

Nyirongo, J., S. Mughogho, and J. Kumwenda (1999). "Soil Fertility Studies with Compost and Igneous Phosphate Rock Amendments in Malawi." *African Crop Science Journal*, 7(4), pp. 415-422.

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, December.

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

Udry, Christopher (2010). "The Economics of Agriculture in Africa: Notes Toward a Research Program." *African Journal of Agricultural and Resource Economics* forthcoming.

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

Table 1: Seed and Shadow Characteristics by Optimal Treatment

	Wealth Measures		Social Network Measures			
	Farm Size (1)	Total Index (PCA) (2)	Degree (3)	Betweenness Centrality (4)	Closeness (5)	Eigenvector Centrality (6)
Treatment						
Simple	-0.130	0.083	0.371	130.486 **	0.004	0.009
	(0.19)	(0.23)	(1.04)	(63.05)	(0.02)	(0.01)
Complex	-0.008	0.348 *	3.668 ***	125.845 **	0.043 **	0.064 ***
	(0.19)	(0.23)	(1.04)	(62.84)	(0.02)	(0.01)
Geographic	-0.591 ***	-0.766 ***	-3.667 ***	-94.184	-0.029	-0.045 ***
	(0.19)	(0.23)	(1.04)	(63.12)	(0.02)	(0.01)
p-values						
Simple = Complex	0.310	0.069	0.000	0.899	0.002	0.000
Complex = Geographic	0.000	0.000	0.000	0.000	0.000	0.000
Simple = Complex = Geographic	0.000	0.000	0.000	0.000	0.000	0.000
N	1241	1241	1225	1225	1225	1225
Mean of Benchmark Partners	2.04	0.649	12	173	0.476	0.173
SD of Benchmark Partners	2.98	1.7	6.85	347	0.134	0.0973

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 6 additional partner farmers (2 Simple partners, 2 Complex partners, and 2 Geo partners) in all 200 villages.

2 *** p<0.01, ** p<0.05, * p<0.1

Table 2: Seeds vs Counterfactual Farmers

	Adopted Pit Planting						Adopted Crop Residue Management		
	(1)		(2)		(3)		(1)		(2)
Seed	0.251	***	0.221	***	0.168	***	0.134	***	0.032
	(0.03)		(0.03)		(0.04)		(0.04)		(0.05)
N	686		672		488		686		467
Mean of Shadows	0.054		0.093		0.139		0.344		0.228
Season	1		2		3		1		2

Notes

1 Also included are village fixed effects. Sample includes only seed and counterfactual seed farmers and excludes Benchmark villages. Standard errors are clustered at the village level.

Table 3: Seed Farmers

	Adopted Pit Planting			Adopted Crop Residue Management	
	(1)	(2)	(3)	(1)	(2)
Simple	-0.018	0.132 *	0.159 *	0.089	-0.105
	(0.07)	(0.07)	(0.09)	(0.08)	(0.09)
Complex	-0.030	0.036	0.011	0.023	-0.111
	(0.08)	(0.07)	(0.08)	(0.08)	(0.10)
Geographic	-0.105	-0.057	-0.032	0.000	-0.101
	(0.08)	(0.07)	(0.08)	(0.08)	(0.10)
N	342	330	247	342	232
Mean of Benchmark	0.346	0.269	0.246	0.432	0.382
Simple = Complex	0.876	0.190	0.097	0.377	0.935
Complex = Geographic	0.363	0.205	0.581	0.755	0.911
Joint test of 3 treatments	0.584	0.071	0.163	0.585	0.641
Season	1	2	3	1	2

Notes

1 Also included are stratification controls (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 ssquare; and district fixed effects. Only seed farmers are included. Standard errors are clustered at the village level.

Table 4: Conversations about Pit Planting

	with Simple Partner		with Complex Partner		with Geo Partner	
	(1)		(2)		(3)	
Simple	0.046	***	0.019		0.005	
	(0.015)		(0.012)		(0.009)	
Complex	0.019	*	0.036	***	0.000	
	(0.011)		(0.014)		(0.008)	
Geographic	0.003		0.005		0.031	**
	(0.012)		(0.009)		(0.016)	
N	3733		3659		3720	
Mean of Benchmark	0.020		0.026		0.018	
SD of Benchmark	0.139		0.159		0.133	
Test: Simple = Complex	0.075		0.209		0.543	
Test: Complex = Geo	0.139		0.008		0.028	
Test: Simple = Geo	0.004		0.160		0.079	
Season	1		1		1	

Notes

1 Sample excludes seed and shadow farmers.

2 Also included are stratification controls (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; district fixed effects; and controls for the number of partner farmers (of the type asked about in the respective column) we asked about in the questionnaire by including a dummy variable for each number of partner farmers from 0 to 4.

