Network-Based Targeting with Heterogeneous Agents for Improving Technology Adoption

Aranya Chakraborty

Ahmedabad University

Ahmedabad University

Motivation

- Low adoption of modern technologies in developing countries (Foster and Rosenzweig, 2010).
- One of the key reasons: information constraints (Magruder, 2018).
- Social networks can facilitate technology adoption by improving diffusion (Foster and Rosenzweig, 1995).
- Most effective use of social ties to improve diffusion?
    - Network-based targeting vs. random seeding. (Akbarpour et al., 2020)
    - For network-based targeting, seed agents *solely* based on their positions in the network. (Beaman et al., 2021)
    - **Key Assumption:** The diffusion depends *only* on the agents' positions in the network.

This Study

If agents differ in the benefits of a new technology and this heterogeneity affects the diffusion of information:

- Can we still use network-based targeting to improve diffusion?
- Recommended network-based targeting strategies still optimal? If not, what works better in such a scenario?

To answer these questions:

- Theoretically model agents learning about heterogeneous benefits from each other.
- Use simulations to characterize the outcomes of different targeting strategies.
- Test predictions using data on the diffusion of pit planting in Malawi.

Contributions

1. Using networks to improve technology adoption
Banerjee et al. (2013, 2019), Beaman et al. (2021)

- Evidence that the success of network-based targeting strategies depend on the population level heterogeneity.

2. Effect of population heterogeneity in social learning
Munshi (2004), Conley and Udry (2010)

- Formalize agents learning from their network about a technology having heterogeneous benefits.

3. Characterizing opinion leaders in diffusing new knowledge
Feder and Savastano (2006), Maertens (2017)

- Based on population heterogeneity, characterize opinion leaders in network-based targeting.

Elements of the Model

- Risk-neutral and myopic households.
- Two stage decision process: first learning, then adoption.
- Traditional technology has a sure payoff of π^T, where the new technology provides a payoff of $\pi^N(\omega_{it})$, $\omega_{it} \in \Omega$.
- Draws depend on the true distribution $p_i^*(\omega_{it})$ for household i. Independent draws every period.
- Uninformed households $\Rightarrow$ p_i^*s are unknown. Need to be fully informed (know p_i^*) before adoption.
- If uninformed, can become informed by putting effort $e_{it} \in \{0, 1\}$ at cost η_i.
- Costly effort: network ties help make this decision.
- Networks are assortative: $G_{ij} \neq 0$ if $|p_i^* - p_j^*| < \delta$. Example

Two-step adoption decision

1. Households decide whether or not to get informed, based on the following rule:

$$e_{it} = \begin{cases} 1 \; if \;\; \int_{\omega_{it} \in \Omega} \hat{p}_{it}(\omega_{it}) \pi^N(\omega_{it}) - c_i - \pi^T \geq \eta_i \\ 0 \; otherwise. \end{cases}$$

2. Conditional on being informed, they decide whether or not to adopt the new technology:

$$Adopt_{it} = \begin{cases} 1 \; if \;\; \int_{\omega_{it} \in \Omega} p_i^*(\omega_{it}) \pi^N(\omega_{it}) - c_i \geq \pi^T \\ 0 \; otherwise. \end{cases}$$

▸ Full Model

Timeline of Decisions

1. At each t, uninformed household i decide whether or not to get informed.
2. To decide, they collect information on beliefs (p_{jt-1}) from their peers $j \in \mathcal{I}$, formed in the last period. Household i use DeGroot averaging to calculate $\hat{p}_{it} = \sum_{j \in \mathcal{I}} G_{ij} p_{jt-1}$.
3. Based on $\hat{p}_{it}$, they decide whether or not to become informed.
4. If not informed $(e_{it} = 0)$: $p_{it} = \hat{p}_{it}$, and next period repeat from 1. If informed $(e_{it} = 1)$: p_i^* is known and adoption decisions are made based on that, and $p_{is} = p_i^* \; \forall s \geq t$.

Implications

- Let's simplify: $\Omega = \{\omega_H, \omega_L\}$ and $p^*_{iH} := p^*_i(\omega_H)$.
- In step 2 the household will adopt the new technology iff:

$$p^*_{iH} \geq \frac{c_i + (\pi^T - \pi^N(\omega_L))}{(\pi^N(\omega_H) - \pi^N(\omega_L))} = \bar{p}^*_{iH}.$$

- In step 1 the household i will choose to get informed at time t iff:

$$p^H_{it} \geq \bar{p}^*_{iH} + \frac{\eta_i}{(\pi^N(\omega_H) - \pi^N(\omega_L))} = \bar{p}^*_{iH} + \bar{\eta}_i.$$

- Under efficient diffusion of information:

$$p^*_{iH} \geq \bar{p}^*_{iH} + \bar{\eta}_i.$$

- Multiple possible equilibria: depends on the initial beliefs.
- If everyone is uninformed and $p^H_{it} \approx 0 \; \forall it$, can network-based targeting help?

METHODOLOGY

Simulate networks:

- Homogeneous non-assortative networks: $p^*_{iH} = p^*_H, \ \forall i \in \mathcal{I}$.
- Heterogeneous networks: p^*_{iH}s vary:
  - **Non-assortative:** G_{ij}s are not formed on the basis of p^*_{iH}s.
  - **Assortative:** G_{ij}s are formed on the basis of p^*_{iH}s.

Select information entry points (initially $p^H_{it} \approx 0 \ \forall it$):

- Centrality-Based
- Probability-Based
- Random

Methodology (continued)

- Let the diffusion take place for a few periods. ▸ Example
- Measure the efficiency of a targeting strategy κ:

$$Efficiency_{\kappa} = \underbrace{\frac{Informed_{\kappa}^{T}}{Informed^{T}}}_{A_{\kappa}} - \underbrace{\frac{Informed_{\kappa}^{F}}{Uninformed^{T}}}_{B_{\kappa}}$$

- Ranges between -1 and 1 (both inclusive).
- Repeat procedure for multiple networks and evaluate results *on average*.

Homogeneous vs. Heterogeneous Networks

Table 1: Efficiency Scores for Simulations using Different Targeting Strategies

		Homogeneous	Heterogeneous	
			Non-Assortative	Assortative
Targeting Strategy	Statistic	(1)	(2)	(3)
Eigenvector Centrality-Based	Mean	0.455	-0.003	0.412
	Variance	0.223	0.002	0.228
Probability-Based	Mean	0.189	-0.040	0.956
	Variance	0.125	0.023	0.004
Random	Mean	0.000	0.000	0.438
	Variance	0.000	0.000	0.228
	Observations†	239	200	200

Notes: † Simulations are done for 400 networks with homogeneous probabilities and 200 networks with heterogeneous probabilities. Upon generation of the true probabilities, some networks are dropped as they contained 0% of informed households under full efficiency. Columns (2) and (3) use the efficiency measure $Efficiency_{\kappa}$ to measure the efficiency of the targeting strategy κ. Column (1) uses the term A_{κ} of $Efficiency_{\kappa}$ for that purpose. All networks contain 30 households, and the threshold probability of learning is assumed to be 0.4 for all of them. For assortative networks, each pair of households having a success probability difference of 0.1 or less is assumed to be connected.

