Social networks analysis in agricultural economies

Zachary Barnett-Howell[a,*] and Ahmed Mushfiq Mobarak[b]
[a]*Yale University and Y-RISE, New Haven, CT, United States*
[b]*Yale University, Deakin University, Y-RISE, NBER, CEPR and IGC, New Haven, CT, United States*
**Corresponding author: e-mail address: zachary.barnett-howell@yale.edu*

Chapter outline

Handbook of Agricultural Economics, ISSN 1574-0072, https://doi.org/10.1016/bs.hesagr.2021.10.010
Copyright © 2021 Elsevier B.V. All rights reserved.

1 Introduction

Economic activity in agricultural economies requires the involvement and cooperation of multiple people. An individual's economic behavior should not be understood in isolation: a single farmer or laborer will allocate their time taking into account the needs and preferences of their household. Households may comprise a large number of people, extending beyond a nuclear definition, producing a large set of interconnected relationships between the constituent members and hierarchies within the household that are defined along gender, age, and marital lines. The household's choices, too, are affected by relationships with others in the community: between members within separate households, and the connection of each separate household to each other. The interconnectedness of households is a necessity for agricultural production that takes labor as a primary input, with planting and harvest times that can require more labor than any one farmer or household can supply by themselves. Borrowing and lending relationships link households within a community in a web of complex financial obligations; migration flows connect people from rural to urban areas as they seek employment on a temporary or permanent basis in nearby cities and across countries. It is through these social ties that the economic decisions made by one person can affect others, with each choice spilling over into the broader community.

Agricultural economics studies how social connections serve as conduits for economic activity. Goods, services, and money flow along these social networks. Economic interaction and interdependence complements and structures the social fabric of agricultural communities. The flow of economic goods and services is often studied alongside the flow of information by researchers in behavioral sciences, including economics, political science, and sociology. Information is a crucial component to the study of economic activity within social networks for a number of reasons. Information provides the mechanism by which economic decisions are realized: choices over agricultural production, such as whether to adopt a new technology, or which destination to migrate to, are shaped by an individual's and household's

ARTICLE IN PRESS

information and beliefs. Where people acquire this information is, therefore, a first-order research question. In areas where people have little formal education and the provision of information by the government or other state institutions is scarce, community ties, trust, and information exchange form the nexus of social learning. The cost of exchanging information is low compared to other economic transactions. People exchange significant information about their experiences and beliefs over social events in a way that encounters fewer frictions when compared to the costs of transmitting cash or goods. Moreover, an individual's or household's beliefs are also relatively unconstrained, in the sense that new and convincing information can lead to a person diametrically shifting their beliefs on a topic. In comparison, material change is costly. Farmers may be constrained in their ability to spend more on fertilizer or adopt a new seed varietal, but are less constrained in believing that there is economic value to either decision. Even people who cannot perform the underlying action—e.g., cannot afford the upfront costs associated with migrating to a nearby city—can transmit the belief to others that such a move would be profitable. Accounting for the network effects of goods and information transmission can therefore lend insights to understanding individual and community level outcomes within agricultural societies. If we can understand how information flows within a community and how consensuses are reached, we may be able to offer policy-relevant advice to organizations conducting interventions in these areas.

1.1 When do you need network analysis?

Network analysis is most useful as a tool to study the behavior of people who depend on the cooperation of others, who lack full information about their environment, and when what each of them knows and does can affect the others. This is a departure from more traditional microeconomic models, which may assume atomized economic actors, where each of their decisions will have no substantial effect on other actors or the market at large. Economic models often assume that agents possess full information about their environment, or that if agents are uncertain, that there are no significant spillovers of information or choices onto other agents. If you are considering an environment which is known to its agents, so that there is no uncertainty about the state of the world, then equilibrium or game theoretic analysis may be more appropriate. If you are studying a case where agents are uncertain about their environment, but cannot effectively transmit information to each other, or the impact of their actions and choices are largely confined to themselves, then the use of reinforcement learning methods and heuristics (e.g., *multiarmed bandit* algorithms) is appropriate.

Social network analysis is the most appropriate in instances where there is considerable uncertainty over the state of the world, where information can flow between individuals, and where economic and social behavior is inherently interdependent such that connections between individuals and households regularly form durable networks. This includes a broad set of cases where people rely not just on their own experiences, but also the choices, information, and outcomes gained via the experiences of others to form more complete, accurate beliefs about the world. Information can be gleaned from others directly via conversation or text, or indirectly

through observation. Information and goods flow between people in a community, causally linking the beliefs and actions of a person to others to whom they are either directly or indirectly connected. Social network analysis is likely to be most valuable in cases where:

1. Information and beliefs vary widely between agents, and are communicable between them.
2. There are few formal markets mediating the production and delivery of relevant information, so that individual connections are crucial to determining choices.
3. And the number of actors is not so large that we cannot define and extract detailed information on each one.