Table 5: Aggregate Pit Planting Adoption

	Adoption Rate for non-seeds			Number of non-seed Adopters			Any non-seed adopters		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Network Treatment	0.011	0.031 **	0.022	0.184	1.711 **	1.350	0.026	0.184 **	0.248 ***
	(0.009)	(0.014)	(0.021)	(0.501)	(0.733)	(1.199)	(0.078)	(0.085)	(0.094)
Geo Treatment	0.018	0.038	0.015	0.419	0.547	-0.715	0.105	0.068	0.188 *
	(0.014)	(0.026)	(0.030)	(0.590)	(0.733)	(1.065)	(0.095)	(0.096)	(0.109)
N	200	200	141	200	200	141	200	200	141
Mean of Benchmark	0.022	0.044	0.077	1.19	1.94	4.1	0.320	0.460	0.543
SD of Benchmark	0.039	0.079	0.107	3.44	3.51	6.25	0.471	0.503	0.505
P value of test: Network = Geo	0.596	0.766	0.760	0.649	0.163	0.061	0.364	0.147	0.500
Season	1	2	3	1	2	3	1	2	3

Notes

1 Network partners are villages where seeds were selected using the threshold model and the social network data. Geographic partners refers to villages where seeds were selected using the threshold model, but where links were proxied by geographic distance instead of direct solitication of social network links.

2 Columns (4)-(6) include sample weights for village size.

3 Also included are stratification controls as listed in Table 4. Seed and shadow farmers are excluded.

4 Test: Network = Geographic shows the p value of the test of whether the effect of the network parnters treatment is different from the geographic partner treatment.

5 Season refers to the number of seasons following the training of seed farmers. Season 1 is 2010 in Mwanza and Machinga, and 2011 in Nkhotakota. Column (3) includes only villages in Mwanza and Machinga as we have 3 seasons of data only for those two districts.

Table 6: Individual-level analysis of Pit Planting Decisions

	Season 1		Season 2		Season 3	
	(1)	(2)	(3)	(4)	(5)	(6)
	Adopted PP	Heard of PP	Adopted PP	Heard of PP	Adopted PP	Heard of PP
Panel A: Direct connections						
Connected to one seed	0.010	-0.007	0.014	0.033	0.009	0.017
	(0.011)	(0.023)	(0.015)	(0.024)	(0.016)	(0.030)
Connections to two seeds	0.018	0.072 *	0.039 **	0.104 ***	0.019	0.069
	(0.014)	(0.037)	(0.020)	(0.040)	(0.034)	(0.064)
N	4207	4155	3937	4538	3000	3105
Mean of Excluded Group	0.023	0.214	0.056	0.274	0.064	0.387
SD of Excluded Group	0.151	0.41	0.23	0.446	0.244	0.487
Test: 2 connections = 1 connection	0.544	0.020	0.213	0.064	0.788	0.405
Test: 2 connections = 2*one connection	0.920	0.055	0.709	0.442	0.996	0.640
Panel B: Two Path Length Connections						
Is within 2 path length of a seed	0.015 *	-0.019	0.025 **	0.025	0.038 *	0.068
	(0.009)	(0.030)	(0.012)	(0.027)	(0.020)	(0.042)
N	4207	4155	3937	4538	3000	3105
Mean of Excluded Group	0.013	0.227	0.044	0.257	0.043	0.380
SD of Excluded Group	0.113	0.419	0.206	0.437	0.203	0.486

Notes

1 Sample excludes seed and shadow farmers in all village. Seed farmers are either simple, control or geo (no benchmark farmers included).

2 In panel A, 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 and two Geo partners.

3 In panel B, additional controls include indicators for the respondent is: within 2 path length of a Simple partner, within 2 path length of a Complex partner, and within 2 path length of a Geo partner.