▸ Robustness w.r.t different centrality measure ▸ Robustness w.r.t different threshold probability of learning

Assortative Networks with Varying Heterogeneity

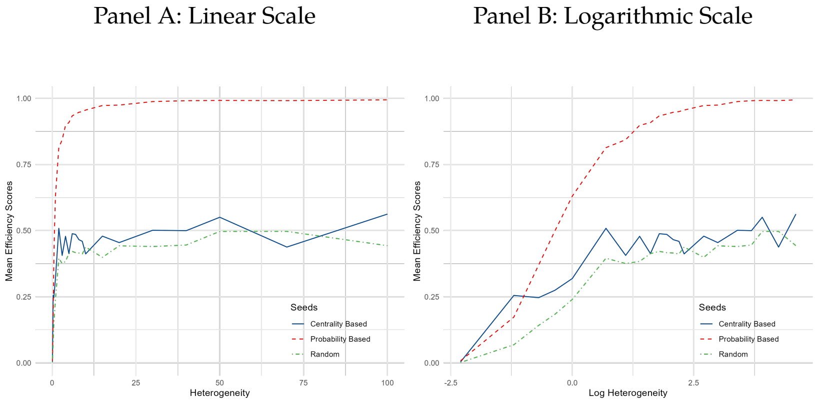


Figure 5: Efficiency scores over increasing levels of heterogeneity (with assortative networks)

▸ Robustness w.r.t different centrality ▸ Robustness w.r.t different δ ▸ Robustness w.r.t different $\bar{p}_i^H$

Hypotheses

Hypothesis 1: As the level of heterogeneity in terms of the benefits from a new technology ↑es, the success of central seeds in terms of diffusing that technology ↓es.

Hypotheses

Hypothesis 1: As the level of heterogeneity in terms of the benefits from a new technology ↑es, the success of central seeds in terms of diffusing that technology ↓es.

Hypothesis 2: As the level of heterogeneity in terms of the benefits from a new technology ↑es, the success of probability-based seeds in terms of diffusing that technology ↑es.

Data

1. Replication data for Beaman et al., 2021 (BBMM):
   - RCT to promote pit planting (PP) for maize farmers in Malawi. Randomized information entry points at the village level.
   - Panel data contains information on adoption, demographics, and network characteristics. Timeline More Details
2. Agricultural Extension Services and Technology Adoption Survey (AESTAS) data collected by International Food Policy Research Institute (IFPRI).
   - Nationally representative survey of farmers in Malawi.
   - Panel data contains information on adoption of different technologies and household demographics. More Details

Identification Using Village-level Variations:

$$Y_{vt} = \beta_0 + \beta_1 Centrality_v + \beta_2 Probability_v + \beta_3 Het_v \\ + \beta_4 Centrality_v \times Het_v + \beta_5 Probability_v \times Het_v + \lambda X_v + \zeta_t + \epsilon_{vt}$$

- Y_{vt}: adoption related outcome for village v at time t (excludes seed households).
- $Centrality_v$: average centrality of the seeds for village v at the baseline (available in the data).
- $Probability_v$: average probability of adoption for the seeds for village v at the baseline (not in the data). Approximation
- Het_v: coefficient of variation (CV) of probability of adoption at the village level.

Descriptive Statistics: Village-level Variations

Table 4: Baseline Village-level Sample Characteristics

Variable	Benchmark	Complex	Simple	Geo	Overall
	Treatment Status				
Adoption Rate (PP)	0.018	0.030	0.029	0.029	0.026
	(0.035)	(0.063)	(0.060)	(0.077)	(0.060)
Any Non-Seed Adopters (PP)	0.300	0.340	0.320	0.420	0.345
	(0.463)	(0.479)	(0.471)	(0.499)	(0.477)
Eigenvector Centrality of Seeds†	0.178	0.235	0.187	0.129	0.182
	(0.090)	(0.077)	(0.096)	(0.090)	(0.096)
Predicted Adoption Index of Seeds‡	0.110	0.114	0.101	0.082	0.101
	(0.034)	(0.036)	(0.041)	(0.025)	(0.036)
CV of Predicted Adoption Index	0.389	0.378	0.379	0.366	0.378
	(0.069)	(0.077)	(0.075)	(0.062)	(0.071)
Observations	50	50	50	50	200

Notes: † Contains 44 observations for the benchmark treatment group, 49 observations for the other treatment groups. Seed level measures are calculated using the average of two seeds, whenever the information on both seeds are available. Otherwise they reflect the information for one seed. Coefficient of Variations (CV) are calculated at the village level for the whole village. Adoption Rate and Any Non-Seed Adopters are calculated excluding seed or shadow farmers in a village.

Descriptive Statistics: Village-level Variations

Table 4: Baseline Village-level Sample Characteristics

	Treatment Status				
Variable	Benchmark	Complex	Simple	Geo	Overall
Adoption Rate (PP)	0.018	0.030	0.029	0.029	0.026
	(0.035)	(0.063)	(0.060)	(0.077)	(0.060)
Any Non-Seed Adopters (PP)	0.300	0.340	0.320	0.420	0.345
	(0.463)	(0.479)	(0.471)	(0.499)	(0.477)
Eigenvector Centrality of Seeds†	0.178	0.235	0.187	0.129	0.182
	(0.090)	(0.077)	(0.096)	(0.090)	(0.096)
Predicted Adoption Index of Seeds‡	0.110	0.114	0.101	0.082	0.101
	(0.034)	(0.036)	(0.041)	(0.025)	(0.036)
CV of Predicted Adoption Index	0.389	0.378	0.379	0.366	0.378
	(0.069)	(0.077)	(0.075)	(0.062)	(0.071)
Observations	50	50	50	50	200

Notes: † Contains 44 observations for the benchmark treatment group, 49 observations for the other treatment groups. Seed level measures are calculated using the average of two seeds, whenever the information on both seeds are available. Otherwise they reflect the information for one seed. Coefficient of Variations (CV) are calculated at the village level for the whole village. Adoption Rate and Any Non-Seed Adopters are calculated excluding seed or shadow farmers in a village.

Regression Results: Village-level Variations

Table 5: Village level Regression 1 of Adoption Outcomes (Pit Planting)

Variables	Adoption Rate (1)	(2)	Any Non-Seed Adopters (3)	(4)
Eigenvector Centrality of Seeds	1.173**	0.917*	1.181	1.235
(= *$Centrality_v$*)	(0.581)	(0.467)	(1.439)	(1.332)
Predicted Adoption Index of Seeds	-2.973**	-2.140	-8.019**	-3.344
(= *$Probability_v$*)	(1.467)	(1.318)	(3.257)	(3.233)
CV of Predicted Adoption Index	-0.296	-0.157	-0.928	0.506
(= *$Heterogeneity_v$*)	(0.208)	(0.214)	(1.079)	(1.053)
$Centrality_v \times Heterogeneity_v$	-2.625**	-2.131**	-2.851	-3.299
	(1.324)	(1.066)	(3.777)	(3.562)
$Probability_v \times Heterogeneity_v$	6.715**	4.762*	18.484***	7.562
	(3.131)	(2.796)	(6.997)	(7.073)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.080	0.180	0.049	0.169

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include a constant term and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Village-level Variations

