MSc Thesis

International Development Studies

Specialization: Economics of Development

Social Networks and Technology Diffusion in DRC

Paul Hofman[a 1]

901114350110

[a] Development Economics Group, Wageningen University

27 January 2016

Course Code: DEC-80439
Thesis Supervisor: Maarten Voors

Abstract: We investigate if incorporating social network characteristics in the selection of lead farmers can improve adoption in the farmer field school approach. We implement a randomized control trial where we select network entry points based on their eigenvector centrality. Communities were randomly assigned a treatment of central lead farmers or isolate lead farmers. We then estimate the causal effect of centrality on uptake, knowledge, spread, willingness to pay and speed of diffusion of an agricultural technology. We find modest effects. Isolate lead farmers distribute their resources faster than more central lead farmers. We find no treatment effect for uptake, willingness to pay, knowledge or diffusion. We also find that adoption and knowledge of the intervention taper off strongly as it moves through the community. Third-hand receivers are no better off than non-receivers. In addition, we find that people prefer to share goods with people that live closer to them, are in their social network or are poorer than them. This is evidence of charitable behaviour. Our findings suggest that centrality is not very useful to improve efficiency of farmer field schools. Using more entry points would reduce attenuation of the intervention. This should increase social learning and uptake of novel agricultural techniques.

[1] Acknowledgements: this master thesis was completed as part of a larger research project with Maarten Voors (WUR), Martha Ross (WUR), Peter van der Windt (NYU, WUR), Jenn Larsson (NYU) and Neelanjan Sircar (CPR Delhi). They contributed to the research design, and aided in the analysis approach. I contributed to the design and was responsible for the implementation, analysis and writing. Special thanks to John Quattrochi, together with whom I collected the data in DRC. Great thanks to all the local enumerators that helped collect the data. Thanks also to Remi van Balen and Kevin Grieco for excellent feedback. Finally and most importantly, thanks to my family for their invaluable support.

1. Introduction

Development organizations strive to distribute limited resources to a wide audience. An efficient distribution of resources is critical for development in isolated communities to compensate for high transport and distribution costs. Furthermore, uptake of novel agricultural techniques is often low, as farmers are generally risk averse. A popular technique for many agriculture-focused aid programs is to group potential recipients in so called farmer field schools. The core idea is that so called 'lead farmers' are identified and trained in improved agricultural practices. These farmers are then responsible to spread this new knowledge, as well as to distribute further any inputs/goods provided by the development agency (e.g. Feder et al, 2003). There are two main advantages to this technique: 1) Distribution costs are lower, as only a few farmers need to be trained, and 2) the uptake of novel agricultural techniques is higher if communicated by (influential) peers. The effectiveness of this approach is thus closely linked with the size of the lead farmers' social network.

It is likely that the social network of a farmer can be useful to determine who is the optimal lead farmer. That is, the farmer that causes the highest uptake and diffusion. The social network literature uses the concept of centrality to find the most influential people in a network. Bonacich's (2007) eigenvector centrality calculates what person is best at achieving attitudinal change. It incorporates all the paths that information can flow through in a network, unlike other centrality measures. Borgatti (2005) argues that it is the best centrality measure to identify the most influential members of a network. However, no empirical research has used eigenvector centrality to calculate the optimal lead farmer. Earlier papers that use social network data to improve diffusion use other measures of centrality (Banerjee et al, 2013) or calculate the optimal lead farmer using a diffusion model (Beaman et al, 2015). Our design explicitly examines the effect of eigenvector centrality on the wideness and speed of distribution, knowledge retention, willingness to pay and adoption. We can thus test whether incorporating centrality in the lead farmer selection process can improve the efficiency of the farmer field school approach. Besides the social network effect, we also gain insight into the optimum number of network entry points, by examining the attenuation of the intervention impact. We also examine what factors determine an exchange relationship in the farmer field school approach.

Data for this project was collected between February and April 2015. We mapped full social networks in 40 remote villages in Eastern DRC. Examining full social networks is unique, as almost all previous studies examined partial networks where randomly selected members are interviewed. This can lead to bias in the social network map. In these villages of around 65 households, we trained a subset of village members (lead farmers) in the use of a chemical fertilizer. We then provided ten kilograms of fertilizer to each of these entry points. Chemical fertilizer is useful because it can improve low yields. The Congolese wars have left the area devoid of agricultural investments for many years. Chemical fertilizer can contribute to increasing farmers' yields. These lead farmers were tasked with spreading the knowledge gained from the training and the fertilizer throughout the village. Specifically, we asked them to train more famers in the village as lead farmers. This allows the intervention to spread even further. We randomly assigned villages to treatments of either lead farmers with a high eigenvector centrality (Centrals) or lead farmers with a low eigenvector centrality (Isolates). Our first research question is thus '*Does the eigenvector centrality of the lead farmer affect the dissemination effect?'* We collected extensive post-intervention data on the dissemination process within each village. This allows us to examine the attenuation of the intervention. Our second research question is: '*Does the impact of an intervention taper off as it diffuses away from the lead farmer?'* The lead farmers were free to distribute to whomever they wanted. By examining individual characteristics and relational data we can try to predict who lead farmers will distribute. Our third research question is: *'Can we predict distribution behavior?'*

We find moderate effects of varying the eigenvector centrality of the lead farmers. Isolate lead farmers distribute their resources faster than central lead farmers. This might be because they do not need a lot of time to decide whom to distribute to. By definition they should be less connected, easing their decision process. We find no treatment effect for uptake or knowledge. It might be that eigenvector centrality is not so well suited to predict attitudinal change. Earlier papers found larger effects of incorporating social network data (e.g. Banerjee et al (2013) and Beaman et al (2015)). This lack of results is surprising. Eigenvector centrality was specifically designed to predict which person in a network is best at achieving attitudinal change – the goal of our intervention. This results are not driven by the fact that either group of lead farmers were non-compliant with our instructions (e.g. whether they decided to keep all the fertilizer for themselves). Low power does not drive our results either: the impact of the treatment is low but precise. We also find evidence that our social network measures are valid, so the centrality measure should be able to effectively identify the most influential individuals in the network.

We clearly see that the magnitude of the impact of the intervention tapers off as it diffuses away from the lead farmer. This is the case for both uptake and knowledge. Third-hand receivers are no better off than non-receivers. This suggests that relying on more network entry points will increase uptake and knowledge retention. We also see several factors that predict who lead farmers decide to share with. Lead farmers prefer to share with people that live closer to them, that are family, that are field neighbors or that they discuss agriculture with. This is shows that diffusion follows existing (informational) exchange relationships. Lead farmers also prefer to give to people that are poorer than them. This is evidence of charitable behavior. These results are useful to predict how an intervention will spread within a village.

This research was preregistered. Our analyses and hypothesis before examining the data, but after data collection, can be found here: http://egap.org/registration/1636. Where we deviate from this plan we mention this using footnotes. We report no large deviations from our pre-analysis plan.

The rest of this thesis is organized as follows. Section 2 briefly reviews several earlier studies related to social networks and Technology Diffusion. Section 3 provides a detailed description of the study design. Section 4 contains the results, followed by robustness analysis in section 5. Section 6 provides a short discussion on problems that remain, and concludes.

2. Agricultural Development and Social Networks

Agricultural yields in Africa are low. Table 1 shows FAO data for cereal yield per hectare. The yield for cereals in Africa is half that of the next lowest continent, Asia. Low yields are associated with food insecurity and increased poverty levels. Historically, the main approach to increase yields is by introducing new technologies. An example is the green revolution which consisted of the introduction of hybrid seeds and fertilizer. This approach led to huge rates of growth of agricultural production in Asia (Evenson & Gollin, 2003). To improve efficiency over traditional extension agencies, the more participatory Farmer Field School approach has gained popularity in recent years. Then, the responsibility for information transmission is moved to local farmers, dubbed lead farmers. Besides expected higher uptake because information is communicated by peers, it also leads to significant cost savings for the diffuser of knowledge. This is generally the government or an NGO. (Feder, Murgai & Quizon, 2003)

Table 1: Agricultural Production

	Cereals Yield (Kg/Ha)
World	32,343
Americas	52,241
Europe	36,767
Asia	30,273
Africa	14,678

Data from 2010. Source: FAOStat

One important technology that can increase yields is chemical fertilizer. There is considerable discussion on the effectiveness of fertilizer in increasing profits (Beaman et al, 2013; Duflo, Kremer & Robinson, 2008). The theoretical increases in yields are impressive, however. Especially when soil fertility is low it can have long-lasting positive benefits. While fertilizer use is not complicated, instructions on proper usage are required. This makes Farmer Field Schools an ideal method to distribute fertilizer. So which lead farmer maximizes the diffusion in a village?

One way to approach this problem is by examining social networks. Social Network analysis uses knowledge of nodes (e.g. people, firms) and edges (e.g. friendship, trade relations) between these nodes to examine patterns of diffusion and dependence. In the context of a village, edges can represent which node (a villager) communicates with what other nodes. It can be expected that technology diffusion will then follow social network lines (Durlauf, 2004).

The concept of centrality in social networks uses social network data to determine the most important person in a network. For each node in a network the centrality can be calculated based on the pattern of edges and nodes. There are many popular measures of centrality[2]. Borgatti (2005) states that each centrality measure makes implicit assumptions about the manner in which traffic flows in a network. For studying attitudinal change, Bonacich's (2007) eigenvector centrality is most appropriate since it assumes that information does not necessarily follow the most direct path in a network. Eigenvector centrality not only counts the number of edges leading to and from each node, but also weighs each edge based on the importance of the node it originates from. It thus incorporates the entire structure of the network to decide the most influential person. The person with the highest eigenvector centrality is thus the person most effective at achieving attitudinal change, and is therefore the ideal lead farmer. The least ideal lead farmer is the person with the lowest eigenvector centrality. Other centrality measures assume that information must travel along the shortest path possible, or that information cannot reach a person twice. So far, eigenvector centrality and its applications has been mostly a theoretical affair. To our knowledge, it has not been used in an experimental design. We are

[2] Examples are degree centrality, betweenness centrality and closeness centrality.

therefore uncertain of its causal effect. In this study we look for the causal effect of eigenvector centrality on technology diffusion and attitudinal change. Our expectation is that diffusion, uptake and knowledge retention will be higher for the central ambassadors, as the theory predicts.