4 Also included in both panels are village fixed effects.

5 The excluded group in Panel A is comprised of individuals with no connections to a seed farmer.

6 The excluded group in Panel B is comprised of individuals who are not within a 2 path length of either seed.

Table 7: Simulation of Complex and Simple Contagion

	Simulated Adoption Rate		Simulated Number of Adopters		Simulated Any Adopters	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Simulations Assuming Farmers learn by Simple Contagion						
Simple Treatment	0.089 **	0.070	6.350	5.548	0.033	0.015
	(0.045)	(0.065)	(4.769)	(9.546)	(0.032)	(0.034)
Complex Treatment	0.066	0.001	3.486	-0.925	-0.001	-0.008
	(0.054)	(0.074)	(4.898)	(8.883)	(0.040)	(0.043)
Geo treatment	-0.092 *	-0.100	-7.931 *	-7.802	-0.063	-0.085 *
	(0.051)	(0.068)	(4.328)	(8.440)	(0.047)	(0.051)
Year	2	3	2	3	2	3
N	186	138	186	138	186	138
Mean Benchmark Partners	0.517	0.706	36.2	59.3	0.935	0.953
SD Benchmark Partners	0.301	0.322	21.1	42.4	0.175	0.165
Test: Simple = Complex	0.587	0.294	0.532	0.486	0.363	0.567
Test: Complex = Geo	0.002	0.141	0.007	0.417	0.202	0.152
Test: Simple = Geo	0.000	0.003	0.000	0.134	0.016	0.035
Panel B: Simulations Assuming Farmers Learn by Complex Contagion						
Simple Treatment	-0.032	-0.077	-3.571 **	-8.934 **	-0.102	-0.081
	(0.036)	(0.058)	(1.606)	(3.784)	(0.066)	(0.080)
Complex Treatment	0.132 ***	0.190 ***	6.318 ***	12.784 **	0.224 ***	0.263 ***
	(0.040)	(0.067)	(2.159)	(5.250)	(0.070)	(0.083)
Geo treatment	-0.006	-0.078	-2.623	-5.828	-0.042	-0.045
	(0.037)	(0.061)	(1.631)	(4.007)	(0.071)	(0.085)
Season	2	3	2	3	2	3
N	187	138	187	138	187	138
Mean Benchmark Partners	0.151	0.277	7.91	17.1	0.566	0.563
SD Benchmark Partners	0.197	0.324	8.94	18.8	0.39	0.398
Test: Simple = Complex	0.000	0.000	0.000	0.000	0.000	0.000
Test: Complex = Geo	0.001	0.000	0.000	0.000	0.000	0.000
Test: Simple = Geo	0.438	0.979	0.384	0.261	0.297	0.621

Notes

1 Only Includes Control Villages where we had both Seeds in Census.

Table 8: Simple and Complex Learning in Pit Planting

	Adoption Rate				Number Adopters				Any Non-Seed Adopters			
	(1)		(2)		(3)		(4)		(5)		(6)	
Simple Treatment	0.035	**	0.006		1.041		0.434		0.158		0.189	*
	(0.017)		(0.022)		(0.747)		(1.297)		(0.101)		(0.111)	
Complex Treatment	0.027	*	0.038		2.369	**	2.231		0.210	**	0.304	***
	(0.016)		(0.026)		(1.172)		(1.716)		(0.095)		(0.101)	
Geo treatment	0.038		0.015		0.540		-0.726		0.068		0.188	*
	(0.026)		(0.030)		(0.736)		(1.071)		(0.096)		(0.110)	
Year	2		3		2		3		2		3	
N	200		141		200		141		200		141	
Mean of Benchmark	0.044		0.077		1.940		4.100		0.46		0.543	
SD of Benchmark	0.079		0.107		3.510		6.250		0.503		0.505	
Test: Simple = Complex	0.684		0.177		0.313		0.341		0.581		0.240	
Test: Complex = Geo	0.670		0.442		0.142		0.077		0.113		0.220	
Test: Simple = Geo	0.898		0.723		0.552		0.331		0.352		0.990	

Table 9: Any non-seed Adopters: Actual Results if Less than Median baseline familliarity with Pit Planting (<0.0432 ever tried)