Table 5: Village level Regression 1 of Adoption Outcomes (Pit Planting)

Variables	Adoption Rate (1)	(2)	Any Non-Seed Adopters (3)	(4)
Eigenvector Centrality of Seeds	1.173**	0.917*	1.181	1.235
(=*Centrality$_v$*)	(0.581)	(0.467)	(1.439)	(1.332)
Predicted Adoption Index of Seeds	-2.973**	-2.140	-8.019**	-3.344
(=*Probability$_v$*)	(1.467)	(1.318)	(3.257)	(3.233)
CV of Predicted Adoption Index	-0.296	-0.157	-0.928	0.506
(=*Heterogeneity$_v$*)	(0.208)	(0.214)	(1.079)	(1.053)
$Centrality_v \times Heterogeneity_v$	-2.625**	-2.131**	-2.851	-3.299
	(1.324)	(1.066)	(3.777)	(3.562)
$Probability_v \times Heterogeneity_v$	6.715**	4.762*	18.484***	7.562
	(3.131)	(2.796)	(6.997)	(7.073)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.080	0.180	0.049	0.169

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include a constant term and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Village-level Variations

Table 5: Village level Regression 1 of Adoption Outcomes (Pit Planting)

Variables	Adoption Rate (1)	(2)	Any Non-Seed Adopters (3)	(4)
Eigenvector Centrality of Seeds	1.173**	0.917*	1.181	1.235
(=$Centrality_v$)	(0.581)	(0.467)	(1.439)	(1.332)
Predicted Adoption Index of Seeds	-2.973**	-2.140	-8.019**	-3.344
(=$Probability_v$)	(1.467)	(1.318)	(3.257)	(3.233)
CV of Predicted Adoption Index	-0.296	-0.157	-0.928	0.506
(=$Heterogeneity_v$)	(0.208)	(0.214)	(1.079)	(1.053)
$Centrality_v \times Heterogeneity_v$	-2.625**	-2.131**	-2.851	-3.299
	(1.324)	(1.066)	(3.777)	(3.562)
$Probability_v \times Heterogeneity_v$	6.715**	4.762*	18.484***	7.562
	(3.131)	(2.796)	(6.997)	(7.073)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.080	0.180	0.049	0.169

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include a constant term and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Summary

Key Takeaway: Network-based targeting may require more than identifying central households within a social network.

⇒ We need to have an understanding of possible heterogeneity in benefits across households.

- Under the assumption of assortativity, simulations show that centrality (probability) based targeting perform worse (better) as heterogeneity increase.
- Empirical results show support in favor of my hypotheses:
  - Positive (negative) effect of seeds' centrality (probability) on adoption decrease with increase in village-level heterogeneity. Descriptive Figures Robustness
  - Weaker evidences in favor of my hypotheses are found using the experimental variations in the data. Identification and Results

THANK YOU!

Assortative Network

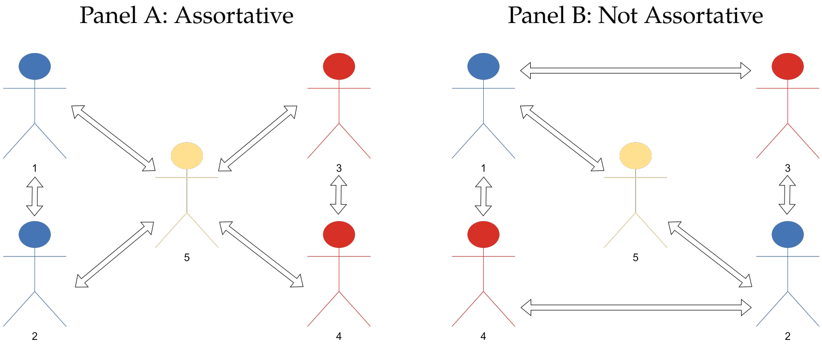


Figure 1: Networks with Heterogeneous Benefits

◂ Back

Theoretical Model

- Two-stage decision process:
  - **Stage 1:** Irreversible investment to learn about an available new technology.
  - **Stage 2:** Conditional on the investment, decide whether to stick to a traditional technology or adopt the new technology.
- Traditional technology has a sure payoff of π^T, where the new technology provides a payoff of $\pi^N(\omega_{it})$, $\omega_{it} \in \Omega$.
- ω_{it} is drawn independently at each period t according to the true distribution $p_i^*(\omega_{it})$ for household i. Draws are not correlated over time within household and between households.

Theoretical Model

- But, true distributions are positively correlated between households according to the existing network structure (more details below).
- $\forall it$, $\exists \omega_{it}, \omega'_{it} \in \Omega$ such that $\pi^N(\omega_{it}) \geq \pi^T \geq \pi^N(\omega'_{it})$.
- $\mathcal{I}$ denotes the set of all households.
- $\exists i, j \in \mathcal{I}$ such that $\int_{\omega_{it} \in \Omega} p_i^*(\omega_{it}) \pi^N(\omega_{it}) - c_i \geq \pi^T$ and $\int_{\omega_{jt} \in \Omega} p_j^*(\omega_{jt}) \pi^N(\omega_{jt}) - c_j \leq \pi^T$, with c_i being the cost of new technology for household i.
- Initially all households are uninformed $\Rightarrow p_i^*$s are unknown.
- The household i has beliefs $p_{it}(\omega_{it})$ over the distribution of ω_{it} at period t.

Theoretical Model

- At period t, uninformed household i has the option to become informed by putting effort $e_{it} \in \{0, 1\}$.
- If $e_{i\tau} = 1$, $e_{it} = 1 \; \forall t \geq \tau$.
- If $e_{it} = 1$, the household learns the true distribution $p_i^*(\omega_{it})$ at cost η_i. The cost of learning is incurred the first time the household gets informed only.
- If $e_{it} = 0$, no effort cost is incurred and the household uses DeGroot averaging to approximate the true distribution.
- Let G denote the $n \times n$ weighted, directed, and non-negative influence matrix ($n = |\mathcal{I}|$), where $G_{ij} \geq 0$ represents the weight i places on j's opinion (with $\sum_{j \in \mathcal{I}} G_{ij} = 1$).

Theoretical Model

- Then $\hat{p}_{it} = \sum_{j \in \mathcal{I}} G_{ij} p_{jt-1}$ denotes household i's approximation based on others' opinion following the DeGroot averaging.
- Networks are assortative: $G_{ij} \neq 0$ if $|p_i^* - p_j^*| < \delta$.
- The belief of household i at period t:

$$p_{it}(\omega_{it}) = e_{it}(p_i^*(\omega_{it})) + (1 - e_{it})\hat{p}_{it}(\omega_{it}).$$

- Assume that households need to be informed before they adopt: helps me explicitly capture the point when the households stop seeking information from their peers.
- Assume the households to be risk-neutral and myopic.