1.2 Networks as substitutes

The study of the effects of social networks in high-income countries often focuses on some of the more pathological outcomes. These include studies of the capacity for social networks to rapidly diffuse misinformation about politics and public health initiatives using social media platforms, for example, Facebook, Twitter, and WhatsApp (Allcott, Braghieri, Eichmeyer, & Gentzkow, 2020; Levy, 2021; Peterson & Kagalwala, 2021). Other areas in higher-income societies in which social networks are frequently studied include how individuals increase their access to employment and educational opportunities through interpersonal connections (Bayer, Ross, & Topa, 2008; Beaman, 2012). While concerns over the degree to which social networks can lead to inequitable outcomes should not be dismissed out of hand, in the case of agrarian societies in low-income countries, social networks may also constitute an essential substitute where formal markets and institutions are incomplete or entirely lacking. If formal channels for communicating relevant and verified information are not robust or otherwise do not exist, then social networks can be drawn upon to provide and aggregate information and to reach a consensus. If there is uncertainty over how a new, unproven technology functions relative to traditional methods, collective experience can lead to a more correct group consensus, a "wisdom of crowds," where any one experience might fail to accurately determine the truth (Munshi, 2014). Social learning is an important phenomenon to study in any context, but within agrarian societies, social learning and networks provide critical functions that are often not provided for by existing markets and institutions.

1.3 Cautious optimism

The ubiquity of social networks in agrarian economies, combined with the complexity inherent to network analysis, has created a promising field of study. The slow pace of technological diffusion, the multiplicity of credit obligations within a community, and the migration decision over where and when to move may be better understood when the networks underpinning these choices are illuminated. However,

it is also increasingly clear that the theoretical properties of models of network diffusion and aggregation often fail to attain in real-world settings. There are important gaps between theoretical models of network formation and function, which have well-studied and understood properties, and the types of networks that emerge and develop within societies. As such, the ultimate value of models of social learning and network spillovers to policy-makers and practitioners remains to be seen. The data requirements to estimate detailed network structure continue to be a particular limitation to studying the properties and impact of the flow of information and goods across social networks. The immense expense that gathering network census data entails and the cultural specificity of network structures may limit the widespread adoption of network analysis. The use of network structure to inform policies, such as optimally seeding information, may be obviated by brute-force strategies, such as simply providing more seeds. Nevertheless, we remain optimistic that understanding the network structure of agricultural economies and how information is passed between people can be crucial to observational and experimental work.

This chapter is intended to serve a number of uses. First, it aims to provide concise and intuitive definitions of a number of technical network properties. For a more complete and thorough treatment, Jackson (2010) is an ideal place to start, and these concepts are similarly described in other review works, including Munshi (2014), Chuang and Schechter (2015), and Breza, Chandrasekhar, Golub, and Parvathaneni (2019). Our goal is to provide an up-to-date resource for graduate students or others who are similarly equipped and new to this discipline, to allow them to gain a sense of the current methods and tools in currency, as well as specific work which they may want to consider when designing their own research projects. The selection of applied papers that we describe is motivated both by our consideration of the subfields in which some the most valuable contributions of network theory have been made *vis-à-vis* agricultural economics, and areas of study that we believe will continue to be promising in the future. Research is therefore highlighted as a function of its historical import, our present belief in the potential promise that it offers the field, and a combination of these two considerations. Given the enormous breadth of the study of networks in agricultural economies, which extends across multiple disciplines, this chapter is necessarily incomplete. Our choice not to address numerous fruitful topics of study, including everything from job search networks to vote buying networks, is a reflection of our preference to present a more focused look at fewer topics than any critical opinion about the research that we have omitted.

2 Network theory, transmission, and learning models

The study of networks can be split between two, not entirely exclusive, categories. There are diffusion and contagion problems, which focus on the spread of simple information or states (typically binary ones). There are aggregation problems, which involve the transmission of higher-dimensional information and continuous-valued

states. This is an inexact typology, and a topic of interest may be placed into either category with some discretion by the researcher. Consider a study of how farmers choose whether to adopt a new technology. The researcher may be interested in studying the diffusion of this technology through a community or a region, where the outcome of interest is whether a given farmer chooses to use the new technology or not. Because adoption is a binary state, this appears to be a diffusion problem. However, the same study can also be posed as an aggregation problem, where the researcher might investigate how farmers learn about multiple dimensions of the new technology, and how they construct posterior beliefs over its value using the experiences and decisions of those around them. The choice of posing a research question as either a diffusion or aggregation problem can be equally correct; it does imply a different modeling approach and set of data necessary in order to adequately investigate and answer the question. This chapter outlines the models, data considerations, and methods that go into the making of network research design decisions.