	Adoption Rate		Number of Adopters		Any Non-Seed Adopters	
	(1)	(2)	(3)	(4)	(5)	(6)
Simple Treatment	0.041	0.014	0.943	1.019	0.155	0.312 **
	(0.03)	(0.02)	(1.32)	(1.38)	(0.153)	(0.151)
Complex Treatment	0.037	0.098 ***	4.142 *	5.745 **	0.254 *	0.458 ***
	(0.03)	(0.03)	(2.10)	(2.37)	(0.138)	(0.131)
Geo treatment	0.0221	0.0457	0.3101	1.4679	0.047	0.350 **
	(0.03)	(0.03)	(1.22)	(1.62)	(0.145)	(0.153)
Season	2	3	2	3	2	3
N	99	82	99	82	99	82
Mean of Benchmark	0.0396	0.0526	1.85	2.86	0.458	0.45
SD of Benchmark	0.093	0.093	3.810	5.450	0.509	0.51
Test: Simple = Complex	0.902	0.011	0.185	0.081	0.440	0.279
Test: Complex = Geo	0.617	0.187	0.097	0.122	0.096	0.428
Test: Simple = Geo	0.539	0.265	0.625	0.732	0.429	0.800

1 The sample is restricted to villages where less than 4.32% of households (the median) ever tried pit planting at baseline.

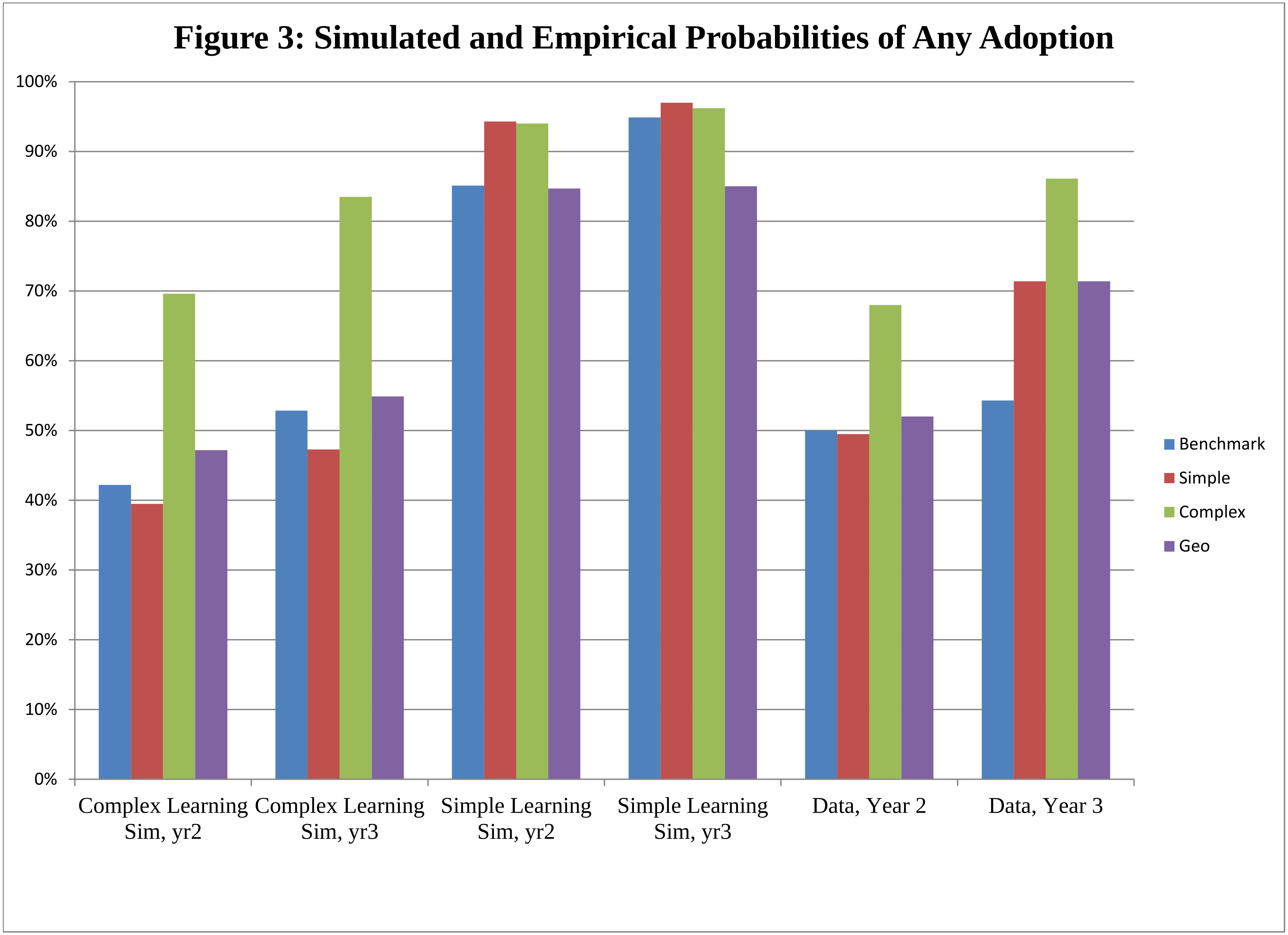
Figure 3: Simulated and Empirical Probabilities of Any Adoption
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Complex Learning Sim, yr2
Complex Learning Sim, yr3
Simple Learning Sim, yr2
Simple Learning Sim, yr3
Data, Year 2
Data, Year 3
Benchmark
Simple
Complex
Geo