Back

Illustrative example

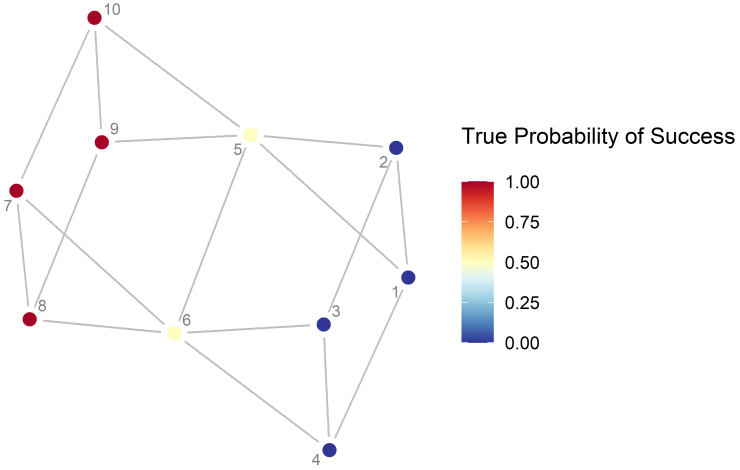


Figure 1: Distribution of True Probability within the network

Illustrative example

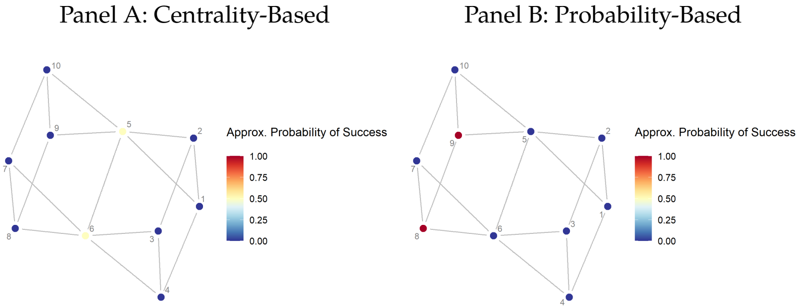


Figure 3: Initial Seeding based on Centrality and Probability

Illustrative example

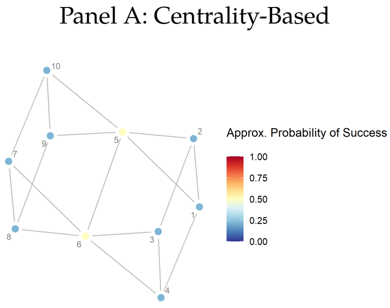


Figure 4: Performance of seeds after three periods

Back

Simulation Robustness 1

◂ Back

Table D.7: Simulation Robustness (w.r.t different centrality measure)

		Homogeneous	Heterogeneous	
			Non-Assortative	Assortative
Targeting Strategy	Statistic	(1)	(2)	(3)
Betweenness Centrality-Based	Mean	0.463	-0.010	0.635
	Variance	0.225	0.002	0.210
Probability-Based	Mean	0.189	-0.040	0.956
	Variance	0.125	0.023	0.004
Random	Mean	0.000	0.000	0.438
	Variance	0.000	0.000	0.228
	Observations†	239	200	200

Notes:† Simulations are done for 400 networks with homogeneous probabilities and 200 networks with heterogeneous probabilities. Upon generation of the true probabilities, some networks are dropped as they contained 0% of informed households under full efficiency. Columns (2) and (3) use the efficiency measure $Efficiency_\kappa$ to measure the efficiency of the targeting strategy κ. Column (1) uses the term A_κ of $Efficiency_\kappa$ for that purpose. All networks contain 30 households, and the threshold probability of learning is assumed to be 0.4 for all of them. For assortative networks, each pair of households having a success probability difference of 0.1 or less is assumed to be connected.

Simulation Robustness 2

◂ Back

Table D.8: Simulation Robustness (w.r.t $\bar{p}_i^H = 0.5$, instead of $\bar{p}_i^H = 0.4$)

		Homogeneous	Heterogeneous	
			Non-Assortative	Assortative
Targeting Strategy	Statistic	(1)	(2)	(3)
Eigenvector Centrality-Based	Mean	0.197	-0.007	0.414
	Variance	0.136	0.006	0.230
Probability-Based	Mean	0.017	-0.009	0.965
	Variance	0.008	0.012	0.003
Random	Mean	0.000	0.000	0.161
	Variance	0.000	0.000	0.129
	Observations[†]	197	200	200

Notes:[†] Simulations are done for 400 networks with homogeneous probabilities and 200 networks with heterogeneous probabilities. Upon generation of the true probabilities, some networks are dropped as they contained 0% of informed households under full efficiency. Columns (2) and (3) use the efficiency measure $\textit{Efficiency}_\kappa$ to measure the efficiency of the targeting strategy κ. Column (1) uses the term A_κ of $\textit{Efficiency}_\kappa$ for that purpose. All networks contain 30 households, and the threshold probability of learning is assumed to be 0.5 for all of them. For assortative networks, each pair of households having a success probability difference of 0.1 or less is assumed to be connected.

Simulation Robustness 3

Back

Table D.9: Simulation Robustness (w.r.t $\bar{p}_i^H = 0.3$, instead of $\bar{p}_i^H = 0.4$)

		Homogeneous	Heterogeneous	
			Non-Assortative	Assortative
Targeting Strategy	Statistic	(1)	(2)	(3)
Eigenvector Centrality-Based	Mean	0.642	-0.004	0.409
	Variance	0.218	0.008	0.224
Probability-Based	Mean	0.481	-0.031	0.948
	Variance	0.236	0.012	0.004
Random	Mean	0.018	0.003	0.469
	Variance	0.010	0.003	0.227
	Observations†	281	200	200

Notes:† Simulations are done for 400 networks with homogeneous probabilities and 200 networks with heterogeneous probabilities. Upon generation of the true probabilities, some networks are dropped as they contained 0% of informed households under full efficiency. Columns (2) and (3) use the efficiency measure $Efficiency_\kappa$ to measure the efficiency of the targeting strategy κ. Column (1) uses the term A_κ of $Efficiency_\kappa$ for that purpose. All networks contain 30 households, and the threshold probability of learning is assumed to be 0.3 for all of them. For assortative networks, each pair of households having a success probability difference of 0.1 or less is assumed to be connected.

Simulation Robustness 4

Back

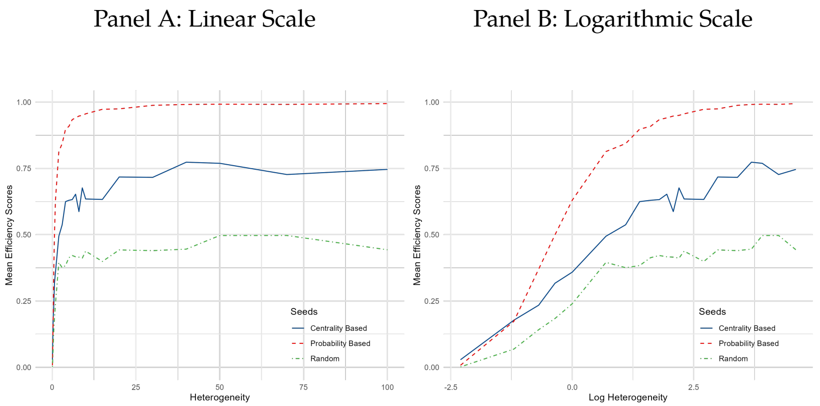


Figure 8: Efficiency scores over increasing levels of heterogeneity (with assortative networks) w.r.t betweenness centrality (instead of eigenvector centrality)