2.1 Terminology

The study of networks has its roots in a number of disparate fields, including sociology, epidemiology, and graph theory, and so a nonstandardized terminology has evolved to describe network phenomena. For our purposes, denote the relevant economic actor or agent, whether an individual or a household, as a *node*. Nodes are distinct from one another, forming the building blocks of a given community. A community of N nodes can be discretely indexed $i \in \{1, 2, \ldots, N\}$. Nodes can form connections with each other called *edges*. The set of immediate connections between a node and other nodes is called a *neighborhood*, and the set of all nodes and edges is the relevant social *graph*. A *walk* defines a sequence of nodes connected by edges, allowing for cycles where nodes and edges can be revisited; a *path* is a walk without repeated nodes. The *diameter* of the network is the shortest distance between the two most distant nodes.

Each node in the network can be defined by a set of attributes, as well as its connection to other nodes and its location within the graph. The *degree* of each node can be calculated by the number of edges that it has with other nodes in the graph. One of the characteristics that nodes may possess can be one or more *state* attributes, representing a set of beliefs or observable choices that can be affected by the beliefs and choices of others in their network. Information and goods in a network are transmitted between nodes along edges. Agents can only access information or goods from other nodes that they have connections with, so the distance a given flow must travel is determined by the number of connections it must traverse.

A network can be represented either graphically or as a matrix. A graphical depiction of a network lays out the nodes as geometric shapes and connects them with lines, allowing a researcher to develop an intuition as to the overall structure of the network. However, even in a relatively small network, nodes can be arranged in a number of ways, and these arrangements may lead to different interpretations. As the number of nodes and edges increases, graphical depictions can be increasingly

confusing and hard to interpret.[1] The same graphical depiction of a network can be described as an $N \times N$ adjacency matrix, where each row and column represents a single node in the network. An edge connecting nodes is coded in adjacency matrix notation as a 1, and the absence of an edge is coded as a 0. The representation of a network as an adjacency matrix is a necessary step to allowing a researcher to perform the matrix algebra used to estimate node and network characteristics.

2.2 Diffusion and contagion

A model of diffusion or contagion is the study of how novel information or a state can spread throughout a network. Many of these models are rooted in the study of communicable diseases (hence "contagion"), but are widely applicable to processes in which new practices or individual states gradually become the norm, or else fail to take root. The means by which this new state enters a social network is through one or more *seeds*, which are the first nodes that display the state in question. In the study of epidemiology this state would be the infection of one or more patient zero(s); for agricultural economics this state could instead refer to the adoption of a new technology, or the use of a new farming practice. Crucially, the state can be transmitted across edges, so that nodes connected to the initial seed are susceptible to changing their state. It is through this process of diffusion that a novel state beginning with an initial can *saturate* a network, an outcome where all other connected nodes exhibit the relevant state. Diffusion processes can fail, which includes cases where the initial state fails to spread beyond the initial seeds, or spreads but does not saturate the network.

The class of diffusion models can be characterized by the workhorse "Susceptible, Infected, Recovered" (SIR) compartmental model. In this model, nodes begin in a susceptible state and become "infected" with some probability through contact with the initial seed. The state propagates through the network, and in the epidemiological case, nodes eventually enter a recovery state. While recovery is obviously a relevant state for the study of disease dynamics, it is potentially less relevant to the study of socioeconomic processes in which "recovery" from learning a new idea or having access to a new technology is not always a relevant state. In most economic diffusion models, the SIR model can then be succinctly reduced into two constituent states: $\{0,1\}$, susceptible and infected.

Consider a farmer who has not yet adopted a new technology, and therefore can be considered susceptible to adopting the technology through contact with others who have adopted it. If, through exposure, our farmer in question too adopts, then they become "infected," and are likewise capable of teaching others in their

[1]It is a nontrivial problem to algorithmically visually draw a network, producing a graphical depiction of the nodes and edges according to both formal and esthetic criteria. See Herman, Melancon, and Marshall (2000) and Beck, Burch, Diehl, and Weiskopf (2017) for a survey of research on algorithms in this area.

neighborhood about the benefits of the technology. This diffusion may occur through direct conversation with others or through observing the adoption choice and resulting outcomes.

A larger set of states can be used to represent more complex phenomenon. For example, Banerjee, Breza, Chandrasekhar, and Mobius (2019) employ a three state construction, $\{\emptyset,0,1\}$, where the empty set denotes individuals with no priors or exposure, a 0 represents a person who is aware of the new technology but has not yet chosen to adopt, and a 1 denotes adoption. However, as the number of states grows, a diffusion model becomes less appropriate and a learning and aggregation model may be more useful to capture the complexities of learning about a state with higher dimensions. It is important to emphasize the causality of the network structure. A frequent assumption is that there should be no spontaneous adoption—absent the initial seeds, no one in the network should exhibit the relevant state except through contact with others. In this way, only nodes in a neighborhood of other adopters can change their state. Someone unaware of the new technology cannot learn about its properties and select into it.

