Estimating Causal Spillover Effects in Smallholder Farmer Networks in Western Kenya

Medha Uppala[1], Bruce A. Desmarais[1], and David P. Hughes[2]

[1]Center for Social Data Analytics, Pennsylvania State University
[2]Department of Entomology & PlantVillage, Pennsylvania State University

Abstract

As farmers share information on practical agricultural techniques with each other, we want to understand what demographic and social network factors lead to greater information sharing and adoption. The goal of this research is to understand and estimate how efficiently information on disease management behaviors (DMB) spreads in smallholder farmer networks. In this research, we are concerned with DMBs with respect to plant and crop disease. More formally, the goal is to causally estimate the network spillover effects of specific DMBs in smallholder farmer networks in Western Kenya. We intend to achieve this goal through a cluster randomized trial. Methods employed include re-randomization during the design phase, randomization tests to detect interference and estimating direct and indirect effect size via exposure networks.

1 Introduction

A majority of the farms (87%) around the world are family-owned, smallholder farms with 83% of them being less than 2 hectares (4.9 acres) (Lowder et al. [2016]). In Sub-Saharan Africa, many of these smallholder farms are predominantly used for subsistence farming, and used to cultivate staples like maize and cassava. Each of these crops suffers from pests, diseases and weeds that lead to loss of farm yields and income. One way to mitigate farm loses, reduce food insecurity and raise farmers out of poverty in these regions is to actively help farmers adopt sustainable and technological advanced farming techniques. Studies have shown that sustainable Disease Management Behaviors (DMB) are successful in reducing farm infection rates and increasing yield by multiple fold. DMB is referred to the behavior an individual expresses that can reduce the incidence of disease. Examples include vaccinations, avoiding sources of infection etc. However, these behaviors are not innate and need to be taught to the farmers.

Many studies in Behavioral and Agricultural Economics have shown that spread and adoption of new technologies among farmers is not immediate, straightforward or homogeneous; and farmers are influenced by many factors such as education, learning level, credit availability, social norms, short term risks and incomplete information [Pannell and Zilberman, 2020; Foster and Rosenzweig, 2010]. Research has also shown that adoption is higher through social network channels, but could be improved by choosing the right extension agents, social leaders and lead farmers to spread the information [Banerjee et al., 2013; Benyishay and Mobarak, 2019; Beaman et al., 2018].

Many farmers in Kenya belong to pre-organized farmer groups; the farmers in each group elect a *lead* farmer (LF) to take up the responsibility of educating and demonstrating agricultural techniques to the rest of the *peer/following* farmers (FF). These groups provide a social network structure that the farmers can rely on to seek advice, knowledge and partnership to improve their farm yields. This research proposes to take advantage of these pre-existing farmer groups in Western Kenya to explore their effectiveness in information spread and DMB adoption. The goal of this work is to harness the network dynamics and block structure of farmer groups to study how effectively lead farmers cascade information and encourage adoption in the rest of the farmer network. This research also intends to identify which of the two treatment types implemented is best suitable to increase farmer knowledge and behaviour change.

With the partnership of PlantVillage at Pennsylvania State University, a cluster randomized study is performed in two counties of Busia and Bungoma in Western Kenya during the long rains season of March to August 2021. The experiment has two treatment types where under treatment A only the lead farmer of a group receives SMS messages on DMBs while under treatment B all the members of a farmer group receive the same SMS messages. See section 4.3 for greater detail on treatment types. Section 2 provides details on what motivates this research, and below is a short introduction to PlantVillage and its work in Kenya.

1.1 PlantVillage

PlantVillage (`https://plantvillage.psu.edu/`) is a research and development unit of Penn State University, that operates with the primary goal of empowering small holder farmers in Africa. This is achieved by providing farming knowledge through affordable, usable technology to smallholder farmers to help increase their farm yield. PlantVillage has built and provided public access to an AI application called NURU that identifies, in real time, multiple diseases on cassava, African maize, wheat and potato crops, and in turn provides advice on disease management techniques. The application is being expanded to include forecasts and advice on climate stressors. To increase access to this application, PlantVillage has and continues to distribute free smartphones equipped with Nuru to smallholder farmers in Kenya. More recently, PlantVillage–in partnership with the Food and Agriculture Organization of the United Nations (FAO)–built the eLocust3M application to track and model the 2020 locust swarms in Kenya and other East African countries. Besides its presence in Kenya, PlantVillage is expanding its ground services to farmers in Tanzania and Burkina Faso. PlantVillage maintains continuous partnerships with governments and organizations to increase food security and climate resilience in smallholder farmers.