There is a growing literature on how social networks, and an individual's position within a network, matter for development outcomes. Experimental studies looking at the causal effect of interventions over social networks include Banerjee et al. (2013), Cai (2012) and Oster and Thornton (2008), amongst others. Each study, though unique, relies on a similar methodology. First, the social network of the population of interest is mapped, either fully or partially. Second, a random subset of individuals receives an intervention (often information). Third, individuals in the network are interviewed. To examine diffusion of knowledge, researchers employ regression analysis to estimate the relationship between the outcome of interest and the level of indirect exposure to the intervention (knowledge) through social networks. This exposure is typically proxied as the number of friends that received the intervention. As expected, greater indirect exposure to the intervention typically leads to higher adoption. This is generally called peer effects or neighborhood effects. Two issues should be highlighted with this general approach. First, this procedure demonstrates that information flows explicitly through the measured network, but it doesn't establish that information flows in a manner that is strongly correlated to the network structure. Second, this empirical strategy may result in biased estimates unless the researcher properly accounts for randomization probabilities (Aral et al, 2009). The first point is not a problem because our measure of centrality accounts for alternative paths in information diffusion. Because we use full networks, unlike most other network studies, the second point is also not a problem.

There are several studies that look for the existence of peer effects. Cai (2012) uses the above described approach in rural China, where the intervention is a briefing on an agricultural insurance scheme (in conjunction with insurance contracts offered). He finds that respondents with a high number of friends who were exposed to the first intensive training session are more likely to take up the offered insurance. Oster and Thornton (2012) distribute menstrual cups (a new technology for the area) to randomly selected girls across four schools in Nepal. They found that having many friends who have the cup matters most as it increased the chance of successful adoption. One major difference of note between our study and the study of Oster and Thornton's (2012) is the amount of learning required for successful adoption of the technology involved. Menstrual cups have a much higher level of learning required compared to chemical fertilizer.

Bandiera and Rasul (2006) look at adoption of sunflower cultivation in Mozambique. They found that the probability of adoption increased with the number of family or friends who also adopted, up to a threshold point. After this threshold, more family and friends who adopted actually *decreased* the probability of adoption. They hypothesize two factors that are believed to influence the decision to adopt. First, there is the learning by doing effect: a farmer will become more proficient with a new crop the longer he uses it. This acts as an incentive to adopt early. Second is the learning from others effect. If there are more people in the network that have already adopted, then further adoption becomes easier and less risky (as the technology is proven). This acts as an incentive to adopt late and to wait for more people to adopt the technology.

These papers all conclude the same thing: peer effects matter for technology adoption. Taken together, we can conclude that apparently information regarding technology diffusion travels along social network lines, and social networks can be used to predict and optimize diffusion.

Two papers have done exactly this. Banerjee et al (2013) collect network data (along twelve dimensions) before the implementation of a microfinance program in villages in India. The implementing microfinance organization invited several key respondents to an informational meeting, and then asked these individual to pass on the program information to the rest of the village. Using data on final participation in the program, they develop a model that predicts participation based on network connections. They validate their model by predicting final participation based on initial seeds. They find that their model can accurately predict participation in the microfinance program.

There is only one paper (to our knowledge) that like us uses social network data to *select* initial seeds. Beaman et al (2015) study agricultural practices and network characteristics. They collected full network data in 200 villages in Malawi on three relationships classifications: consultation on agricultural decisions, food sharing, and friendship. Like us, they use social networks to determine the optimal lead farmers in a village, for an agricultural technique (pit planting and crop residue management). They simulate who the optimal lead farmer is using three models based on their social network data. They randomly assigned these three treatments to villages, and compare them against the status quo of lead farmers being selected through consultations with village leaders. Their results show that all three models are more effective in causing adoption in the villages than the status quo.

We add to this literature in several ways. First, we test a well-known measure of centrality in an experimental setting. Second, we use full networks for social network calculations, which is very rare. Third, we can gain insight into how a technology diffuses within a community. Finally, we can see how far a technological innovation can spread before losing effectiveness. While we make no claims to external validity, these questions apply in all isolated communities.

3. Experimental design

3.1 Context

We collect data from 40 remote villages in eastern DRC[3]. We focus on small villages with population maximums of 100 households. We selected small villages because in larger villages collecting full social network data would take too much time. Figure 1 shows the approximate locations of the villages, as well as altitude lines and the major regional city Bukavu. Eastern Congo has been embroiled in violent conflict for several decades, in the First and Second Congo war. Hostilities remain up to this day. Due to the conflict, developments in farming have been slow for many years. Yields are low and hardly improving, causing malnutrition. Incomes from cash crops are low and most farmers struggle to improve their livelihoods. Soil fertility is low. Chemical fertilizer can greatly improve yields in these circumstances, leading to improvements in nutrition and income.

Furthermore, it makes sense to examine social networks at the village level here, because social support derives mostly from other villagers. Formal forms of insurance and state support are rare or non-existent. Most of the agricultural committees and groups are also formed within a village. Delivering resources and training is very costly, because villages are hard to access. Furthermore, as farmers spend a great deal of their time away from the village when they are farming, villagers are hard to reach. A farmer field school approach is effective in this context because it can save on transport and distribution costs. And, as only some farmers need to be trained not a lot of time is wasted looking for missing people. Several NGOs have used the farmer field school in Eastern DRC for these reasons. Therefore, we thought it appropriate to mimic this approach for our intervention.

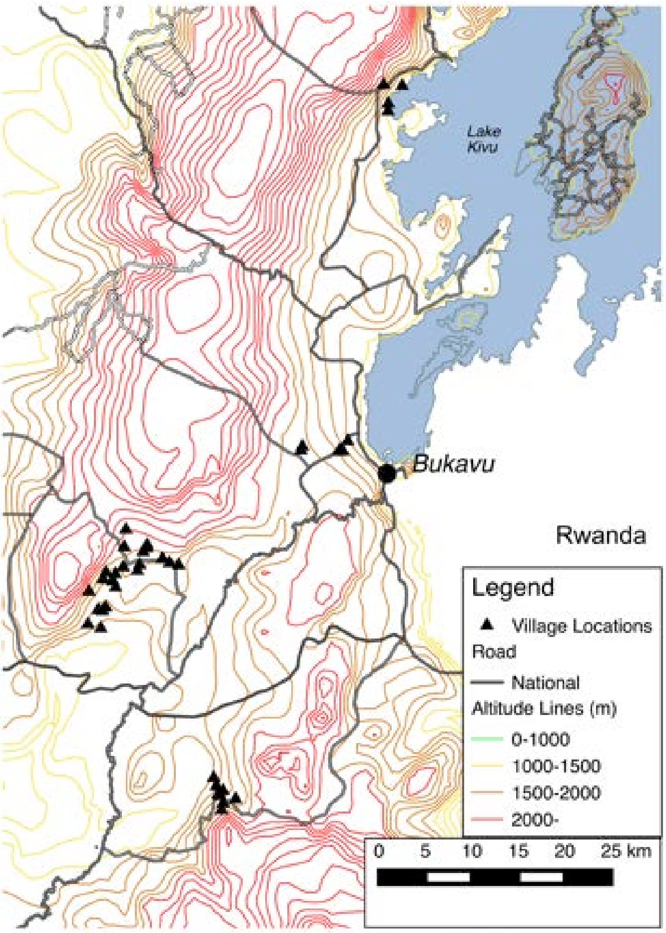

Figure 1: Village Locations

[3] We did not have any priors on sample size. We were unable to conduct power calculations as completing a baseline survey in advance was unfeasible.

3.2 Data

Each village was visited three times: visits A, B and C. The table below gives a brief overview of the research activities conducted during each of these visits.

Visit	Day	Activity
A	1	• Survey: mapping household-level networks • Chief survey to obtain village level information
B	31	• Train ambassadors and distribute technology
C	45	• Track down physical spillovers • Survey: knowledge surrounding technology (cognitive spillovers)

In the first visit (Visit A) research assistants collected data on all heads of households. In this visit we were able to reach 97% of all heads of households, much higher than the 80% collected by Beaman et al (2015) who also mapped full networks. As Aral et al (2009) point out collecting data on a subset of the network introduces bias in the social networks calculations. We therefore tried hard to interview nearly all members of the social network. If the head of household was not available to be interviewed, we returned for another visit, preferably in the weekend as respondents were more likely to be present then. If the respondent remained unreachable, we looked for a replacement, generally the wife of the head of household.

The survey in visit A consisted of two sections. The first section collected basic characteristic and socio-economic information on the individual and their respective household. In the survey we collected data on demography, conflict history, agricultural/production information and knowledge in fertilizer use and optimal application. We collected social network data along three dimensions:

- Blood Family: whether the head of the household is biological family with the other head of the household[4]
- Field neighbors: whether the head of the household's fields border the other head of the household's fields.
- Discuss agriculture: whether the head of the household discusses agricultural related issues with the other head of the household.

We opted to measure social network across fewer dimensions than some earlier studies (e.g. Banerjee, 2013), to prevent survey fatigue. The survey took around 1 hour. The networks we selected are important for several reasons. First, they tie in closely with our research questions of agricultural information & input flows. Second, villagers use family as an informal form of insurance. It is likely that villagers distribute to their family members and expect this favor to be returned at some other point in time. Third, villagers can most easily learn of new agricultural techniques by observing it on another farm. They are mostly likely to see this at a farm they see often: that of their field neighbor. Fourth, new agricultural techniques can also come up when discussing agriculture with other farmers. Finally, an early pilot with more networks showed that these three dimensions were the most distinct from one another and thus captured maximum variation while minimizing the number of network survey questions.

To elicit social network data, we used a household roster that was constructed upon arrival to the village. This roster contained the full names and ages of all heads of household. It also listed if there were any other adults present in the household (for example, the wife of the head of household, or adult children). The research assistant first explained the network under study, and then moved down the village roster asking for each

[4] Specifically, we use whether the other person is biologically related to a maximum of the third degree (this is a well-understood term in Congo). This does not include the wife's family; it has to be through descent.

person in the village if that network applies (starting at the ID of that household head), including other adults. We thus have networks that are identified at the household level[5].