Table A1: Balance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
VARIABLES	Housing[1]	Assets[1]	Livestock[1]	Basal fertiliser (kg)	Top dressing fertiliser (kg)	Pit planting	# of Adults	# of Children	Farm size (acres)	Own land	Yields	Provided Ganyu	Used Ganyu
Benchmark	-0.297	-1.128	-0.205	50.73	61.99	-0.00479	2.280	1.212	1.683	0.934	-15.59	0.157	0.0183
	(0.283)	(0.191)	(0.207)	(13.40)	(12.04)	(0.00724)	(0.0768)	(0.0781)	(0.262)	(0.0387)	(51.92)	(0.0640)	(0.0337)
Simple Treatment	-0.625	-1.250	-0.269	50.69	59.45	-0.00581	2.271	1.246	1.466	0.930	-17.82	0.182	-0.00779
	(0.311)	(0.204)	(0.223)	(15.31)	(13.24)	(0.00718)	(0.0783)	(0.0797)	(0.271)	(0.0415)	(53.49)	(0.0631)	(0.0359)
Complex Treatment	-0.469	-1.221	-0.250	52.44	59.04	-0.00417	2.285	1.241	1.516	0.932	-31.99	0.181	0.00331
	(0.293)	(0.195)	(0.214)	(13.99)	(12.21)	(0.00704)	(0.0790)	(0.0750)	(0.273)	(0.0419)	(51.48)	(0.0592)	(0.0352)
Geo Treatment	-0.416	-1.209	-0.343	50.26	60.64	-0.00614	2.272	1.233	1.622	0.929	-20.21	0.166	0.0234
	(0.322)	(0.205)	(0.209)	(13.10)	(12.27)	(0.00745)	(0.0773)	(0.0753)	(0.279)	(0.0419)	(51.14)	(0.0653)	(0.0354)
Observations	14,089	14,346	14,346	10,427	10,526	14,079	14,103	14,090	14,083	14,346	13,500	14,078	14,078
Control = Simple	0.004	0.108	0.423	0.993	0.443	0.548	0.805	0.219	0.00792	0.830	0.898	0.188	0.0186
Control = Complex	0.193	0.176	0.539	0.664	0.328	0.708	0.875	0.354	0.0572	0.947	0.390	0.271	0.137
Control = Geo	0.545	0.281	0.0539	0.891	0.638	0.439	0.839	0.425	0.550	0.808	0.792	0.717	0.707
Simple = Complex	0.151	0.696	0.782	0.708	0.896	0.303	0.618	0.890	0.555	0.854	0.390	0.944	0.277
Simple = Geo	0.325	0.624	0.270	0.927	0.703	0.835	0.996	0.645	0.123	0.927	0.867	0.493	0.0258
Complex = Geo	0.807	0.874	0.105	0.575	0.568	0.223	0.709	0.796	0.320	0.842	0.483	0.559	0.121
Joint	0.0312	0.376	0.190	0.950	0.775	0.618	0.960	0.623	0.0456	0.992	0.806	0.554	0.0518

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A2: Yields

	(1)		(2)		(3)	
Estimation	OLS		OLS		IV	
Adopted PP					0.449	**
					(0.225)	
Seed	0.109	**	0.538	***		
	(0.054)		(0.140)			
Total precipitation over season (mm)			0.000	***		
			(0.000)			
Seed X total precipitation			-0.001	***		
			(0.000)			
Observations	1119		1119		1119	

No Notes

1 1 All columns include district and season FE and controls for total farm size, village size, and village baseline usage of fertilizer, composting and pit planting. The sample includes only seeds and shadows and excludes Benchmark villages.