2 Motivation

Motivation for this research is two fold, with the first reason emerging from PlantVillage's role as a Research and Development team based in a land-grant university[1]. PlantVillage has established itself to pursue the goal of supporting smallholder farmers through technology and modern sustainable farming practices. The best way to continue this work is to maximize knowledge spread and increase smart behavior adoption among smallholder farmers. Given the network dynamics and block structure of the existing farmer groups in Kenya, PlantVillage is interested in understanding how lead farmers cascade information to the rest of the farmer network. A thorough understanding of how well lead farmers spread information, and identifying the social network factors that influence spillover can help PlantVillage reach and educate more farmers. This knowledge will also inform how best to support these farmer groups in the future. PlantVillage is also interested in exploring the more scalable option of communicating with farmers via SMS rather than smartphones as most farmers already own personal feature phones. Moreover, as mobile penetration is growing in low and middle income countries, digital agricultural advisory services via SMS and IVR (Integrated voice response) systems will become increasing important to reach the 2 billion people living in smallholder farming households around the world [Fabregas et al., 2019].

The second motivation behind this research is to develop an innovative approach to the design of field experiments for identifying 'interference' or 'spill-over' effects in a social network. One of the main assumptions in causal inference is SUTVA (Stable Unit Treatment Value Assumption). This assumption

[1] A land-grant university in the United States is a federally endowed university, historically established to focus on teaching agriculture, science and engineering

implies that there is no interference between units and all units have the same version of each treatment level; that is treatment level of one unit does not affect the outcome of another unit, and the potential outcomes of treatments are well defined. However, most social processes involve some form of interference between units (farmers in our case), making causal inference challenging.

Causation in social networks takes many forms, with the main ones being spillover, homophily and confounding environmental factors. Spillover, which is the focus of this research proposal, is also referred to as social influence, contagion, peer effects or network effects, depending on the field of study. It is the effect of a unit's treatment, and in turn, outcome on its connected or unconnected neighbors' outcomes (See Aronow et al. [2021] for a discussion on how spillover subtly differs from contagion in terms of statistical inference). The network structure between the lead farmer and the following farmers sets the structural stage upon which knowledge and behaviors are shared and adopted. Taking advantage of this existing block network structure, we are interested in causally estimating spillover when just the lead farmer is treated versus the spillover when everyone is treated. This way, we are able to enhance PlantVillage's understanding of farmer interaction while also pursuing a methodologically challenging, causal problem.

3 Previous Work

3.1 Agricultural Economics Literature

There is extensive literature on the adoption of new technologies in the fields of Behavioral Economics and Agricultural Economics [Streletskaya et al., 2020]. These fields adopt an econometric outlook towards modeling farmer choices in terms of risk, utility and gains. A large portion of this literature employs risk utility models and target input models where a farmer's input choices are modeled against the optimal level of input and the utilities they produce. In addition, there is a thriving subfield of econometricians modeling farmer learning and choices through a social network perspective. Banerjee et al. [2013], Beaman et al. [2018] and Benyishay and Mobarak [2019] are a few studies that use farmer social networks to study the spread and adoption of new technologies.

Banerjee et al. [2013] present a model of diffusion of microfinance in households in 43 villages in South India by using a set of village leaders as information injection points. They adopt the method of simulated moments to estimate network diffusion parameters and conclude that adopters of microfinance are seven times more likely than non-adopters to inform other households of microfinance. However, they also show that informed households are not more likely to actually adopt microfinance even when their neighbors adopt microfinance. This raises the question of interplay between homophily and contagion in these households. The authors do not differentiate between the processes of homophily and contagion but term this interplay as 'endorsement effects'. They show that there is no strong evidence of endorsement effects, that is, a neighbor's choice to adopt microfinance does not influence a household's own adoption. Overall, neighbors play a large role in spreading information on microfinance but, whether they themselves are adopters or not, do not really influence adoption in others.

Benyishay and Mobarak [2019] outline a study that runs an RCT on farmers in Malawi with three types of actors that share information on new agricultural technologies. The three actors include pre-existing government extension agents, lead farmers (LF) chosen from the leadership of the village and 5 following farmers (FF) who are closer to the average farmer in the village. This study also incorporates incentives to the actors based on their performance of sharing knowledge in the village. The new technologies introduced to farmers in this study are pit planting and Chinese composting. The authors show that without incentives, the LFs and FFs rarely adopt the technologies nor do they communicate the information to other farmers. However, with incentives, FFs experiment and communicate the technologies at a higher rate while the LFs are not responsive. They conclude that while social learning is an important way to increase technology adoption, low-cost incentives have the power to increase adoption rates. It should be noted that in this study, the LFs and FFs were chosen right before the start of the experiment, similar to the setup in Banerjee et al. [2013], indicating that leadership and new responsibilities were being required of the new LFs. However, in our case farmer groups are already formed and been functioning as social units for quite some time (at least 2 years). This evolved social structure could provide the advantage of stronger social ties among the farmers, allowing for quicker information spread and adoption in our RCT.