During the second visit (visit B), the treatment was implemented (see section 3.3 for more details). During this visit to the village, we also elicited individuals' Willingness to Pay (WTP) for chemical fertilizer, using Randomized Card Sorting (RCS) as used in Smith (2006). 18 participants were selected from the village (6 most central, 6 most middling and 6 most isolate) and were given 10 cards with prices in Congolese Francs on them. The prices ranged from 100 Fc (about 0.90US$) to 5000 Fc (about 4.50US$). The market price for fertilizer in the large regional town Bukavu was 1.60 US$. We asked them which of these amounts they would pay for a 1 kg bag of NPK fertilizer. We asked them to sort them into three piles: would pay, would maybe pay and would not pay.

A big problem with Willingness to Pay measurements is starting point bias (or anchoring). Respondents rarely deviate far from the initial valuation posed by the researchers. By giving all the amounts to the respondent in a random order this is prevented. Smith (2006) compares three approaches that use payment cards, and finds evidence that randomizing the order of the cards yields the most valid results.

The last visit (Visit C) took place about 2 weeks after visit B. This visit consisted of two parts:

1. Tracking the distribution of the fertilizer bags
2. Learn about individuals' knowledge on fertilizer use

Part 1 started by asking the three ambassadors to whom they handed fertilizer. Those that received kits were tracked, and enumerators asked them to whom they gave fertilizer. All individuals were visited to verify receiving the fertilizer and to ask what was done with the fertilizer.
Part 2 was conducted with all heads of household in the village. The Visit C Survey repeated the module on fertilizer knowledge from the Visit A survey. This gives us a panel dataset from before and after the training to look at the impact of the training on resource recipients as well as those who did not receive any fertilizer or training. In addition to knowledge on fertilizer outcomes, we also collected panel data on the willingness to pay for fertilizer using Randomized Card Sorting, but only for those who played this in visit B. This allows us to compare their willingness to pay to a simple survey question that was asked of all participants. Finally, we also asked about whether participants had received fertilizer in the previous weeks. As a result, we have data on fertilizer distribution from both the ambassadors' perspective and the receivers' perspective, which allows us to verify their answers.

3.3 The intervention

The experiment used for this project consisted of two components:

1. Resources: 1kg bags of chemical fertilizer that were to be distributed throughout the village by randomly selected village members;
2. Knowledge: We trained the village members in fertilizer use and optimal application. We also asked the village members to spread this knowledge.

Three selected[6] village members took part in a short training session (one hour in length) on chemical fertilizer led by the agronomists amongst our enumerators. These village members were termed the "Ambassadors". Topics in the training session included the types of fertilizer, benefits, application methods and prices and access points in Bukavu.

[5] Specifically, for those lines that are related to the heads of the households, the question is "Is person [name of the head of that other households] blood family of this household's head? For those lines related to the other adults in the other households, the question is "Is there another person in this other household blood family with your household's head?". We thus obtain information not only if this household's head is connected to the other household's head, but to any adult in that household.

[6] We also had several backup ambassadors if the first selected were not present.

After the training session, each Ambassador received one 1kg bag of chemical fertilizer (NPK) and 3 "kits". Each kit contains three 1 kg of chemical fertilizer, three stickers and three pens. Ambassadors were instructed that they must distribute the three kits to three household heads of their choosing. These chosen household heads may keep one small bag of fertilizer for themselves, but must further distribute the other 2 bags: one to one other head of household, and one to another head of household.[7] By design we thus created the following networks *per* ambassador:

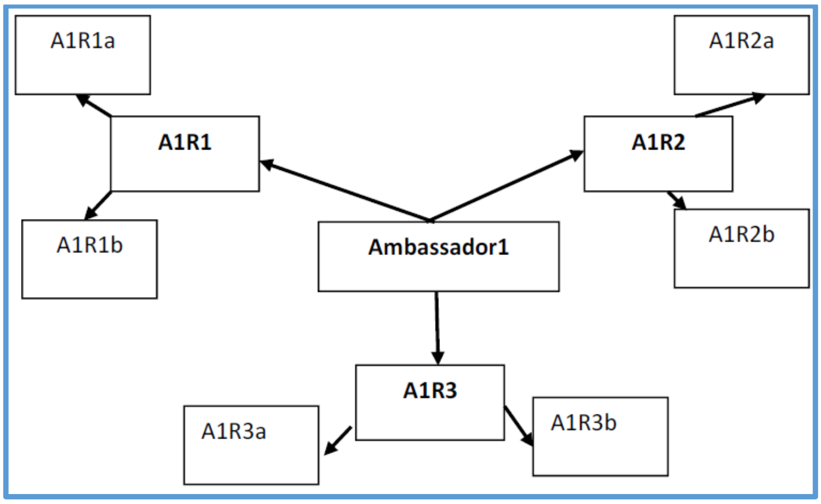

Figure 2: Network per Ambassador

Each Ambassador also received 10 pens. Ambassadors were asked when distributing to put a sticker on each of the three kits they pass on. They were to write on this sticker to whom the fertilizer was being given, and the date and time of the transfer. Moreover, those that received the kit from the ambassador also received pens and the stickers, and so the chain continued. Each ambassador receives 10 of the same pens, but across ambassadors they were different so that the pens acted as another tracker for the distribution flows. Every village received the same training and the same amount of a resource to distribute in order to have comparable diffusion networks mapped for each village.

In-between Visit A and B we used the social network data to calculate the eigenvector centrality of all heads of household. We are interested in flows of information and resources. These might flow through any of our networks (Blood Family, Field Neighbours, Discuss Agriculture). Therefore, for all analyses we use the union of these networks: for an edge (relationship) to be present in the union it needs to be present in any of the networks. Since an edge need not be reciprocated for a node (head of household) to transfer information or resources, we convert our directed network data to undirected network data: if either node claims a connection to another node, we will treat it as an edge between the two. We refer to the network after these conversions as simply the combined network. We opted to utilize this combined network approach as it provided the highest variance in node centrality, and reduced the consequences of potential measurement error from any single dimension.

We used this data to randomly assign villages to two different treatments:

Treatment "Central Ambassadors": In twenty randomly selected villages the ambassadors were the 3 heads of household with the highest eigenvector centrality. Generally, we expect that the central ambassadors are more influential, as Bonacich (2007) predicts. This should lead to higher uptake and knowledge for those trained by central ambassadors.

[7] We thus have links of size two.

Treatment "Isolate Ambassadors": In the other twenty villages the ambassadors were the 3 heads of household with the lowest eigenvector centrality. Here we expect to see lower uptake and usage of chemical fertilizer.

Figure 3 illustrates the two treatments. It shows the network map for two villages. The selected ambassadors are represented by red nodes, while other heads of household are blue nodes. Obviously, in the village that received the isolate treatment the ambassadors are on the fringes of the network while in the central treatment they are more central in the network map.

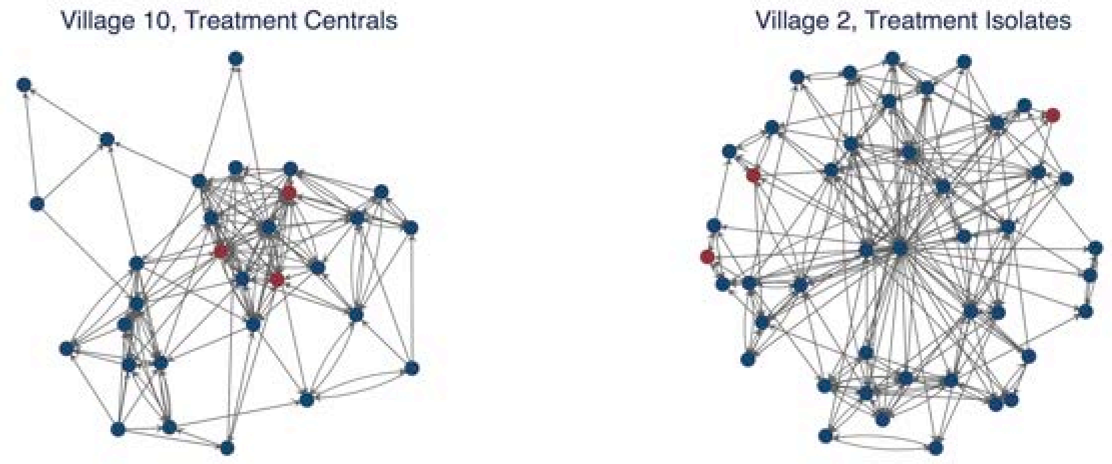

Figure 3: Ambassador Selection

We decided upon chemical fertilizer as the resource to spread for several reasons. While farmers sometimes use fertilizer, it is rare, and we wanted our intervention to be sufficiently new. We wanted a learning component in the intervention as well, and chemical fertilizer needs to be used in a certain way. We also wanted our treatment to have a significant positive effect. Fertilizer can greatly increase yields in areas with low soil fertility. We used a type of fertilizer called NPK because of its flexibility: it can be used throughout the agricultural season and has a positive effect on a range of crops.

The three village members trained by our enumerators are referred to as first-stage ambassadors. Those villagers that received fertilizer and training from these first-stage ambassadors are called second-stage ambassadors *if* they also passed fertilizer or knowledge. If they did not, they are simply referred to as receivers. Villagers that did not receive anything are termed non-receiverers.

3.4 Variable Definition, Descriptive Statistics and Balance

We examine the treatment effect in terms of five outcome variables. For pre-intervention variables summary statistics are in Table 2. It shows the means in both treatment groups, including a p-value that checks if means are equal between the two groups.

1. **Recent Fertilizer Use.** This variable captures if the respondent recently used fertilizer. In Visit A and C, participants self-reported if they used chemical fertilizer in that agricultural season. Given that both visits took place in one agricultural season, if they started using fertilizer between visits, the usage must have happened between our two visits, and can likely be attributed to our intervention. Before

our intervention, around 6% of respondents had ever used fertilizer in the past. This confirms that it is a sufficiently new technique.