Simulation Robustness 5

◂ Back

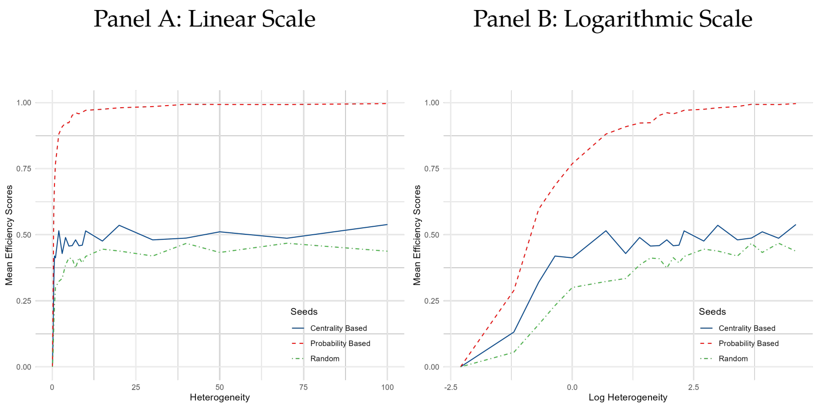


Figure 9: Efficiency scores over increasing levels of heterogeneity (with assortative networks w.r.t $\delta = 0.2$ instead of $\delta = 0.1$)

Simulation Robustness 6

Back

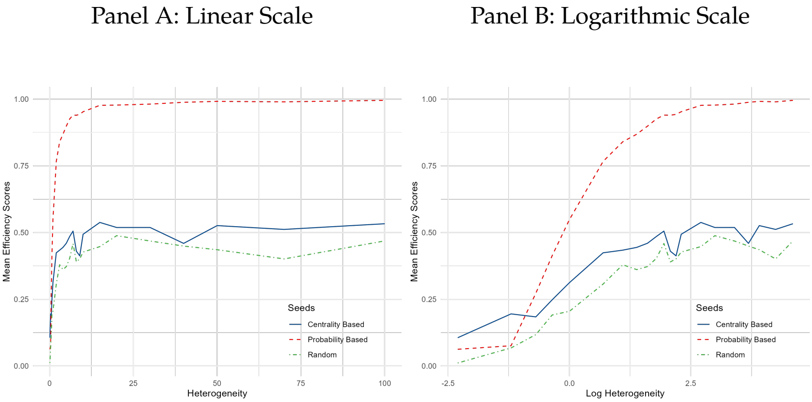


Figure 10: Efficiency scores over increasing levels of heterogeneity (with assortative networks w.r.t $\delta = 0.05$ instead of $\delta = 0.1$)

Simulation Robustness 7

◂ Back

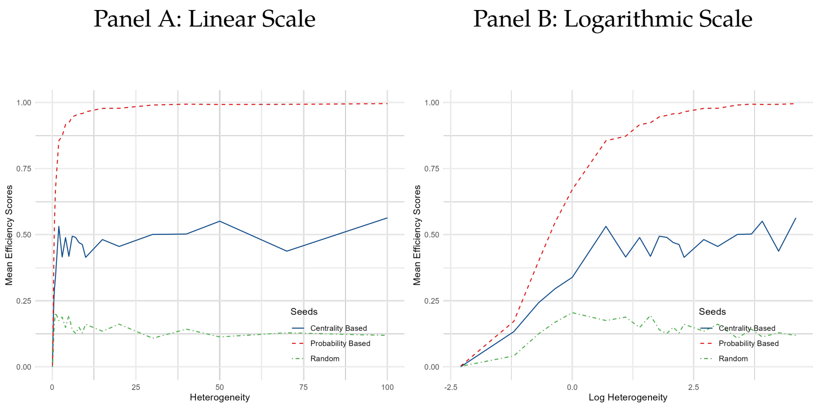


Figure 11: Efficiency scores over increasing levels of heterogeneity (w.r.t $\bar{p}_i^H = 0.5$, instead of $\bar{p}_i^H = 0.4$)

Simulation Robustness 8

Back

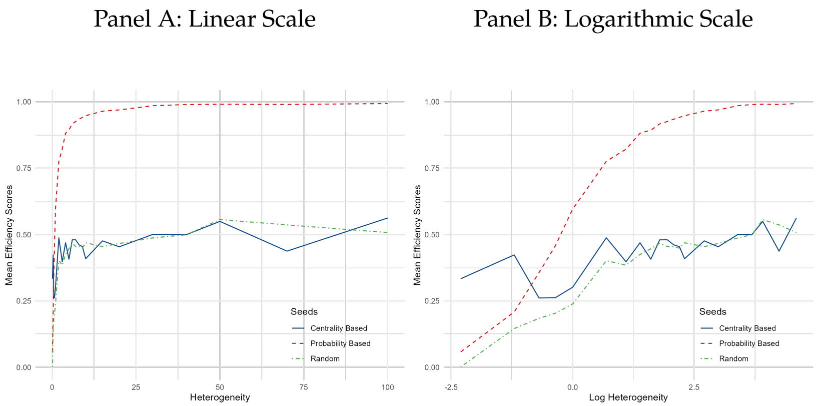
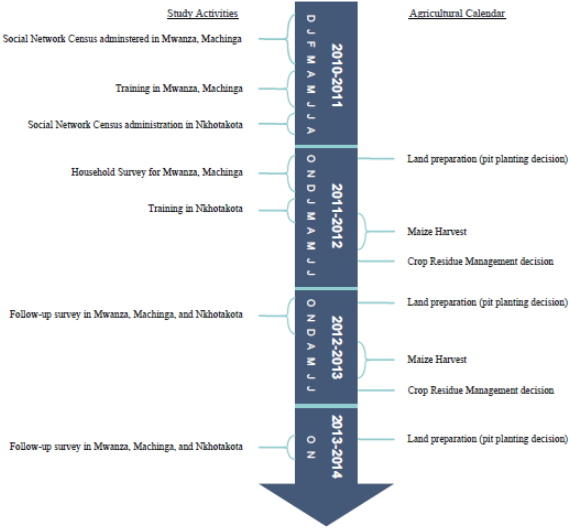


Figure 12: Efficiency scores over increasing levels of heterogeneity (w.r.t $\bar{p}_i^H = 0.3$, instead of $\bar{p}_i^H = 0.4$)

Timeline of BBMM

Back



BBMM Replication Data

◂ Back

- First collected the social network census data to elicit names of people each respondent consults when making agricultural decisions along with some other demographics.
- Used this responses with the village listing to identify links. Considered individuals linked if either party named each other or if they are part of the same household.
- Used simulations with the network information to identify seeds according to different diffusion processes to optimize diffusion after four periods.
- Randomly allocated villages to one of the four treatment groups and selected seeds for training based on that.
- Once the training is complete, randomly surveyed a panel of approximately 30 households per village, including all the seed and shadow farmers.

AESTAS Data

Back

- Objective was to monitor the *Lead Farmer* (LF) program in Malawi.
- Covers all districts of Malawi, except Likoma. Data collected in two waves: 2016 and 2018.
- Three types of interviews: Household, LF, and Community.
- Random sample of around 10 households were selected for interview from randomly selected sections within each district.
- Stratification was done based on whether or not the household had a LF.
- The same households were interviewed in the two waves with very small level of attrition (around 4%).
- For each household, both household head and their spouses were interviewed.