Another study conducted on farmers in Uganda introduced in Shikuku et al. [2019] also explores the power of disseminating farmers to share information on climate smart agriculture. The disseminating farmers are trained and randomly assigned to no incentive, private incentive and public incentive groups. The authors conclude that incentives of social recognition work better than private, individual incentives to increase information diffusion. They also recognize that extrinsic and intrinsic incentives work against each other if the net benefits of the new technology are ambiguous. The study explores how issues of acclaim and altruism affect information diffusion when the disseminating farmers are chosen randomly from the village population. In contrast we are using the network embedded information of farmers' social standing and popularity to estimate rates of information diffusion.

Finally, Beaman et al. [2018] present a study that adopts various 'threshold' models[2] of diffusion to identify optimal seed farmers in 200 Malawi villages. Four different diffusion models are randomly assigned to the villages as four separate treatment levels, and their corresponding seed farmers are estimated on the village's social network data. They show that central seeds with high degree are more successful in disseminating information on pit planting[3]. But the gain of network-theory based seed selection is only significant when the technology adoption is complex. That is, only when the technology requires a complex learning environment is it critical to identify the popular, central seeds to successfully infuse and diffuse knowledge. If the technology is simple enough to be understood by the farmers quickly, dissemination might not improve drastically with central, high degree seeds. The authors also acknowledge that a network-theory based seed selection is costly and not straightforward to implement (as it requires social network data before hand) as policy. A quicker, more feasible way is to identify the one or two central individuals most villagers seek out for agricultural advise. Most often, these particular individuals prove to be central, high degree nodes appropriate to be seed farmers.

Our proposed study benefits from the natural set up of existing farmer groups where the *lead* farmers have adopted and settled into centrally connected roles over time. While our study setup does not allow us to explicitly compare new and old *lead* farmers, we do believe that the long-standing social structures in the farmer groups provides a stronger framework for trusted social learning.

3.2 Statistics Literature

While the literature introduced so far is predominantly econometric, we will pursue this research from a perspective of causal inference on social networks. Causal analysis on social networks is inherently tricky as treatment units (farmers in our case) interact and influence each other, violating SUTVA. Other limiting factors of social networks could include assumptions around dyadic independence and non-static network structures. In addition, limited information from cross-sectional network data makes it difficult to disentangle homophily and contagion in social networks. However, a randomized controlled trial is the best available tool to make causal statements on treatment and spillover effects.

Over the last decade, there has been growing interest in causal inference methods that deal with the presence of interference, and some of these works directly inform this research.

Research in this area begins with attempts to relax the assumption of no interference and much of it is motivated by the study of infectious diseases. Halloran and Struchiner [1995]; Hudgens and Halloran [2008] propose individual level causal estimands in the present of interference by making potential outcomes of an individual be dependent on the vector of treatment assignments of all other units. However, the number of potential outcomes and treatment assignment vectors gets out of control with the number of

[2] A linear threshold model posits that an agent adopts a behavior once at least λ of her connections adopt the same behavior.

[3] Pit planting requires planting seeds in a shallow pit to retain greater moisture in arid regions while minimizing soil disturbance. This technique involves initial labor costs, and is also antithetical to the traditional practice of ridging in Malawi.

individuals and treatment levels. A simplifying assumption to combat this is *partial interference* [Sobel, 2006]; here units are part of smaller groups and interfere only with units within their group. *Partial interference* allows for within group interference while assuming between group independence. Our study setting does assume partial interference such that farmers interfere with individuals in their own group and do not interfere with those in other groups (more on this in Section 5.1).

Works in the field either focus on causal inference on observational social networks [Toulis and Kao, 2013; van der Laan, 2014; Sofrygin and van der Laan, 2016], or on inference in an experimental set up [Rosenbaum, 2007; Ugander et al., 2013; Eckles et al., 2017; Aronow and Samii, 2017]. There is also literature that focuses on developing randomization tests for interference [Bowers et al., 2013, 2016; Basse et al., 2019] (include Athey2018).