2. **Fertilizer Knowledge.** Based on questions surrounding the effects, timing, methods and availability, we construct a score that measures respondent's knowledge surrounding fertilizer. We asked this in visit A and C, and thus have panel data. The score ranges from 0 to 8.5. The score pre-intervention was around 1.27. Fertilizer Knowledge was low pre-intervention.
3. **Willingness to Pay.** This captures the price respondents are willing to pay for a 1 kg bag of NPK chemical fertilizer. We have data for this in two ways: using randomised card sorting (RCS) and as a simple survey question. For RCS we have panel data, but it is not representative for the village as a whole. For the RCS data respondents sorted price cards into three stacks: would pay, would maybe pay and would not pay. We use the highest amount in the would pay stack as the amount they are willing to pay. On average, participants are willing to pay around 1.50 US$ for a 1 kilo bag of NPK when prompted using the RCS method before the intervention.
4. **Longest Path.** We are interested in the wideness of the fertilizer distribution in the village. To examine this, we look at all nodes that received fertilizer (from us or an ambassador), and then calculate the geodesic distance between all receivers. The geodesic is the minimum number of edges that need to be used to connect one node with another. Then, we take the longest of these distances which is divided by the number of nodes in a village to make the results comparable. If there is no longest path possible (because the receivers are in unconnected parts of the network) this variable takes the maximum value, 1. For this variable we only have post-intervention data, so it is not in table 2.
5. **Speed of Diffusion.** Besides wideness, the speed of diffusion is also interesting when distribution needs to happen fast. When ambassadors were giving fertilizer they were asked to note the date and time. In visit C we collected this data. We take the date of the interaction and subtract the date of visit B to calculate the number of days since ambassadors received it. If it was given in the evening (after 17.00), we add 0.5 to the number of days. This variable is not in table 2.

Next is some further information regarding respondents. Respondents are around 46 years old and 32% is female. Individuals respond that they rarely participated in the war. Most respondents were exposed to the wars in several ways though, for example by fleeing, losing a family member or being injured. The household size is around 6 people (including wives and children), and 33% was not born in the village they now live in. This is a result of the wars that have taken place in the area. The average participant farm plot is around 51 km^2, which is very high. When we restrict the maximum land size to 100 km^2, the average drops to 5 km^2, which is more realistic. Generally, fields lie around the villages and can thus be quite large[8]. The indegree is the number of edges that lead towards a node. We see that the average indegree in central treatment villages is lower, 8 instead of 10. Since our treatment is based on social network calculations, we are worried that this imbalance in indegree might affect the treatment effect. For our treatment analyses (in section 4.1) we add indegree as a control variable. The density is the percentage of possible edges that is realized. We see that it averages 15% in our villages. The clustering coefficient is a measure of clustering within a network. We see that it is relatively high, at 61%. The average path length measures overall connectedness, divided by the maximum path length. Villages have around 65 heads of household.

[8] For subsequent analyses that require land size as a predictor we will use the measure without outliers. We did not account for this in the pre-analysis plan.

Table 2: Descriptive statistics

	Treatment Centrals			Treatment isolates			
Name	Observations	Mean	SD	Observations	Mean	SD	P-value
Ever used fertilizer (yes=1)	1158	0.06	0.24	1285	0.07	0.26	0.424
Fertilizer Knowledge Score	1166	1.27	1.47	1299	1.26	1.44	0.954
Willingness to Pay (RCS)	335	1.49	1.34	339	1.59	1.46	0.664
Age	1191	47	18	1336	45	17	0.149
Gender (Female=1)	1191	0.32	0.47	1336	0.32	0.47	0.858
War Participation (participated=1)	1170	0.02	0.14	1301	0.02	0.13	0.666
War Exposure	1171	2.41	1.07	1304	2.22	1.01	0.249
Household Size	1171	6.29	3.14	1304	6.37	3.18	0.748
Migrant (yes=1)	1191	0.35	0.48	1335	0.31	0.46	0.181
Land Size	1171	37	504	1304	64	1219	0.476
Land size without outliers	1149	5.11	11.16	1266	4.80	9.91	0.722
Indegree	1195	7.79	5.08	1338	10.02	7.10	0.029***
Density	20	0.14	0.07	20	0.15	0.07	0.620
Clustering coefficient	20	0.60	0.18	20	0.62	0.20	0.783
Average Path length	1195	0.03	0.01	1338	0.03	0.01	0.921
Village Size	20	62	19	20	68	13	0.246

- P-values calculate differences in means between treatments. P-values from regressions with standard errors clustered at the village level. We deviate from the PAP here. We said we would do simple t-tests, but that does not control for clustering within a village.

3.5 Treatment Implementation

To confidently speak of a causal effect of our treatment, we need to assure that the treatment was implemented properly. A total of 97 out of 120 first-stage ambassadors actually distributed fertilizer, which is 80% of the intended amount of ambassadors. The other 20% of first-stage ambassadors decided to keep all fertilizer for themselves, or decided not to distribute yet. Some said they would wait until the next agricultural season to distribute. We did not come back to each village after the same amount of days. So some ambassadors had less time to distribute their goods. Each village had at least 14 days. The number of days is distributed equally across treatments. The fertilizer was generally distributed within 5 days, though our number of observations is low because it was only registered in about half of the cases. We see significant variation. Some ambassadors distributed fertilizer on the same day they received it, while others did not distribute until the day we returned to the village, 28 days later. Given that each first-stage ambassador was asked to share with three people we should have a total of 360 second-stage ambassadors. We find only 102 second stage ambassadors that also distributed fertilizer (28% of the total). The rest decided to keep all the fertilizer they received for themselves, or said they would distribute at a later date[9].

Finally, we look at the centrality of our ambassadors. This is a check to see if the ambassadors that ended up distributing in a village were indeed of the centrality we wanted. The centrality of first-stage ambassadors in

[9] It might be that while first-stage receivers received so much (10 kg) they decided to share some, this was not the case for second-stage ambassadors, who preferred to hold what they received for themselves. It might also be that later stage ambassadors felt less of an obligation to follow our instructions, as we did not directly train them.

villages where we selected central ambassadors is much higher (0.87) than in villages where we selected isolate ambassadors (0.08). This is by design. However, we also see significant variation, with one supposed isolate ambassador having almost the maximum centrality score. It might be that our enumerators trained the wrong person to be an ambassador. It might also be that someone took the fertilizer from an ambassador, and then claimed to be an ambassador himself.

Overall, we see our intervention did not happen exactly as we intended: the amount of diffusion was lower, and not all ambassadors were of the correct centrality. As such, we will not be able to evaluate an Average Treatment Effect (ATE) as we proposed in our Pre-Analysis Plan. For an ATE we would need to be confident that our treatment worked exactly as intended. Instead, we examine an Intent to Treat Effect (ITT). Now we simply look an effect of what we *intended* the treatment to be. By simply referring to it as the intention to assign a treatment, we can still make causal inferences. Because we still have significant variation in the centrality of ambassadors, our treatment could still have an effect. This is the focus of section 4.1.

4 Results

4.1 Intent to Treat Effect

We first look for the causal effect of eigenvector centrality on resource spread, technology adoption and knowledge. To examine this question, we look for the effect of treatment assignment on several outcome variables. For Fertilizer use, Speed of diffusion, Wideness of distribution and the Willingness to Pay (survey question) we have cross-sectional data. For Fertilizer knowledge we have panel data and we also use the difference-in-difference approach to get at the treatment effect. We use the following models:

$$\boldsymbol{Y_{ij}} = \beta_0 + \beta_1 T_j + \varepsilon_{ij} \qquad (1)$$

T_j=Treatment

$$Y_j = \beta_0 + \beta_1 T_j + \varepsilon_j \qquad (2)$$

$$\boldsymbol{Y_{ijt}} = \beta_0 + \beta_1 T_j + \beta_2 \, Post_t + \beta_3 \, Post_t * T_j + \varepsilon_{ij} \qquad (3)$$

i is the individual level, j is the village level and t is the time period (pre-intervention t=0, post-intervention t=1). **Y** in these models are the five treatment variables defined above. T_j is the the treatment, which is at the village level. Models 1 and 2 both look for the intent to treat effect on the outcome variable, but model 1 is at the individual level while model 2 is at the village level. Model 3 examines the Difference-in-Difference effect for those variables where we have panel data[10]. Post is a dummy that signifies pre or post treatment. β_1 gives the effect of the treatment, β_2 gives the effect over time and β_3 gives the effect of treatment over time. This last coefficient is where we find the effect of the treatment.

In Table 3 we examine the causal effect of eigenvector centrality on technology adoption, speed and wideness of distribution, willingness to pay for for fertilizer and fertilizer knowledge. Columns 1, 2, 4 and 5 use model 1. Column 3 uses model 2 and columns 5-6 use model 3. Generally, we expect to see negative coefficients on the isolate treatment variables. Bonacich (2007) and Borgatti (2005) predict that central ambassadors are more effective at increasing adoption and communicating knowledge. For regressions 5-6 we examine the DID effect. Then the treatment variable of interest is the interaction between Treatment and a Later Visit dummy (Visit C*Treatment). The Visit C dummy reports the change over time. We report Newey-West standard errors robust

[10] While the pre-analysis plan does mention using a difference-in-difference approach, it was not specified correctly. We correct this now.

to heteroskedasticity and autocorrelation, clustered at the village level. For DID analyses we report both Fixed Effects and Random Effects regressions[11].