2 Robust standard errors clustered by village in parentheses.

Table A3: Aggregate CRM Adoption

	Adoption Rate for non-seeds		Number of non-seed Adopters		Any non-seed adopters	
	(1)	(2)	(3)	(4)	(5)	(6)
Network Treatment	-0.013	-0.021	0.541	-0.410	-0.044 **	-0.062 **
	(0.025)	(0.022)	(1.645)	(1.503)	(0.021)	(0.030)
Geo Treatment	-0.001	-0.042	-0.226	-2.066	-0.044	-0.093 *
	(0.032)	(0.029)	(1.674)	(1.641)	(0.030)	(0.049)
N	200	141	200	141	200	141
Mean of Benchmark	0.308	0.227	14	12.1	1.000	1.000
SD of Benchmark	0.217	0.105	12.1	11.1	0.000	0.000
P value of test: Network = Geo	0.688	0.468	0.624	0.315	0.998	0.554
Season	1	2	1	2	1	2

s

Notes

1 Network partners are villages where seeds were selected using the threshold model and the social network data. Geographic partners refers to villages where seeds were selected using the threshold model, but where links were proxied by geographic distance instead of direct solitication of social network links.

2 Columns (4)-(6) include sample weights for village size.

3 Also included are stratification controls as listed in Table 4. Seed and shadow farmers are excluded.

4 Test: Network = Geographic shows the p value of the test of whether the effect of the network parnters treatment is different from the geographic partner treatment.

5 Season refers to the number of seasons following the training of seed farmers. Season 1 is 2010 in Mwanza and Machinga, and 2011 in Nkhotakota. Column (3) includes only villages in Mwanza and Machinga as we have 3 seasons of data only for those two districts

Table A4: Individual-level analysis of CRM Decisions

	Season 1		Season 2	
	(1)	(2)	(3)	(4)
	Adopted CRM	Heard of CRM	Adopted CRM	Heard of CRM
Panel A: Direct connections				
Connected to one seed	-0.018	-0.006	-0.019	-0.011
	(0.023)	(0.029)	(0.032)	(0.025)
Connections to two seeds	-0.024	0.034	0.017	-0.050
	(0.041)	(0.045)	(0.050)	(0.046)
N	3220	3183	2041	3444
Mean of Excluded Group	0.259	0.613	0.182	0.637
SD of Excluded Group	0.438	0.487	0.386	0.481
Test: 2 connections = 1 connection	0.896	0.340	0.398	0.330
Test: 2 connections = 2*one connection	0.807	0.417	0.329	0.563
Panel B: Two Path Length Connections				
Is within 2 path length of a seed	0.007	-0.004	-0.019	0.005
	(0.039)	(0.041)	(0.043)	(0.041)
N	3220	3183	2041	3444
Mean of Excluded Group	0.203	0.608	0.174	0.611
SD of Excluded Group	0.402	0.488	0.38	0.488

Notes

1 Sample excludes seed and shadow farmers in all villages, and excludes control villages. Seed farmers are either simple, control or geo (no control farmers included).

2 In panel A, 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 and two Geo partners.

3 In panel B, additional controls include indicators for the respondent is: within 2 path length of a Simple partner, within 2 path length of a Complex partner, and within 2 path length of a Geo

4 Also included in both panels are village fixed effects.

5 The excluded group in Panel A is comprised of individuals with no connections to a seed farmer.

6 The excluded group in Panel B is comprised of individuals who are not within a 2 path length of either seed.

Table A5: Simple and Complex Learning in CRM

	Adoption Rate	Number Adopters	Any Non-Seed Adopters	
	(1)	(2)	(3)	
Simple Treatment	-0.021	-1.228	-0.070	
	(0.026)	(1.580)	(0.043)	
Complex Treatment	-0.022	0.376	-0.054	
	(0.027)	(2.153)	(0.040)	
Geo treatment	-0.042	-2.076	-0.093	*
	(0.029)	(1.649)	(0.049)	
Year	2	2	2	
N	141	141	141	
mean	0.227	12.100	1	
sd	0.105	11.100	0	
Test: Simple = Complex	0.950	0.487	0.787	
Test: Complex = Geo	0.552	0.283	0.525	
Test: Simple = Geo	0.492	0.618	0.694	