Please see Halloran and Hudgens [2016] and Aronow et al. [2021] for a comprehensive review of current methods in the causal inference with interference literature.

Come back to this section.

4 Research Design

4.1 Data

The initial survey of the farmers is to collect covariate and social network information from the farmer groups. This allows us to implement a thorough randomization design that blocks and balances on the relevant variables. The covariate data, collected at the farmer level, includes farmer, household and farm characteristics. The social network data, also collected at the farmer level, maps the relationships between the members in the farmer group. More specifically, each farmer is asked about their knowledge of and interaction with every other member in their group. Ties between farmers indicate different forms of interactions or relationships (such as family ties, common social/agricultural groups, church group etc), and these relationship types are recorded for every dyad. This produces a directed, weighted sociomatrix for each of the farmer groups. We made sure to choose farmer groups that are geographically far apart and do not suspect interference *between* two or more farmer groups.

Data is collected from 40 farmer groups from the two counties of Busia and Bungoma in Western Kenya. Each group has a single *lead* farmer and anywhere from 5 to 17 following farmers. The study comprises of 410 farmers in total.

The second and the final survey will collect data on whether the farmers learned and applied the DMBs on their farms. Due to the nature of the DMBs and disease incidence on the farms, it is not straightforward to confirm behaviors on the farms. So, most of the analysis will rely on self-reported data. However, for one of the DMBs that introduces the use of parasitoids to control Fall Army Worm infection in Maize, we did ask the farmers to place themselves on a waitlist (via phonecall) to receive the parasitoids free of cost. This list is an added source to the final survey, confirming real interest and intiative on behalf of the farmers.

4.2 Rerandomization and Covariate Balance

The farmer groups in Busia and Bungoma formed much before PlantVillage began serving these counties. In the contest of long-standing farmer groups in this region, it is important to understand how the group unit functions in the realm of knowledge dissemination and hence the group is chosen as the treatment unit. As a result, the treatment unit is the farmer group while the measurement unit is the individual farmer (giving rise to a cluster randomized trial).

PlantVillage has also served the farmers of Busia county (2 years) longer than those of Bungoma county (4 months) at the time of the study, and hence the treatment assignment is blocked by county.

As we had access to farmer covariates before treatment assignment, rerandomization [Morgan and Rubin, 2012] was implemented to achieve the highest possible balance between the relavant covariates.

Rerandomization allows the experimenter to improve covariate balance in a randomized experiment by pre-specifying a covariate balance criterion; this procedure allows one to re-randomize the treatment vector until the best covariate balance (determined by the criterion) is achieved. The balance criterion used in this study, presented in [Hansen and Bowers, 2008], allows for blocking, and takes into account the cluster and individual level values. The balance criterion for a single covariate is:

$$d(\mathbf{z},\mathbf{x}) = \left(\sum h_b \bar{m}_b\right)^{-1} \cdot \left[\sum_{b=1}^{B} \mathbf{Z}_b^t \mathbf{x}_b - \sum_{b=1}^{B} n_{tb}(\mathbf{1}^t \mathbf{x}_b / n_b)\right],$$

where block $b = 1, \ldots, B$, within which n_{tb} of n_b clusters (farmer groups) are chosen for treatment while the rest are assigned to control. **z, x** and **m** represent the n-vectors of treatment, total of x-values and cluster size of each assignment unit. Finally, $h_b \bar{m}_b = \bar{m}_b n_{tb}(1 - n_{tb}/n_b)$ represents the optimal block weights taken to be proportional to the product of block's mean cluster size and the harmonic mean of n_{tb} and $n_b - n_{tb}$.

The balance criterion used in our study, used to balance k covariates simultaneously is:

$$\begin{aligned} d^2(\mathbf{z}; \mathbf{x}_1, \ldots, \mathbf{x}_k) :=& [d(\mathbf{z}, \mathbf{x}_1), \ldots, d(\mathbf{z}, \mathbf{x}_k)] \\ & \cdot \left\{ \operatorname{Cov}\left(\begin{bmatrix} d(\mathbf{Z}, \mathbf{x}_1) \\ \vdots \\ d(\mathbf{Z}, \mathbf{x}_k) \end{bmatrix} \right) \right\}^{-} \begin{bmatrix} d(\mathbf{z}, \mathbf{x}_1) \\ \vdots \\ d(\mathbf{z}, \mathbf{x}_k) \end{bmatrix}, \end{aligned} \tag{1}$$

where $\operatorname{Cov}(d(\mathbf{Z}, \mathbf{x}), d(\mathbf{Z}, \mathbf{v})) = (\sum h_b \bar{m}_b)^2 \cdot \sum_{b=1}^{B} h_b \bar{m}_b \frac{s(\mathbf{x}_b; \mathbf{v}_b)}{\bar{m}_b}$ and M^- is a generalized inverse of M. It is shown that the test statistic in (1) roughly follows a χ^2 distribution assuming $d(\mathbf{Z}, \mathbf{x}_1), \ldots, d(\mathbf{Z}, \mathbf{x}_k)$ are approximately normal. This test statistic also balances on all linear combinations of the covariates.