Table 3: Intent to Treat Analysis of Centrality

	(1) Fertilizer Use	(2) Speed of Diffusion	(3) Wideness of Distribution	(4) WTP Survey	(5) Fertilizer Knowledge, Visit C	(5) Fertilizer Knowledge, FE	(6) Fertilizer Knowledge, RE
Treatment (isolates=1)	0.004 (0.031)	-4.299** (1.680)	0.078 (0.075)	-0.194 (0.140)	-0.061 (0.132)		-0.072 (0.141)
Indegree	0.005** (0.002)	-0.104 (0.066)	-0.005 (0.004)	0.025*** (0.009)	0.026*** (0.008)		0.028*** (0.006)
Visit C*Treatment						0.015 (0.144)	0.009 (0.145)
Visit C						0.701*** (0.110)	0.702*** (0.110)
Constant	0.081*** (0.028)	8.602*** (1.494)	0.149* (0.084)	0.859*** (0.139)	1.768*** (0.105)	1.267*** (0.035)	1.057*** (0.113)
Observations	2208	223	40	2304	2305	4770	4770
Number of Clusters	40	33		40	40	40	40
Correlation WTP Survey v RCS				0.47***			
P Hausman							0.80

- Models 1-5 use OLS regressions. Robust standard errors in parentheses clustered at the village level.
- p < 0.10, ** p < 0.05, *** p < 0.01.
- We deviate from the PAP here. Indegree was added to control for imbalance after the randomisation

We find one interesting effect of varying the network entry point on our outcome variables. Column 1 shows the impact of isolate ambassadors on the fertilizer use by all villagers. We see that isolate ambassadors lead to slightly higher fertilizer uptake in the village, but the result is not significant. While our power is not very high (40 villages), the magnitude of the effect is very small, so it is unlikely that more power should improve this. This result is surprising, as a higher eigenvector centrality predicts higher average adoption. Villagers with a higher average indegree are more likely to adopt. This might be because they are more likely to be connected to an ambassador. Column 2 shows the effect on the speed of diffusion[12]. Isolate ambassadors take around 4 days less to distribute their fertilizer, which is opposite to the result we expected. We expected that central ambassadors would distribute faster, as they had more connections and thus more potential people to distribute fertilizer to. It might instead be that because isolate ambassadors had less potential people they wanted to distribute to, their decision was easier and thus faster. Column 3 shows that isolate ambassadors appear to distribute wider than central ambassadors. In central villages the longest shortest path between all receivers was 15% of the maximum length possible, while in isolate villages it was 23%. If this percentage is higher it means that the treatment has diffused further in the village. It makes sense that isolates lead to further diffusion: centrals are often clustered together in the centre of the network map, while isolate ambassadors can be anywhere on the fringes of the network. This result is not significant. As this variable is at the village level, our statistical power is very low, and we find high standard errors. This might drive the lack of significant result.

Column 4 examines the survey question for WTP. It does not show a difference between treatments. While more knowledge surrounding the advantages of fertilizer is likely to increase the WTP, higher local availability should depreciate the price. We cannot unpack these two separate effects, but combined either groups of

[11] The Random Effects regression has more power but might be biased. We check if this is a problem using a hausman test. A significant p-value means rejecting the RE results.

[12] The number of observations is lower here because it was not registered in all cases. This might be because ambassadors forgot to write the time down or because they did not know what the time was.

ambassador did not lead to a higher or lower willingness to pay. The correlation across the two methods we used (a survey question and Randomized Card Sorting) is high (0.47), and highly significant. This means that both methods are likely to be valid measures of willingness to pay. We continue to use only the survey measure for subsequent analyses, as this was done for all villagers. Column 5 shows that the treatment did not cause a significant increase in average fertilizer knowledge in the villages. We expected that knowledge retention would be higher in central villages. The magnitude of the effect is again low which suggests that this result is not driven by low statistical power. A higher indegree increases fertilizer knowledge. This increases the probability to be connected to an ambassador.

The final two regressions show the difference-in-difference effect. The variable of interest is Visit C*Treatment. We expect this to be negative, as central ambassadors should be more convincing. The FE and RE results are similar, and since the p-value of the hausman test is insignificant we prefer the RE result. We see for fertilizer knowledge there is a strong increase in knowledge between Visit A and visit C, because the coefficient for Visit C is positive. Our intervention has clearly increased the level of knowledge regarding fertilizer in the villages. We find no evidence that the increase is lower for isolate ambassadors.

Next, we look for attenuation (reduction of effect) of the intervention. We use the following model, which analyses whether the progression across groups of receivers differs between isolate and central ambassadors.

$$\boldsymbol{Y_{ij}} = \beta_0 + \boldsymbol{\beta^{cent}} \boldsymbol{X}_i + \boldsymbol{\beta^{iso}} \boldsymbol{X_i} * T_j + \beta_1 T_j + \varepsilon_{ij} \qquad (4)$$

X_i = (fromfirst, fromsecond, nonreceivers)

To do so we compare four groups: first-stage ambassadors, those who received from first-stage ambassadors, those who received from 2nd (or higher) stage ambassadors and non-receivers. Dummy variables that indicate membership of a group are in the row vector X_i. The difference in individual measures of the dissemination effect between the groups is caused by their exposure to the treatment. We expect that some of the message is lost as it spreads throughout the village. As a result, we expect that 1st stage ambassadors know most, those who received from them know less, those who received from 2nd stage ambassadors know even less, and non-receivers should know the least. Figure 4 illustrates the four groups. By interacting the coefficients for these groups with our treatment variable, we can see if the tapering off of the effect is lower for central ambassadors, shown in β^{iso}. This is presented in table 4. The main variables of interest here are Treatment and the interactions with Treatment. We expect them to be negative, which would mean that the dissemination effect tapers off more for isolate ambassadors.

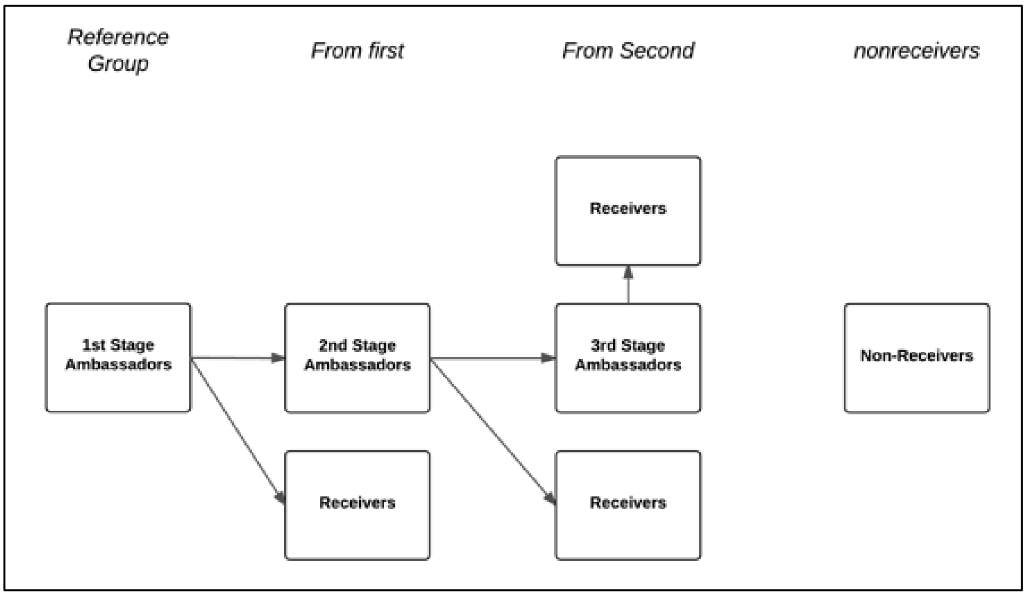

Figure 4: Intervention progression in groups

Table 4: Centrality and intervention attenuation

	(1) Fertilizer Use	(2) Fertilizer Knowledge	(3) WTP Survey
Received from 1^{st} stage ambassadors	-0.062	-0.423**	0.374
	(0.079)	(0.188)	(0.540)
Received from 2^{nd} stage ambssadors	-0.160*	-0.688**	-0.069
	(0.084)	(0.294)	(0.329)
Non-Receivers	-0.261***	-1.194***	-0.391
	(0.080)	(0.194)	(0.242)
Treatment (1=isolates)	0.027	0.080	-0.076
	(0.121)	(0.280)	(0.286)
Treatment*Received from 1^{st} stage ambassadors	-0.045	0.045	-0.380
	(0.124)	(0.310)	(0.581)
Treatment*Received from 2^{nd} stage ambssadors	-0.065	0.130	-0.198
	(0.118)	(0.402)	(0.395)
Treatment*Non-Receivers	-0.005	-0.150	-0.037
	(0.114)	(0.300)	(0.291)
Indegree	0.003	0.019**	0.021**
	(0.003)	(0.008)	(0.009)
Constant	0.305***	2.819***	1.133***
	(0.088)	(0.198)	(0.252)
Observations	2208	2305	2304
Number of Clusters	40	40	40

OLS regressions. Robust standard errors in parentheses clustered at the village level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We again see no effect of our treatment: in no case do the variables intereacted with the treatment variable significantly differ from zero. This is the case for fertilizer use, fertilizer knowledge and willingness to pay. So isolate ambassadors were no worse than central ambassadors at sustaining the effect of the intervention. Centrality makes no predictions on the teaching quality of network members. Therefore it is not surprising that the groups are equally good at sustaining the effect. The other variables do enter significantly, these show intervention attenuation. This is examined in the next section.

4.2 Intervention Attenuation

We conjecture that the effect of our intervention, where we distributed fertilizer and information regarding fertilizer, attenuates as it spreads through the village. We again look at the four groups in Figure 4: 1^{st} stage ambassadors, those who received from 1st stage ambassadors, those who received from 2nd stage ambassadors and non-receivers, but now we will directly compare between these four groups. This is shown in Figure 5. The top-left panel shows the percentage that has recently used fertilizer. Fertilizer uptake clearly goes down as the intervention moves away from the entry point. The same holds for the top-right panel, which shows the fertilizer knowledge score. This also decreases down during diffusion (the maximum possible score is 8.5). For willingness to pay the effect is less strong. The effect even appears to be non-linear: first-stage receivers value fertilizer higher than ambassadors. We analyze these results more formally now.