Approximating Probability of Adoption

Back

- How to calculate probability of adoption?
- Proxy for probability of adoption using predicted adoption index.
- Calculate the index at the baseline, conditional on household demographics: number of adults and children, housing, livestock, and assets. Description of Variables
- Calculation uses estimates from following regressions using AESTAS data: $Adoption\ Index_{it} = f(X_{it}; \mu_{it})$. Results
- Based on a set of assumptions. All Assumptions

Description of Key Demographic Variables

◂ Back

- **Adults**: Number of adults in the household.
- **Children**: Number of children in the household.
- **Housing**: Standardized first principal component (PC). Includes information on materials walls are made of, roof materials, floor materials (0- Traditional, 1- Modern), and whether the household has a toilet (only in the BBMM sample).
- **Livestock**: Standardized first PC. Includes the number of sheep, goats, chickens, cows, pigs the household owns. The BBMM sample also includes number of guinea fowl and doves.
- **Assets**: Standardized first PC. Includes the number of bicycles, radios and cell phones the household owns.

Approximating Probability: Assumptions

Back

- **Assumption 1:** Adoption and Usage indices are good proxies for the probability of adoption.
- **Assumption 2:** The variation in adoption and usage indices, conditional on the observable demographics, is sufficient for my analysis. Actual and Predicted Variations
- **Assumption 3:** The mapping of observable characteristics to the adoption probability is the same across the datasets I use in this study. Sample Comparison
- **Assumption 4:** Any bias in the estimated relationship between adoption probability and observable characteristics is independent of the unobserved village-level learning in the BBMM sample.

Sample Comparison

Back

Table 2: Baseline Demographics Across Datasets

		Variables				
Dataset	Statistic	Adults	Children	Housing	Livestock	Assets
AESTAS	Mean	2.14	3.00	-0.09	-0.03	-0.03
	(SD)	(1.00)	(2.00)	(0.98)	(0.99)	(1.00)
	Median	2.00	3.00	-0.29	-0.40	-0.29
	Observations	2820	2820	2803	2820	2820
BBMM	Mean	2.36	2.77	-0.02	0.02	0.09
	(SD)	(0.95)	(1.86)	(0.99)	(1.02)	(1.03)
	Median	2.00	3.00	-0.24	-0.31	-0.10
	Observations	5384	5407	5382	5407	5407

Notes: The variables *Adults* and *Children* represent number of adults and children in a household, respectively. The variables *Housing*, *Livestock*, and *Assets* were standardized first principal components. More details available in the paper.

Approximating Probabilities of Adoption

◂ Back

Table 3: OLS Regression Results for Adoption and Usage Indices

	Adoption Index			Usage Index		
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Adults	0.008***	0.008***	0.005**	0.011***	0.011***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Children	0.003***	0.003***	0.002	0.003***	0.003***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Housing	0.009***	0.009***	0.008***	0.003	0.003	0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Livestock	0.010***	0.010***	0.005*	0.014***	0.014***	0.009***
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Assets	0.024***	0.024***	0.017***	0.020***	0.020***	0.014***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Year Fixed-Effects	No	Yes	Yes	No	Yes	Yes
Household Controls	No	No	Yes	No	No	Yes
Observations	5610	5608	5604	5610	5608	5604
R-squared	0.096	0.096	0.150	0.085	0.131	0.169

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors clustered at the section level are in parentheses. All regressions use a constant term and sample weights. The variables *Adults* and *Children* represent number of adults and children in a household, respectively. The variables *Housing*, *Livestock*, and *Assets* were standardized first principal components.

Approximating Probabilities of Adoption

◂ Back

Panel A: Adoption Index | Panel B: Usage Index

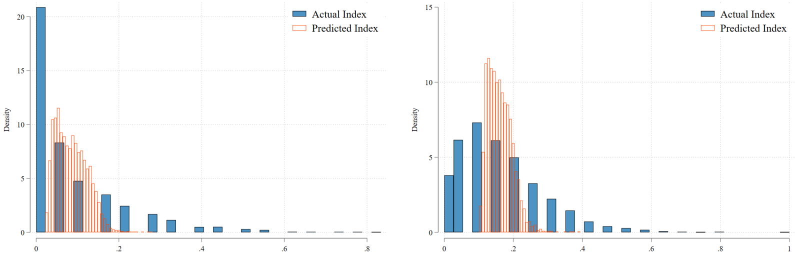


Figure 7: Actual and Predicted Adoption and Usage Indices

Descriptive Figures: Village-level Variations

◂ Back

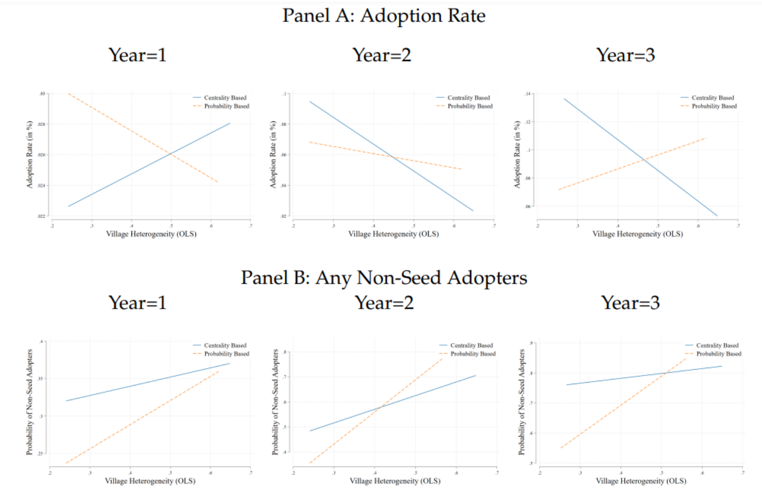


Figure 8: Outcomes for Different Seeding Strategies with respect to Village Heterogeneity

Regression Results

Table A1: Village level Regression 1 with Different Measure of Probability

Variables	Adoption Rate (1)	(2)	Any Non-Seed Adopters (3)	(4)
Eigenvector Centrality of Seeds	0.999*	0.817*	0.984	1.067
(= $Centrality_v$)	(0.565)	(0.480)	(1.302)	(1.191)
Predicted Usage Index of Seeds	-2.174	-1.511	-4.599	-0.0836
(= $Probability_v$)	(1.410)	(1.279)	(3.317)	(3.053)
CV of Predicted Usage Index	-1.091	-0.631	-2.549	2.142
(= $Heterogeneity_v$)	(0.805)	(0.779)	(2.905)	(2.823)
$Centrality_v \times Heterogeneity_v$	-4.481*	-3.936*	-4.874	-5.907
	(2.623)	(2.281)	(6.889)	(6.438)
$Probability_v \times Heterogeneity_v$	10.330*	7.276	23.13	0.889
	(6.160)	(5.623)	(14.19)	(13.40)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.063	0.174	0.037	0.164

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include a constant term and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results