With this balance criterion (that is χ^2 distributed) bounded at the lower 5% significance level, the 40 farmer groups were blocked by county and balanced on 11 relevant farmer covariates. This resulted in 7 out of the 14 farmers groups in Bungoma and 13 out of the 26 farmers groups in Busia begin assigned to treatment 1 and the rest to treatment 2.

Note that while the treatment wings in the study are equal in size, the balance criteria of the wings are not equal(check again), possibly resulting in a biased average treatment effect estimate (Theorem 2.1 in Morgan and Rubin [2012]).

4.3 Treatment Content & Types

The choice of DMBs as treatment is chosen with close collaboration with the PlantVillage team. Careful consideration of treatment kind is necessary as some DMBs call for roguing[4], and such treatment uptake(adoption) by individual farmers is highly heterogenous depending on how reliant the farmer is on the crop.

To control for this adoption heterogeneity, three DMBs that are relatively risk free were chosen as treatments. The first one encourages the use of mulch in the farms to reduce soil erosion and moisture loss. As about 92% of all the study farmers were growing maize as their main crop and Fall Army Worm (FAW) is already known to infect maize in East Africa, it was chosen as the focus of the next two DMBs. The second DMB encourages FAW scouting in the farm, where the farmer is required to look for FAW infection and destroy the egg masses. Simple, cost effective methods such as using soap solution or neem oil to destroy the egg masses is encouraged and the use of pesticides is actively discouraged. The final DMB introduces the idea of parasitoids to combat FAW infection; the farmers are informed about the wasp parasitoid that attacks and feeds on the FAW egg masses without causing any harm to the crop. Farmers are encouraged to reach out to PlantVillage for a free supply of parasitoid eggs to use on their

[4]In agriculture, roguing refers to identifying and removing(uprooting) plants with disease or undesirable characteristics.

farms. These three DMBs and their advantages are communicated via SMS messaging to the farmers every weekday for 10 weeks during the long rains planting season. Note that while the first two DMBs may be familiar to some of the study farmers, the idea of parasitoids is completely novel to the study farmers.

Given the SMS content, we decide on two treatment types as follows:

Treatment 1 (T1): Under treatment 1, *only* the lead farmer will receive SMS messaging about the various DMBs from PlantVillage during the study period. That is, the treatment saturation in the farmer group is limited to just the lead farmer.

Treatment 2 (T2): Under treatment 2, *all* the farmers in the group will the same SMS messaging about the various DMBs from PlantVillage. The treatment saturation in these farmer groups is a 100%.

We deliberately refer to the treatment types as T1 and T2 to avoid confusion regarding the presence of a true control wing in the experiment. PlantVillage is interested in the comparison between T1 and T2 instead of their effect against a true control of no treatment; the status quo of the farmers before the study is that of a true control where most farmers in this region are unaware of sustainable practices that will prepare them for climate change. Need to expand more?

5 Methodology

This section outlines the assumptions we make regarding the spillover mechanism within the farmer groups and also outlines the tests and methodology we intend to implement on the final survey data.