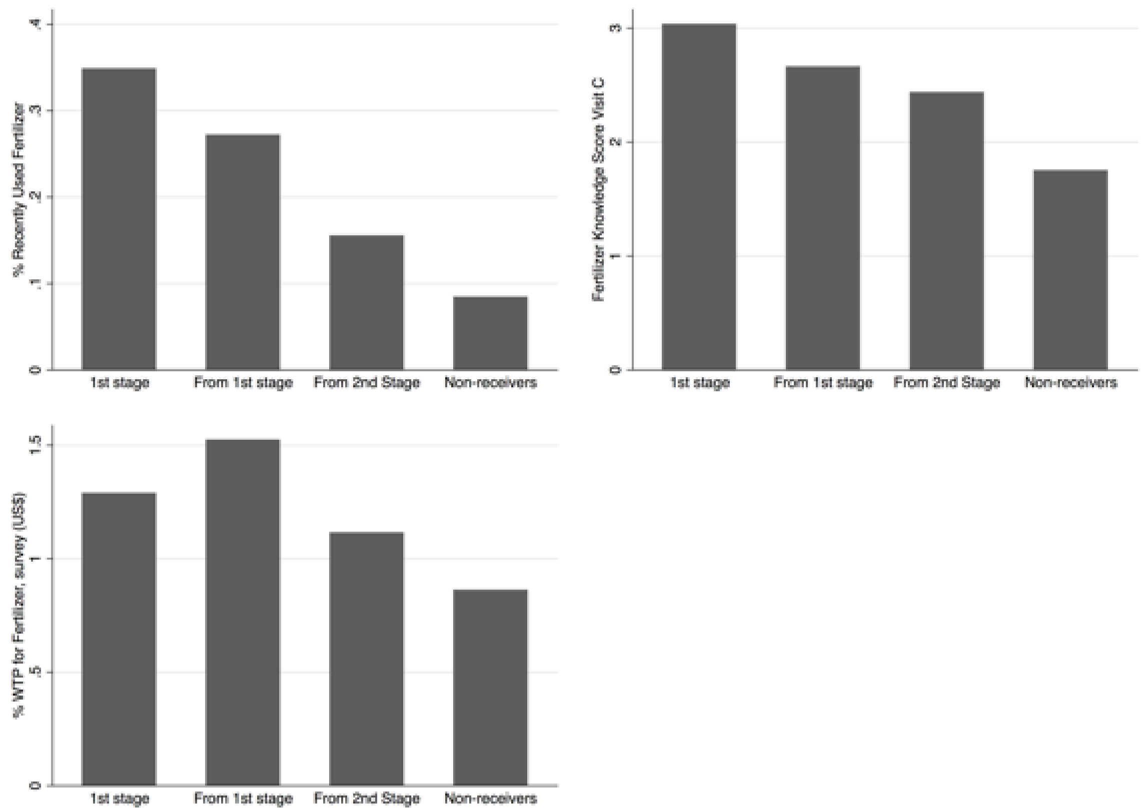

Figure 5: Intervention Attenuation

We use the following model:

$$\boldsymbol{Y_{ij}} = \beta_0 + \beta_1 fromfirst_{ij} + \beta_2 fromsecond_{ij} + \beta_3 nonreceivers_{ij} + \varepsilon_{ij} \qquad (5)$$

To analyze whether the treatment attenuates in a regression framework, we use 1st stage ambassador as the reference group, as we expect them to be most knowledgeable. Their average values will show up in β_0, the constant. For non-1st stage ambassadors we expected the results to be lower, so all other coefficients should be negative. We also expect the attenuation to be higher further away from the entry point ($\beta_1 > \beta_2 > \beta_3$). This is done by individual t-tests between the coefficients, shown in the bottom of the table. These results are in Table 5.

Table 5: Intervention attenuation

	(1) Fertilizer Use	(2) WTP survey	(3) Fertilizer Knowledge, FE	(4) Fertilizer Knowledge, RE
Received from 1st stage ambassadors	-0.077	0.231	-0.629***	-0.391***
	(0.058)	(0.285)	(0.194)	(0.144)
Received from 2nd stage ambassadors	-0.193***	-0.171	-1.029***	-0.641***
	(0.056)	(0.183)	(0.257)	(0.202)
Non-Receivers	-0.265***	-0.429***	-1.270***	-1.284***
	(0.053)	(0.136)	(0.210)	(0.152)
Visit C			1.816***	1.770***
			(0.195)	(0.154)
Constant	0.349***	1.290***	1.269***	1.267***
	(0.056)	(0.131)	(0.034)	(0.074)
Observations	2208	2304	4770	4770
Number of Clusters	40	40	40	40
P fromfirst=fromsecond	0.0022	0.18	0.039	0.066
P fromsecond=nonreceiver	0.077	0.083	0.24	0.000
P Hausman				0.017

Models 1-2 use OLS regressions. Robust standard errors in parentheses clustered at the village level.
* p < 0.10, ** p < 0.05, *** p < 0.01.

We see that there is attenuation for adoption and knowledge, and less so for willingness to pay. Looking at fertilizer use (Column 1), we see that all groups are less likely to use fertilizer than 1st stage ambassadors, though the difference between ambassadors and those receiving from them is not significant. The difference between those who received from 1st stage ambassadors group and those that received from 2nd stage ambassadors is significant, and so is the difference between the latter group and non-receivers. The difference between those who received from 1st stage ambassadors and 2nd stage ambassadors is also significant, though between the latter and non-receivers it just isn't. This shows that relying on multi-step diffusion is not very effective: third-hand receivers do not adopt more than non-receivers.

This effect is not as clear for willingness to pay, in column 2. It is not consistenly negative, as we expected, and only the difference between non-receivers and 1st stage ambassadors is significant. The valuation of fertilizer is lower for non-receivers than 1st stage ambassadors, as we expected. We cannot clearly conclude attenuation in this case. For fertilizer knowledge (Column 3-4) we again see attenuation. We ignore the RE results, as the hausman test is significant. 1st stage ambassadors are most knowledgeable regarding fertilizer, then those who received from them, then those who received from 2nd stage ambassadors. Only between this last group and non-receivers is there no significant difference. Like for adoption, third-hand receivers are no better off than non-receivers. Again, we see that knowledge regarding fertilizer has gone up over time, as by the Visit C coefficient. Overall, we see attenuation. While a multi-step distribution approach has some merit to increase diffusion, the magnitude of the intervention drops. Those at the end of the receiving chain are no better off than non-receivers.

4.3 Predicting Diffusion

We attempt to determine what criteria drive the ambassadors in their choice of recipient for the fertilizer[13]. We use the following model for this analysis:

$$\boldsymbol{Y}_{aij} = \boldsymbol{\beta X_{aij}} + \varepsilon_{aij} \qquad (6)$$

a refers to all ambassadors (1st and 2nd stage) in a village, and i to all other villagers. **Y** is a column vector of our two decision variables for ambassadors: who to give fertilizer to and who to give information to. **X** is a row

[13] For this analysis, we reshape our data to ambassador level, so that each ambassador is repeated for n_j rows, the number of people in their village.

vector of all the variables we expect can predict giving behavior. We examine the decision of the ambassador by looking at characteristics of the (possible) receivers, and the difference in several characteristics between the ambassador and the receiver. Positive coefficients mean that an increase in this variable increase the likelihood that this person is given fertilizer. We do this separately for receiving fertilizer (Column 1) and receiving information regarding fertilizer (Column 2)[14]. Besides that, we also interact all variables with our treatment variable to see if for isolate ambassadors other characteristics matter (Columns 3 & 4). These results are in table 6.

Table 6: Predicting Diffusion Behaviour

	(1) Fertilizer Given, ambassador perspective	(2) Information Given, receiver perspective	(3) Fertilizer Given, ambassador perspective	(4) Information Given, receiver perspective
Geodesic	-0.071 (0.110)	0.051 (0.104)	-0.139 (0.085)	-0.161 (0.130)
Physical Distance	-2.227*** (0.477)	-1.789*** (0.367)	-2.877*** (0.731)	-1.596** (0.662)
Blood Family	0.448*** (0.146)	0.465*** (0.110)	0.633*** (0.152)	0.463*** (0.145)
Field Neighbours	0.359* (0.183)	0.561*** (0.191)	0.257 (0.277)	0.382 (0.262)
Discuss Agriculture	0.499*** (0.152)	0.422*** (0.140)	0.771*** (0.202)	0.497** (0.237)
Centrality of receiver	0.186 (0.268)	-0.014 (0.262)	-0.071 (0.329)	0.187 (0.488)
Difference in centrality	-0.099 (0.229)	-0.067 (0.307)	1.272*** (0.193)	1.443*** (0.268)
Age of receiver	0.006* (0.003)	0.000 (0.003)	0.003 (0.005)	-0.003 (0.004)
Gender of receiver	-0.207 (0.204)	-0.081 (0.145)	-0.805*** (0.298)	-0.340 (0.244)
Whether the receiver is opposite gender	0.204 (0.166)	-0.145 (0.134)	-0.438 (0.280)	-0.288 (0.199)
Migrant	0.132 (0.118)	0.118 (0.105)	0.056 (0.197)	0.305* (0.162)
Income Index	-0.108* (0.066)	-0.019 (0.088)	-0.075 (0.067)	-0.007 (0.116)
Difference in Income Index	-0.155** (0.062)	-0.061 (0.067)	-0.150*** (0.057)	-0.039 (0.073)
Being in a similar agricultural group	0.248* (0.146)	-0.061 (0.174)	0.318* (0.174)	-0.147 (0.310)
Village Size	-0.010* (0.006)	-0.014* (0.008)	-0.010 (0.007)	-0.008 (0.007)
Treatment (isolates=1)			-0.107 (0.759)	1.035 (1.270)
Geodesic* Treatment			-0.050 (0.189)	0.243 (0.181)
Physical Distance* Treatment			0.764 (0.999)	-0.626 (0.801)
Blood Family* Treatment			-0.029 (0.298)	0.206 (0.213)
Field Neighbours* Treatment			0.156 (0.377)	0.334 (0.391)
Discuss Agriculture* Treatment			-0.530 (0.333)	-0.087 (0.299)
Centrality of receiver* Treatment			-0.073 (0.513)	-0.669 (0.608)

[14] We examine the results from the perspective of the ambassadors because this makes more sense: they should know the best who they gave fertilizer to. Unfortunately for information giving separately we only have data from the receiver perspective. In the robustness section we examine the receiver perspective for fertilizer receiving.