Back

Table A2: Village level Regression 1 with Different Measure of Centrality

Variables	Adoption Rate (1)	(2)	Any Non-Seed Adopters (3)	(4)
Closeness Centrality of Seeds	0.609**	0.454*	0.571	0.617
(=$Centrality_v$)	(0.306)	(0.234)	(0.709)	(0.659)
Predicted Adoption Index of Seeds	-2.438**	-1.709	-7.555**	-2.904
(=$Probability_v$)	(1.230)	(1.134)	(3.201)	(3.152)
CV of Predicted Adoption Index	-0.0774	-0.007	-0.677	0.887
(=$Heterogeneity_v$)	(0.214)	(0.202)	(1.196)	(1.158)
$Centrality_v \times Heterogeneity_v$	-1.325*	-1.020*	-1.552	-1.997
	(0.716)	(0.558)	(1.896)	(1.823)
$Probability_v \times Heterogeneity_v$	5.610**	3.814	17.55**	6.849
	(2.660)	(2.439)	(6.873)	(6.940)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.087	0.179	0.048	0.170

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include a constant term and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Identification Using Experimental Variation

Back

$$
\begin{aligned}
Y_{vt} = &\psi_0 + \psi_1 Cent_v + \psi_2 Prob_v + \psi_3 Het_v + \psi_4^0 Cent_v \times Het_v \\
&+ \psi_4^T Cent_v \times Het_v \times Treat_v + \psi_5^0 Prob_v \times Het_v \\
&+ \psi_5^T Prob_v \times Het_v \times Treat_v + \gamma X_v + \rho_t + \eta_{vt}.
\end{aligned}
$$

- $Treat_v$: captures whether the village v belongs to complex, simple or geo treatment arm.
- Effects are measured in terms of the omitted category (benchmark treatment arm).
- Villages are less (or, same level of) heterogeneous in other treatment arms (compared to benchmark). That implies:
    - Y_{vt} $\uparrow$es with centrality and $\downarrow$es with probability.
    - No prediction for seeds with less centrality and probability.

Descriptive Statistics: Experimental Variations

Table 4: Baseline Village-level Sample Characteristics

Variable	Treatment Status Benchmark	Complex	Simple	Geo	Overall
Adoption Rate (PP)	0.018	0.030	0.029	0.029	0.026
	(0.035)	(0.063)	(0.060)	(0.077)	(0.060)
Any Non-Seed Adopters (PP)	0.300	0.340	0.320	0.420	0.345
	(0.463)	(0.479)	(0.471)	(0.499)	(0.477)
Eigenvector Centrality of Seeds†	0.178	0.235	0.187	0.129	0.182
	(0.090)	(0.077)	(0.096)	(0.090)	(0.096)
Predicted Adoption Index of Seeds‡	0.110	0.114	0.101	0.082	0.101
	(0.034)	(0.036)	(0.041)	(0.025)	(0.036)
CV of Predicted Adoption Index	0.389	0.378	0.379	0.366	0.378
	(0.069)	(0.077)	(0.075)	(0.062)	(0.071)
Observations	50	50	50	50	200

Notes: † Contains 44 observations for the benchmark treatment group, 49 observations for the other treatment groups. Seed level measures are calculated using the average of two seeds, whenever the information on both seeds are available. Otherwise they reflect the information for one seed. Coefficient of Variations (CV) are calculated at the village level for the whole village. Adoption Rate and Any Non-Seed Adopters are calculated excluding seed or shadow farmers in a village.

Descriptive Statistics: Experimental Variations

Table 4: Baseline Village-level Sample Characteristics

	Treatment Status				
Variable	Benchmark	Complex	Simple	Geo	Overall
Adoption Rate (PP)	0.018	0.030	0.029	0.029	0.026
	(0.035)	(0.063)	(0.060)	(0.077)	(0.060)
Any Non-Seed Adopters (PP)	0.300	0.340	0.320	0.420	0.345
	(0.463)	(0.479)	(0.471)	(0.499)	(0.477)
Eigenvector Centrality of Seeds†	0.178	0.235	0.187	0.129	0.182
	(0.090)	(0.077)	(0.096)	(0.090)	(0.096)
Predicted Adoption Index of Seeds‡	0.110	0.114	0.101	0.082	0.101
	(0.034)	(0.036)	(0.041)	(0.025)	(0.036)
CV of Predicted Adoption Index	0.389	0.378	0.379	0.366	0.378
	(0.069)	(0.077)	(0.075)	(0.062)	(0.071)
Observations	50	50	50	50	200

Notes: † Contains 44 observations for the benchmark treatment group, 49 observations for the other treatment groups. Seed level measures are calculated using the average of two seeds, whenever the information on both seeds are available. Otherwise they reflect the information for one seed. Coefficient of Variations (CV) are calculated at the village level for the whole village. Adoption Rate and Any Non-Seed Adopters are calculated excluding seed or shadow farmers in a village.

Descriptive Statistics: Experimental Variations

Table 4: Baseline Village-level Sample Characteristics

	Treatment Status				
Variable	Benchmark	Complex	Simple	Geo	Overall
Adoption Rate (PP)	0.018	0.030	0.029	0.029	0.026
	(0.035)	(0.063)	(0.060)	(0.077)	(0.060)
Any Non-Seed Adopters (PP)	0.300	0.340	0.320	0.420	0.345
	(0.463)	(0.479)	(0.471)	(0.499)	(0.477)
Eigenvector Centrality of Seeds†	0.178	0.235	0.187	0.129	0.182
	(0.090)	(0.077)	(0.096)	(0.090)	(0.096)
Predicted Adoption Index of Seeds‡	0.110	0.114	0.101	0.082	0.101
	(0.034)	(0.036)	(0.041)	(0.025)	(0.036)
CV of Predicted Adoption Index	0.389	0.378	0.379	0.366	0.378
	(0.069)	(0.077)	(0.075)	(0.062)	(0.071)
Observations	50	50	50	50	200

Notes: † Contains 44 observations for the benchmark treatment group, 49 observations for the other treatment groups. Seed level measures are calculated using the average of two seeds, whenever the information on both seeds are available. Otherwise they reflect the information for one seed. Coefficient of Variations (CV) are calculated at the village level for the whole village. Adoption Rate and Any Non-Seed Adopters are calculated excluding seed or shadow farmers in a village.