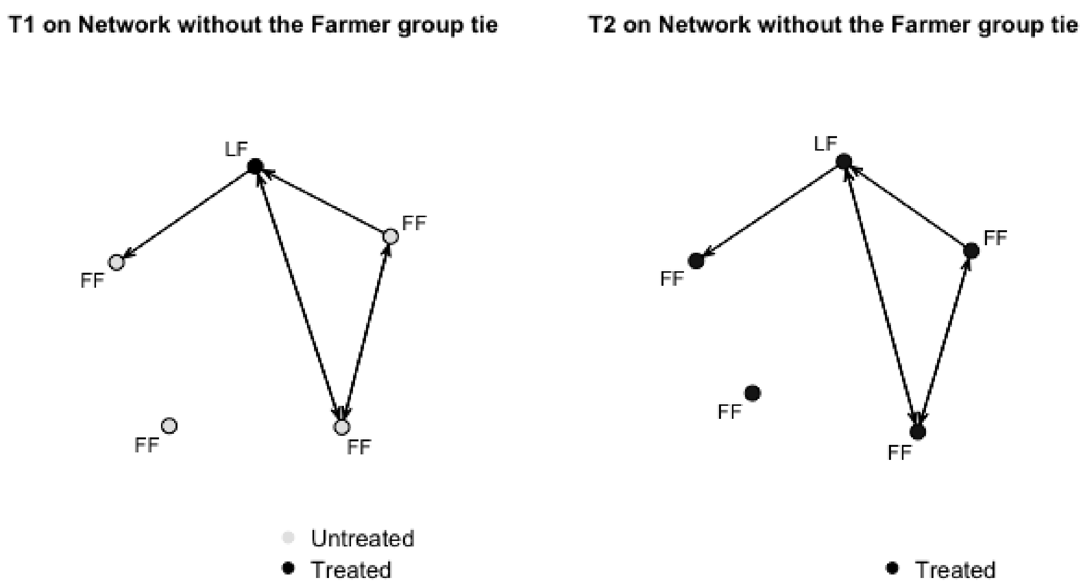
5.1 Assumptions

1. **Partial Interference**: We assume that interference between units only occurs within the farmer group and not between farmer groups. While this may not be true for farmer groups in the same village, we made sure to choose farmer groups that are geographically far apart to adhere to this assumption.

2. **Non Farmer Group ties**: Currently, every farmer group's social network is a completely connected component as every member knows every other member through the minimum context of a farmer group. We know that farmer groups only meet 1 to 3 times a month, and the LFs try to visit every FF once every couple weeks. We also know that LF visitation to FF farms is highly variable and some do check in on their FFs via phonecalls. Besides the farmer group, members know and interact with each other in alternate contexts including family, church, table banking groups etc. Most farmer group networks reveal smaller clusters when the group tie type is eliminated (or when the other tie types are highlighted), revealing that subsets of members interact with each other more than the 1-3 times they meet as a group. We believe that while the farmer group is a source of new information and advice, farmers are more likely to pay heed to information that is circulating predominantly in their immediate social groups. As a result, we think that ties outside of the farmer group are of greater importance for spillover; we believe that the activities outside the farmer group are also of higher frequency, such as family and neighbors. This assumption will appear in the methodology where we only weigh the non farmer group ties as the channels for spillover in the network and disregard the farmer group tie. Farmers have anywhere from 1 to 3 other types of ties besides the farmer group tie.

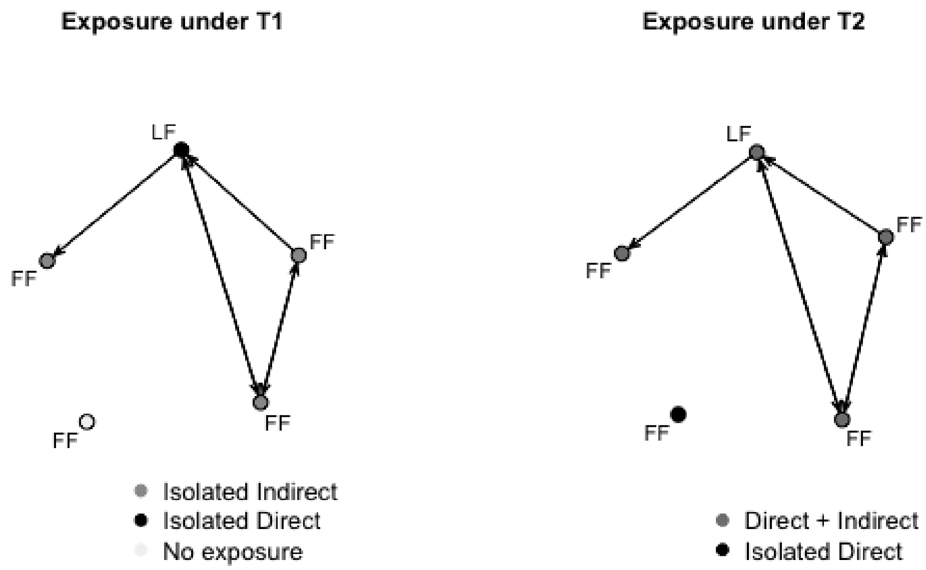
3. **Directionality**: The way the social network data was collected, farmers were asked to nominate individuals they know and interact with. As a result, the adjacency matrix currently represents a directed network. However, there is near complete reciprocity of ties in all the farmer groups. As a result, we will work with undirected versions of the networks for modelling. As we only know the different contexts farmers interact in and not the nature of those interactions, we do not want to impose a direction of influence within the farmers by using a directed network model. While a very

small number of ties in the networks are strongly one-sided (eg. Identifying a Village Elder), most of the tie contexts represent a membership form (eg. family, clan, table banking or church group, neighbor etc.). So the assumption going forward is that spillover can be modelled on the undirected networks of the farmer groups.