Difference in centrality * Treatment			-2.610*** (0.297)	-2.831*** (0.366)
Age of receiver * Treatment			0.006 (0.006)	0.006 (0.005)
Gender of receiver * Treatment			0.793** (0.400)	0.429 (0.327)
Opposite Gender * Treatment			0.872*** (0.326)	0.117 (0.267)
Migrant* Treatment			0.043 (0.260)	-0.506** (0.222)
Income Index* Treatment			0.034 (0.097)	0.083 (0.178)
Difference in Income Index* Treatment			0.034 (0.079)	0.003 (0.121)
Being in a similar agricultural group * Treatment			-0.179 (0.280)	0.086 (0.358)
Village Size * Treatment			-0.000 (0.010)	-0.014 (0.016)
Constant	-2.271*** (0.426)	-2.137*** (0.593)	-2.165*** (0.585)	-2.615*** (0.698)
Observations	9906	12371	9906	12371

- Logistic regressions. Robust standard errors in parentheses clustered at the village level and ambassador level.
- p < 0.10, ** p < 0.05, *** p < 0.01.
- The geodesic is the minimum number of edges that need to be used to connect one node with another
- Income index is calculated using the Stata program WMEANEFFECTS and comprises of the following variables: # chickens owned, # goats or sheep owned, # cows owned, land size, land access
- The PAP said we would check for differences across treatment using a Chow test. However, this is not possible when clustering at two levels, so we opted to simply add interactions with treatments to look for differences in coefficients between ambassador types. We also said we would weight for the probability to be selected based on village size. This is not possible when clustering at two levels, so we simply added it as a control variable

We find a number of factors that predict sharing behavior. We see that people are more likely to give to people that live closer to them physically, as expected. People are dependent on and are close to their neighbors, or they move closer to people they like. We also see that if either the ambassador or the receiver claims to share a network with the other, they are more likely to give fertilizer to this person. This reassures us that our social networks variables make sense and are important in daily life. Ambassadors are more likely to give to their family. This is probably out of loyalty or because of reciprocal exchanges. We also see that ambassadors prefer to share with their field neighbors and people they discuss agriculture with. This is probably because they already share more agricultural goods and information with these individuals. The centrality of the receiver or the difference in centrality does not have an effect, also as we expected. We expected that ambassadors would give to people with similar centrality, as these people would be more like them. Ambassadors are more likely to give to receivers that are older than them, as predicted, though the coefficient is not very significant. It might be that ambassadors gave to older people as a sign of respect. We see no effect of gender or migration status. The income index difference coefficient is negative, which means that ambassadors are more likely to give to receivers that have a lower income. What we might be seeing is charitable behavior to help those less well off. Being in a similar agricultural group is barely not significant, but the magnitude of the effect is high, indicating high noise. We do not have membership on specific agricultural groups, only if they were in a similar one (an example is an agricultural cooperative) which makes our measure noisier. It appears that ambassadors also shared within these groups. The number of nodes in a village is added as a control, and is not easily interpretable. Information passing, in column 2, is much harder to predict. Only being in one of the social networks can predict information distribution. The results are not caused by low power, as the number of observations is higher and standard errors are similar. We see that predicting information passing is much harder.

Column 3 interacts each variable with our treatment variable. The interacted variables show whether the motivations differed for isolates. We see that isolate ambassadors are more likely to give to people that are more central than them. Central ambassadors are more likely to give to people that are less central than them. This is logical, as a larger proportion of the village will have a lower centrality score than central ambassadors. Isolates are also more likely to give to women, and are more likely to give to the opposite gender. We do not know what might cause this. Column 4 does the same, but for information passing. The results are similar though we see that isolate ambassadors are less likely to give to migrants. This might be because migrants are generally excluded more. However, why this effect should not be present for central ambassadors is puzzling. It might be that there are social norms to not exclude migrants, but isolates choose to ignore these norms. Overall, we see that there are several variables that can predict sharing behavior. Especially the social network variables are very robust and highly significant. There is only a modest treatment effect: there are few strong differences between the motivations of isolate and central ambassadors.

5. Robustness

There are several conditions that need to be met to ensure that the results presented above are robust. First, we want to check the validity of our social network measures. Second, there should be no differences in ambassador compliance across treatment. Third, we discuss the implication of lower than expected diffusion. Fourth, the attrition between Visit A and C needs to be random and not correlated with other variables. Fifth, some villagers were not found in any of the visits. We need to check if this absenteeism is random. Finally, we check whether the results for fertilizer giving differ if this is examined from the receiver's perspective.

The question of validity is often ignored in social network studies. This is unfortunate, as the effectiveness of using a survey to measure social networks is debateable. Social relations represent some of the most private aspects of life, and a survey does not allow for trust to develop between researcher and subject. Traditionally, researching social networks has been the field of ethnographers, who spent a considerable amount of time gaining the trust of their subjects. We have tried to overcome this in several ways during the study, and have critically examined our results afterwards. During the survey, the social network questions were asked about halfway through – before survey fatigue could be a problem and after some notion of trust could be developed. Furthermore, we always emphasised that results would remain anonymous, and always conducted the surveys in a private location. Besides that, we find several results that indicate the validity of the measures. We see that ambassadors often give to people that are in their network (Table 6). It is logical that sharing behaviour should follow social network lines. Furthermore, we also see that geodesic distance (distance as measured through social networks) and physical distance are very strongly correlated, as shown in Table 7, column 1. This effect could go both ways: people might choose people they are close with as neighbours, or they become close with their neighbours. Regardless of the direction of causality, it indicates the validity of our measure.

We examined the differences between our two groups of ambassadors, caused by differences in network position. However, any difference could be caused by differences in non-compliance. For example, it might be that isolate ambassadors did not choose to follow our instructions of creating second-stage ambassadors, but distributed all the 1-kilo bags themselves. Then it might seem that they are faster at distributing, but we would not be able to compare them to the central ambassadors. So we need to check if ambassador type can predict non-compliance. We define compliance as an ambassador who gave to exactly 3 people, as instructed. This is shown in table 7, column 2. We see no evidence that either group was less compliant with our instructions. As a further check, we also look at the number of people an ambassador gave fertilizer to. This is shown in column 3. Again, we see no effect of treatment. This contrasts with Larson's (2014) prediction that isolate members of communities are more likely to cheat. The expectation would then be that isolates gave to less members and kept more for themselves.

We find very modest effects of varying the centrality of network entry points. While a conclusion might be that centrality does not matter, this is not necessarily so. We found that the number of ambassadors that shared their knowledge and fertilizer was lower than we expected. Perhaps if we had seen more diffusion in the village the treatment effect would have become more apparent. Ambassadors had several reasons for not sharing yet. Several ambassadors pointed out that they wanted to wait until the next agricultural season, because they believed the fertilizer would have a larger effect when used earlier in the agricultural season. We emphasized that fertilizer is also effective when used later in the planting season. Furthermore, we only gave the ambassadors 14 -28 days to spread fertilizer, which we thought was sufficient time to distribute. Perhaps it was not. Furthermore, we mostly saw low diffusion for 2nd stage ambassadors. The treatment effect should be driven by the 1st stage ambassadors, because they were selected based on their centrality. 1st stage ambassadors did distribute in most cases (80%).

Since some of these analyses require panel data from the first and third visit, we do an attrition analysis on any respondents that were present in visit A but not in visit C. We compare several characteristics from our baseline data of our two groups: the returnees and the non-returnees. This is shown in column 4. We see that dropouts have had slightly lower war exposure. This is not very worrying as the difference is small and we do not expect it to be correlated with our main dependent variables

Furthermore, we also collected some data on participants we could not reach in any of the visits. This data was collected from a neighbor, family, or the village chief. We will compare this group against all who participated in visit 3. This is shown in column 5. We see that migrants are less likely to be absent, and farmers are also less likely to be absent. While the first result is puzzling, the latter is reassuring. For this research we were mostly interested in reaching farmers, it appears we succeeded. Regardless, since absence is so low (3%) we are not worried that this affected the precision of our social network maps.

Finally, we also asked all participants in the final visit if they received fertilizer from anyone. This allows us to check whether their answers line up with whom the ambassadors said they gave fertilizer to. We see that the two perspectives do not completely match up. In table A1 we check if using another perspective changes the results from table 6. Column 1 gives the results from the receiver's perspective. Column 2 gives the results when we look at interactions given by either or both receivers and ambassadors. Column 3 shows only the interactions that were present according to both receivers and ambassadors. Columns 4-6 are the same, but now the variables are interacted with the treatment variable. The results are largely unchanged. The physical distance remains important, as do the social network variables. Looking at the difference in income, we see that ambassadors are still more likely to give to poorer villagers, but the result is not very significant.

Table 7: Robustness Analyses

	(1) Geodesic Distance	(2) Compliance	(3) # People Given to	(4) Panel Attrition	(5) Absence
Physical Distance	1.073*** (0.150)				
Treatment (Isolate=1)		-0.214 (0.412)	-0.209 (0.452)		
Age				-0.005 (0.005)	0.007 (0.010)
Female				0.303 (0.236)	0.272 (0.218)
Centrality				0.075 (0.283)	
War Participation				0.302 (0.424)	
War Exposure				-0.183** (0.083)	
Household Size				0.005 (0.022)	
Migrant				-0.015 (0.200)	-1.233*** (0.409)

Fertilizer Knowledge (Visit A)				-0.060	
				(0.046)	
Land Size				-0.001	
				(0.001)	
Student					0.026
					(0.998)
Farmer					-1.244***
					(0.238)
Wage Labourer					-0.530
					(0.420)
Petty Trader					0.794**
					(0.334)
Miner					-0.030
					(0.310)
Unemployed					-1.299*
					(0.787)
Other Work					-0.356
					(0.470)
Constant	1.539***	-0.666**	3.054***	-1.763***	-2.519***
	(0.049)	(0.332)	(0.321)	(0.478)	(0.547)
Observations	10604	114	114	2461	2330
% attrited				0.091	
% absent					0.031

Robust standard errors in parentheses clustered at the village level. (Regression 1 also at ambassador level)
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6. Conclusion

We conduct a large-scale field experiment mimicking a farmer field school approach. We randomly assigned villages to treatments where we varied the eigenvector centrality of lead farmers. Villages received a treatment of three lead farmers with a high centrality or three lead farmers with a low centrality. The purpose of this was to see if integrating social network data in the farmer field school approach can increase adoption and social learning. Lead farmers were tasked with distributing 1 kg bags of chemical fertilizer and training farmers in proper usage and advantages. This study was implemented in 40 rural villages in Eastern DRC, where yields are low and roads are of bad quality. A farmer field school approach can be cheaper than conventional dissemination approaches here. Social network analysis at the village also makes sense as social linkages at the village level are important.