Regression Results: Experimental Variations

Table 6: Village level Regression 2 of Adoption Outcomes

Variables	Adoption Rate (5)	Adoption Rate (6)	Any Non-Seed Adopters (7)	Any Non-Seed Adopters (8)
$Centrality_v \times Heterogeneity_v$	-2.423**	-2.237**	-6.692	-6.574
	(1.093)	(0.996)	(4.503)	(4.119)
$Centrality_v \times Heterogeneity_v \times Complex$	0.657**	0.664**	4.328**	3.756**
	(0.306)	(0.282)	(1.775)	(1.664)
$Centrality_v \times Heterogeneity_v \times Simple$	0.416	0.428	1.078	0.431
	(0.337)	(0.320)	(2.060)	(1.947)
$Centrality_v \times Heterogeneity_v \times Geo$	2.026**	1.942**	0.103	-0.070
	(0.940)	(0.839)	(2.235)	(2.098)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.133	0.224	0.113	0.222

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include seed centrality, seed probability, village-level heterogeneity, a constant term, and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Experimental Variations

Table 6: Village level Regression 2 of Adoption Outcomes

Variables	Adoption Rate (5)	(6)	Any Non-Seed Adopters (7)	(8)
$Centrality_v \times Heterogeneity_v$	-2.423**	-2.237**	-6.692	-6.574
	(1.093)	(0.996)	(4.503)	(4.119)
$Centrality_v \times Heterogeneity_v \times Complex$	0.657**	0.664**	4.328**	3.756**
	(0.306)	(0.282)	(1.775)	(1.664)
$Centrality_v \times Heterogeneity_v \times Simple$	0.416	0.428	1.078	0.431
	(0.337)	(0.320)	(2.060)	(1.947)
$Centrality_v \times Heterogeneity_v \times Geo$	2.026**	1.942**	0.103	-0.070
	(0.940)	(0.839)	(2.235)	(2.098)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.133	0.224	0.113	0.222

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include seed centrality, seed probability, village-level heterogeneity, a constant term, and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Experimental Variations

Table 6: Village level Regression 2 of Adoption Outcomes

Variables	Adoption Rate (5)	(6)	Any Non-Seed Adopters (7)	(8)
$Centrality_v \times Heterogeneity_v$	-2.423**	-2.237**	-6.692	-6.574
	(1.093)	(0.996)	(4.503)	(4.119)
$Centrality_v \times Heterogeneity_v \times Complex$	0.657**	0.664**	4.328**	3.756**
	(0.306)	(0.282)	(1.775)	(1.664)
$Centrality_v \times Heterogeneity_v \times Simple$	0.416	0.428	1.078	0.431
	(0.337)	(0.320)	(2.060)	(1.947)
$Centrality_v \times Heterogeneity_v \times Geo$	2.026**	1.942**	0.103	-0.070
	(0.940)	(0.839)	(2.235)	(2.098)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.133	0.224	0.113	0.222

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include seed centrality, seed probability, village-level heterogeneity, a constant term, and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Experimental Variations

Table 6: Village level Regression 2 of Adoption Outcomes

	Adoption Rate		Any Non-Seed Adopters	
Variables	(5)	(6)	(7)	(8)
$Centrality_v \times Heterogeneity_v$	-2.423**	-2.237**	-6.692	-6.574
	(1.093)	(0.996)	(4.503)	(4.119)
$Centrality_v \times Heterogeneity_v \times Complex$	0.657**	0.664**	4.328**	3.756**
	(0.306)	(0.282)	(1.775)	(1.664)
$Centrality_v \times Heterogeneity_v \times Simple$	0.416	0.428	1.078	0.431
	(0.337)	(0.320)	(2.060)	(1.947)
$Centrality_v \times Heterogeneity_v \times Geo$	2.026**	1.942**	0.103	-0.070
	(0.940)	(0.839)	(2.235)	(2.098)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.133	0.224	0.113	0.222

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include seed centrality, seed probability, village-level heterogeneity, a constant term, and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Experimental Variations

Table 6: Village level Regression 2 of Adoption Outcomes (continued)

Variables	Adoption Rate (5)	(6)	Any Non-Seed Adopters (7)	(8)
$Probability_v \times Heterogeneity_v$	5.881**	4.104*	22.97***	12.35
	(2.437)	(2.286)	(7.720)	(7.626)
$Probability_v \times Heterogeneity_v \times Complex$	-0.155	-0.232	-1.275	-0.679
	(0.520)	(0.497)	(2.765)	(2.654)
$Probability_v \times Heterogeneity_v \times Simple$	-0.121	-0.110	1.941	3.511
	(0.642)	(0.571)	(3.572)	(3.333)
$Probability_v \times Heterogeneity_v \times Geo$	-2.588**	-2.562**	-0.391	0.538
	(1.131)	(1.039)	(4.028)	(3.618)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.133	0.224	0.113	0.222

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include seed centrality, seed probability, village-level heterogeneity, a constant term, and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Experimental Variations

Table 6: Village level Regression 2 of Adoption Outcomes (continued)

Variables	Adoption Rate (5)	(6)	Any Non-Seed Adopters (7)	(8)
Probability$_v$ × *Heterogeneity*$_v$	5.881**	4.104*	22.97***	12.35
	(2.437)	(2.286)	(7.720)	(7.626)
Probability$_v$ × *Heterogeneity*$_v$ × *Complex*	-0.155	-0.232	-1.275	-0.679
	(0.520)	(0.497)	(2.765)	(2.654)
Probability$_v$ × *Heterogeneity*$_v$ × *Simple*	-0.121	-0.110	1.941	3.511
	(0.642)	(0.571)	(3.572)	(3.333)
Probability$_v$ × *Heterogeneity*$_v$ × *Geo*	-2.588**	-2.562**	-0.391	0.538
	(1.131)	(1.039)	(4.028)	(3.618)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.133	0.224	0.113	0.222

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include seed centrality, seed probability, village-level heterogeneity, a constant term, and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Experimental Variations

Table 6: Village level Regression 2 of Adoption Outcomes (continued)

Variables	Adoption Rate (5)	(6)	Any Non-Seed Adopters (7)	(8)
$Probability_v \times Heterogeneity_v$	5.881**	4.104*	22.97***	12.35
	(2.437)	(2.286)	(7.720)	(7.626)
$Probability_v \times Heterogeneity_v \times Complex$	-0.155	-0.232	-1.275	-0.679
	(0.520)	(0.497)	(2.765)	(2.654)
$Probability_v \times Heterogeneity_v \times Simple$	-0.121	-0.110	1.941	3.511
	(0.642)	(0.571)	(3.572)	(3.333)
$Probability_v \times Heterogeneity_v \times Geo$	-2.588**	-2.562**	-0.391	0.538
	(1.131)	(1.039)	(4.028)	(3.618)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.133	0.224	0.113	0.222

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include seed centrality, seed probability, village-level heterogeneity, a constant term, and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.

Regression Results: Experimental Variations

Table 6: Village level Regression 2 of Adoption Outcomes (continued)

Variables	Adoption Rate (5)	(6)	Any Non-Seed Adopters (7)	(8)
$Probability_v \times Heterogeneity_v$	5.881**	4.104*	22.97***	12.35
	(2.437)	(2.286)	(7.720)	(7.626)
$Probability_v \times Heterogeneity_v \times Complex$	-0.155	-0.232	-1.275	-0.679
	(0.520)	(0.497)	(2.765)	(2.654)
$Probability_v \times Heterogeneity_v \times Simple$	-0.121	-0.110	1.941	3.511
	(0.642)	(0.571)	(3.572)	(3.333)
$Probability_v \times Heterogeneity_v \times Geo$	-2.588**	-2.562**	-0.391	0.538
	(1.131)	(1.039)	(4.028)	(3.618)
Village-level Controls	No	Yes	No	Yes
Observations	324	324	324	324
R-squared	0.133	0.224	0.113	0.222

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors are in parentheses. All regressions include seed centrality, seed probability, village-level heterogeneity, a constant term, and year fixed effects. Village-level controls include percentage of village using pit planting at baseline, percentage of village using compost at baseline, percentage of village using fertilizer at baseline, village size, the square of village size, and district fixed effects.