5.2 Methods

- **Testing**: The first method we intend to apply is the testing framework introduced in Bowers et al. [2018]. This framework allows one to test the validity of a spillover model by converting the observed data (from the two treatment wings) to that of a uniformity trial, and testing the closeness of the treatment wing samples via a Fisherian test. The two models we want to test are 1) positive spillover between individuals connected outside the farmer group and 2) positive spillover between individuals connected 2 or more ways outside the farmer group.
- **Exposure Mapping**: The second method we intend to implement is the exposure mapping framework introduced in Aronow and Samii [2017]. This method allows us to decide on a specific exposure mapping (partly outlined in the Non Farmer Group tie Assumption) given a network structure and estimate direct and indirect combination effects of the treatment from the exposure probabilities. The exposure mapping assumed when we consider all tie types excluding the farmer group tie is illustrated in figure 5.1. While the figure indicates directed ties, we intend to run the model using an undirected network. We also intend to use two different exposure mappings where cut off for positive exposure will change depending on the number of connections, just as the two models being tested above. The two mappings are 1) one or more connections outside the farmer group and 2) two or more connections outside the farmer group.



References

P. M. Aronow and C. Samii. Interference , With Application To a Social Network Experiment. *The Annals of Applied Statistics*, 11(4):1912–1947, 2017.

P. M. Aronow, D. Eckles, C. Samii, and S. Zonszein. Spillover Effects in Experimental Data. In J. Druckman and D. P. Green, editors, *Advances in Experimental Political Science*, pages 289–319. Cambridge University Press, 2021. doi: 10.1017/9781108777919.021.

A. Banerjee, A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. The diffusion of microfinance. *Science*, 341(6144), 2013. ISSN 10959203. doi: 10.1126/science.1236498.

G. W. Basse and E. M. Airoldi. Model-assisted design of experiments in the presence of network-correlated outcomes. *Biometrika*, 105(4):849–858, 2018. ISSN 14643510. doi: 10.1093/biomet/asy036.

G. W. Basse, A. Feller, and P. Toulis. Randomization tests of causal effects under interference. *Biometrika*, 106(2):487–494, 2019. ISSN 14643510. doi: 10.1093/biomet/asy072.

L. Beaman, A. Benyishay, J. Magruder, and A. M. Mobarak. Can network theory-based targeting increase technology adoption? *Economic Growth Center, Yale University*, No. 1062, 2018.

A. Benyishay and A. M. Mobarak. Social learning and incentives for experimentation and communication. *Review of Economic Studies*, 86(3):976–1009, 2019. ISSN 1467937X. doi: 10.1093/restud/rdy039.

J. Bowers, M. M. Fredrickson, and C. Panagopoulos. Reasoning about interference between units: A general framework. *Political Analysis*, 21(1):97–124, 2013. ISSN 14764989. doi: 10.1093/pan/mps038.

J. Bowers, M. M. Fredrickson, and P. M. Aronow. Research note: A more powerful test statistic for reasoning about interference between units. *Political Analysis*, 24(3):395–403, 2016. ISSN 14764989. doi: 10.1093/pan/mpw018.

J. Bowers, B. A. Desmarais, M. Frederickson, N. Ichino, H. W. Lee, and S. Wang. Models, methods and network topology: Experimental design for the study of interference. *Social Networks*, 54:196–208, 2018. ISSN 03788733. doi: 10.1016/j.socnet.2018.01.010. URL `https://doi.org/10.1016/j.socnet.2018.01.010`.

D. Eckles, R. F. Kizilcec, and E. Bakshy. Estimating peer effects in networks with peer encouragement designs. *Proceedings of the National Academy of Sciences of the United States of America*, 113(27): 7316–7322, 2016. ISSN 10916490. doi: 10.1073/pnas.1511201113.

D. Eckles, B. Karrer, and J. Ugander. Design and Analysis of Experiments in Networks: Reducing Bias from Interference. *Journal of Causal Inference*, 5(1):0–32, 2017. doi: 10.1515/jci-2015-0021.