We find modest causal effects. Isolate ambassadors are somewhat faster at distributing resources. It might be that isolate ambassadors have fewer people they'd like to distribute to, easing their decision process. Central ambassadors do not cause higher uptake, knowledge retention or willingness to pay. This is contrary to the prediction that more central farmers are best at engendering attitudinal change. Central ambassadors cannot reduce attenuation as the intervention diffuses away from the network entry point. It might be that conventional measures of centrality (Eigenvector, Betweenness, Degree, etc) which are all highly correlated, are not very effective at identifying the most influential members of a network, as Banerjee et al (2013) already found. Alternative measures of centrality such as Diffusion Centrality are perhaps more useful. These have not been tested empirically yet.

We find strong evidence that fertilizer usage and knowledge retention decrease as the intervention diffuses away from the lead farmer. One way to overcome this is to by increasing the number of network entry points. Expecting that information will be passed on for many steps without losing quality is not realistic. We find that the third-hand receiver of the intervention is no better off than a non-receiver. This effect is not present for willingness to pay, although we find no clear effect of our intervention on willingness to pay in general.

We find several factors that predict to whom lead farmers decide to give resources. Social network characteristics matter strongly. Being family, discussing agriculture, being field neighbours or even living close to an ambassador increase the chance of receiving fertilizer. It is likely that they handed to these individuals because it matches existing patterns of gift and information exchange. Beyond that, older villagers are also more likely to receive, indicating respect for the elderly. Finally, we find that ambassadors prefer to share with villagers that are less well off than them. This is evidence of charitable behaviour of lead farmers. Predicting information passing is significantly harder. Again the social network variables are the most important predictor.

7. References

Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, *106*(51), 21544–21549.

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).

Beaman, L. A., Karlan, D., Thuysbaert, B., & Udry, C. R. (2013). *Profitability of fertilizer: Experimental evidence from female rice farmers in Mali*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w18778

Beaman, L., BenYishay, A., Magruder, J., & Mushfiq Mubarak, A. (2015). Can Network Theory-based Targeting Increase Technology Adoption? *Working Paper*.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, *29*(4), 555–564. http://doi.org/10.1016/j.socnet.2007.04.002

Borgatti, S. P. (2005). Centrality and network flow. *Social Networks*, *27*(1), 55–71. http://doi.org/10.1016/j.socnet.2004.11.008

Cai, J. (2012). Social networks and the decision to insure: Evidence from randomized experiments in china. *Job Market Paper*.

Duflo, E., Kremer, M., & Robinson, J. (2008). How High Are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya. *American Economic Review*, *98*(2), 482–488. http://doi.org/10.1257/aer.98.2.482

Duflo, E., & Saez, E. (2002). *The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment*. National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w8885

Durlauf, S. N. (2004). Neighborhood effects. *Handbook of Regional and Urban Economics*, *4*, 2173–2242.

Evenson, R. E., & Gollin, D. (2003). Assessing the impact of the Green Revolution, 1960 to 2000. *Science*, *300*(5620), 758–762.

Feder, G., Murgai, R., & Quizon, J. B. (2003). *Sending farmers back to school: The impact of farmer field schools in Indonesia* (Vol. 3022). World Bank Publications. Retrieved from http://books.google.com/books?hl=en&lr=&id=lSNlKz8wMwoC&oi=fnd&pg=PA7&dq=info:RuH0DVnjJBEJ:scholar.google.com&ots=KuPs6uBMK_&sig=QGaCygd5cHRjw83FwMDOGGEd9eE

Grund, T. (2014). nwcommands: Stata module for social network analysis. Retrieved from www.nwcommands.org

Kantor, D. (2013). carryforward: Stata module to carry forward previous observations. Retrieved from https://ideas.repec.org/c/boc/bocode/s444902.html

Larson, J. M. (2014). A Failure to Communicate: The Role of Networks in Inter- and Intra-Group Cooperation. *Working Paper*.

Oster, E., & Thornton, R. (2012). Determinants of technology adoption: Peer effects in menstrual cup take-up. *Journal of the European Economic Association*, *10*(6), 1263–1293.

Picard, R. (2012). geodist: Stata module to compute geodetic distances. Retrieved from https://ideas.repec.org/c/boc/bocode/s457147.html

Schaffer, M. E., & Stillman, S. (2010). xtoverid: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2 and xthtaylor. Retrieved from http://ideas.repec.org/c/boc/bocode/s456779.html

Smith, R. D. (2006). It's not just what you do, it's the way that you do it: the effect of different payment card formats and survey administration on willingness to pay for health gain. *Health Economics*, *15*(3), 281–293. http://doi.org/10.1002/hec.1055

8. Appendix

Table A1: Predicting Diffusion Behaviour, alternative perspectives

	(1) Fertilizer Given, receiver perspective	(2) Fertilizer Given, summed perspective	(3) Fertilizer Given, matched perspective	(4) Fertilizer Given, receiver perspective	(5) Fertilizer Given, summed perspective	(6) Fertilizer Given, matched perspective
Geodesic	0.221^{*}	0.057	0.102	0.052	-0.045	-0.085
	(0.119)	(0.113)	(0.126)	(0.112)	(0.086)	(0.144)
Physical Distance	-2.602^{***}	-2.211^{***}	-2.851^{***}	-3.142^{***}	-2.779^{***}	-3.464^{***}
	(0.449)	(0.435)	(0.470)	(0.721)	(0.789)	(0.667)
Blood Family	0.360^{***}	0.421^{***}	0.355^{**}	0.390^{**}	0.544^{***}	0.489^{**}
	(0.139)	(0.123)	(0.175)	(0.157)	(0.128)	(0.222)
Field Neighbours	0.448^{**}	0.466^{***}	0.229	0.426	0.316	0.338
	(0.227)	(0.174)	(0.265)	(0.277)	(0.246)	(0.343)
Discuss Agriculture	0.625^{***}	0.564^{***}	0.608^{***}	0.724^{***}	0.687^{***}	0.889^{***}
	(0.157)	(0.123)	(0.212)	(0.270)	(0.199)	(0.310)
Centrality of receiver	0.366	0.176	0.487	0.166	0.193	0.020
	(0.252)	(0.213)	(0.320)	(0.473)	(0.385)	(0.475)
Difference in centrality	-0.008	-0.124	0.109	1.703^{***}	1.493^{***}	1.793^{***}
	(0.287)	(0.255)	(0.270)	(0.207)	(0.169)	(0.327)
Age of receiver	0.003	0.003	0.005	0.002	0.002	0.003
	(0.003)	(0.003)	(0.004)	(0.005)	(0.004)	(0.006)
Gender of receiver	-0.163	-0.048	-0.475^{*}	-0.427	-0.594^{**}	-0.611^{*}
	(0.168)	(0.186)	(0.252)	(0.329)	(0.289)	(0.326)
Whether the receiver is opposite gender	-0.109	0.092	-0.053	-0.270	-0.421	-0.207
	(0.163)	(0.173)	(0.157)	(0.192)	(0.257)	(0.170)
Migrant	0.099	0.128	0.108	0.143	0.142	0.022
	(0.109)	(0.108)	(0.124)	(0.217)	(0.207)	(0.213)
Income Index	0.017	-0.036	-0.078	0.086	0.016	-0.048
	(0.087)	(0.068)	(0.087)	(0.109)	(0.071)	(0.107)
Difference in Income Index	-0.057	-0.097^{*}	-0.144^{*}	-0.002	-0.068	-0.134
	(0.071)	(0.058)	(0.076)	(0.084)	(0.056)	(0.086)
Being in a similar agricultural group	0.081	0.126	0.275	0.075	0.231	0.183
	(0.167)	(0.134)	(0.168)	(0.229)	(0.179)	(0.198)
Village Size	-0.014	-0.013^{*}	-0.008	-0.010	-0.008	-0.011^{***}
	(0.009)	(0.007)	(0.005)	(0.009)	(0.009)	(0.004)
Treatment (isolates=1)				1.124	1.101	-0.000
				(1.315)	(1.179)	(0.931)
Geodesic* Treatment				0.132	0.045	0.089
				(0.194)	(0.186)	(0.229)
Physical Distance* Treatment				0.377	0.570	0.502
				(0.929)	(0.960)	(0.969)
Blood Family* Treatment				0.300	0.065	0.311

				(0.259)	(0.229)	(0.373)
Field Neighbours* Treatment				0.032	0.281	-0.379
				(0.476)	(0.371)	(0.575)
Discuss Agriculture* Treatment				-0.142	-0.196	-0.579
				(0.341)	(0.273)	(0.435)
Centrality of receiver* Treatment				-0.155	-0.462	0.101
				(0.633)	(0.514)	(0.750)
Difference in centrality * Treatment				-3.152***	-2.985***	-3.277***
				(0.350)	(0.292)	(0.507)
Age of receiver * Treatment				0.002	0.003	0.006
				(0.006)	(0.006)	(0.008)
Gender of receiver * Treatment				0.419	0.760*	0.179
				(0.407)	(0.401)	(0.483)
Opposite Gender * Treatment				0.129	0.617*	0.210
				(0.288)	(0.339)	(0.262)
Migrant* Treatment				-0.257	-0.190	0.078
				(0.244)	(0.244)	(0.275)
Income Index* Treatment				0.069	0.033	0.123
				(0.170)	(0.126)	(0.153)
Difference in Income Index* Treatment				0.020	0.023	0.080
				(0.122)	(0.095)	(0.124)
Being in a similar agricultural group * Treatment				-0.039	-0.240	0.145
				(0.332)	(0.260)	(0.345)
Village Size * Treatment				-0.013	-0.014	-0.001
				(0.017)	(0.016)	(0.011)
Constant	-2.392***	-2.100***	-2.876***	-2.795***	-2.581***	-2.621***
	(0.654)	(0.535)	(0.406)	(0.713)	(0.774)	(0.554)
Observations	9784	12268	7422	9784	12268	7422

- Logistic Regressions. Robust standard errors in parentheses clustered at the village level and ambassador level." "* p < 0.10, ** p < 0.05, *** p < 0.01.