R. Fabregas, M. Kremer, and F. Schilbach. Realizing the potential of digital development: The case of agricultural advice. *Science*, 366(6471), 2019. ISSN 10959203. doi: 10.1126/science.aay3038.

A. D. Foster and M. R. Rosenzweig. Microeconomics of technology adoption. *Annual Review of Economics*, 2:395–424, 2010. ISSN 19411383. doi: 10.1146/annurev.economics.102308.124433.

M. E. Halloran and M. G. Hudgens. Dependent Happenings: a Recent Methodological Review. *Current Epidemiology Reports*, 3(4):297–305, 2016. ISSN 2196-2995. doi: 10.1007/s40471-016-0086-4. URL `http://dx.doi.org/10.1007/s40471-016-0086-4`.

M. E. Halloran and C. J. Struchiner. Causal Inference in Infectious Diseases. *Epidemiology*, March: 142–151, 1995.

B. B. Hansen and J. Bowers. Covariate balance in simple, stratified and clustered comparative studies. *Statistical Science*, 23(2):219–236, 2008. ISSN 08834237. doi: 10.1214/08-STS254.

M. G. Hudgens and M. E. Halloran. Toward Causal Inference With Interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008. ISSN 15378276. doi: 10.1038/jid.2014.371. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf`.

D. P. Hughes and D. R. Hunter. PlantVillage NSF-EEID Proposal. Technical report.

S. K. Lowder, J. Skoet, and T. Raney. The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide. *World Development*, 87:16–29, 2016. ISSN 18735991. doi: 10.1016/j.worlddev.2015.10.041. URL `http://dx.doi.org/10.1016/j.worlddev.2015.10.041`.

K. L. Morgan and D. B. Rubin. Rerandomization to improve covariate balance in experiments. *Annals of Statistics*, 40(2):1263–1282, 2012. ISSN 00905364. doi: 10.1214/12-AOS1008.

E. L. Ogburn and T. J. VanderWeele. Vaccines, contagion, and social networks. *Annals of Applied Statistics*, 11(2):919–948, 2017. ISSN 19417330. doi: 10.1214/17-AOAS1023.

D. Pannell and D. Zilberman. Understanding Adoption of Innovations and Behavior Change to Improve Agricultural Policy. *Applied Economic Perspectives and Policy*, 42(1):3–7, 2020. ISSN 2040-5790. doi: 10.1002/aepp.13013.

P. R. Rosenbaum. Interference between units in randomized experiments. *Journal of the American Statistical Association*, 102(477):191–200, 2007. ISSN 01621459. doi: 10.1198/016214506000001112.

K. M. Shikuku, J. Pieters, E. Bulte, and P. Läderach. Incentives and the Diffusion of Agricultural Knowledge: Experimental Evidence from Northern Uganda. *American Journal of Agricultural Economics*, 101(4):1164–1180, 2019. ISSN 14678276. doi: 10.1093/ajae/aaz010.

M. E. Sobel. What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *Journal of the American Statistical Association*, 101(476):1398–1407, 2006. ISSN 01621459. doi: 10.1198/016214506000000636.

O. Sofrygin and M. J. van der Laan. Semi-Parametric Estimation and Inference for the Mean Outcome of the Single Time-Point Intervention in a Causally Connected Population. *Journal of Causal Inference*, 5(1), 2016. ISSN 2193-3677. doi: 10.1515/jci-2016-0003.

N. A. Streletskaya, S. D. Bell, M. Kecinski, T. Li, S. Banerjee, L. H. Palm-Forster, and D. Pannell. Agricultural Adoption and Behavioral Economics: Bridging the Gap. *Applied Economic Perspectives and Policy*, 42(1):54–66, 2020. ISSN 2040-5790. doi: 10.1002/aepp.13006.

E. J. Tchetgen Tchetgen and T. J. Vanderweele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, 2012. ISSN 09622802. doi: 10.1177/0962280210386779.

P. Toulis and E. Kao. Estimation of causal peer influence effects. *30th International Conference on Machine Learning, ICML 2013*, 28:2539–2547, 2013.

J. Ugander, B. Karrer, L. Backstrom, and J. Kleinberg. Graph cluster randomization: Network exposure to multiple universes. In *KDD '13: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 329–337, Chicago, 2013. ISBN 9781450321747. doi: 10.1145/2487575.2487695.

M. J. van der Laan. Causal Inference for a Population of Causally Connected Units. *Journal of Causal Inference*, 2(1):13–74, 2014. doi: 10.1111/mec.13536.Application.