Transmission between nodes is a probabilistic event, where the likelihood of transmission in a given period t lies within the interval, $p \in (0,1]$. Clearly, for $p = 0$ the information or state cannot be transmitted between nodes, and so there are no network effects to study. As the transmission probability approaches 1, the process diffuses more rapidly throughout the graph. Transmission is a mechanical process; in the most basic construction of this model, agents do not engage in any strategies that either hinder or accelerate the diffusion process.

A difficult modeling question is whether the probability of transmission changes over time. In the epidemiological case, people are typically only capable of transmitting the relevant state for a fixed period of time. However, there is no a priori reason why social learning over the benefits of a new technology would be bounded to a fixed period. A consequence of an unbounded infectious period, however, is that even with small transmission probabilities, we would expect to see complete diffusion through a social network over any meaningful length of time. In many parts of the world, we have seen this type of diffusion process occur with any number of everyday technologies (e.g., telephones), but in many agrarian societies in low-income countries, the incomplete diffusion of many potentially valuable technologies may be the norm.

2.3 Virality

The number of people each node can transmit their state to is determined by the probability of transmission, the number of other people in their neighborhood, and the period of infectivity. The expected number of people that each person will infect is denoted as R_0, also known as the basic reproductive number. Networks with a high degree of connectivity between individuals will transmit information faster and therefore saturate the network quickly, while networks with more limited connections between members will take longer for the information or behavior to reach

everyone. For a specific diffusion process the R_0 value that we estimate for any given network will be a function of the characteristics of that network. Importantly, this means that the R_0 is not inherent to the information or choice that is diffusing through the network, but is a jointly determined with the specific network in question. Changes in behavior can likewise change the basic reproductive number, where the object of interest would be defined as infectivity in a given period, R_t instead of R_0. A technology that increases social connections between people may be transmitted with increasing speed over time. A sufficiently attractive idea or behavior may even induce people to forge connections with others as a means of gaining access, thereby increasing its diffusion rate.

When the R_0 of a process is less than 1 the diffusion process is self-limiting, in that each person transmits their state to less than one person on average, and therefore the process is unlikely to diffuse to an entire network without additional seeding. When the value of R_0 is equal to or greater than 1, each person infects at least one other person on average, and the process is expected to propagate through the network. As the value of R_0 increases above 1, the diffusion process is expected to accelerate over time, as each marginal infection accelerates the rate at which the process saturates the community.

2.3.1 Complex contagion

A logical endpoint of a diffusion model is the case of a successful diffusion process, instances where new information or a new state saturate a network. Unlike in the epidemiological context where transmission can be bounded due to biological factors, there is often no convincing reason to establish bounds on the time period over which economic ideas and behaviors can be transmitted. If an agricultural technology is truly an improvement over preexisting methods, we should expect adoption to become the norm. Yet in many agrarian societies we find diffusion processes to be incomplete: diffusion of new technologies can start with a few seeds, permeate further into the network, but ultimately fail to saturate the network. There can be disadoption too, where nodes that have adopted a new technology later abandon it. How to best explain incomplete diffusion has motivated the development of new models of diffusion.

A modification to the diffusion model to explain instances of slow or incomplete diffusion is the introduction of thresholds, which are used in models of *complex* contagion. In this model of diffusion a nodes cannot independently transmit the relevant state to another node, but can only do when $k > 1$ connected nodes have this state. Intuitively, consider a case where a farmer only adopts the new technology if they are connected to two other farmers ($k = 2$) that have also adopted this technology, this is the minimum number of adoption events that they find convincing—observing a single adoption event is insufficient evidence of the value of the new technology. If the threshold value of k is equal to 1, then this reduces to a model of simple contagion. As the threshold value k becomes large, the speed with which the process diffuses slows down and can even halt if there are bottlenecks in a network

where an insufficient number of connections exist between groups for an idea to spread (Acemoglu, Ozdaglar, & Yildiz, 2011; Centola & Macy, 2007; Granovetter, 1978).

2.3.2 Herd models

Diffusion processes allow new information, as well as observable choices and other states to saturate a network. The diffusion process may exhibit pathological qualities, whereby all members of a network conform to a common choice, regardless of whether it is an optimal choice for them. The metaphor suggests a herd starting to move in a specific direction, leading all other members it encounters to follow in the same direction regardless of their own beliefs and preferences. Consider the choice over whether to adopt a new seed varietal: farmers begin with heterogenous beliefs over whether it will be profitable. The adoption choice by one farmer informs others that they at least believe it to be profitable; the choice not to adopt sends the opposite signal. A sequence of decisions that is sufficiently informative so that subsequent farmers ignore their own prior beliefs over the value of adoption constitutes a herd, a point at which later choices no longer convey new information to others in the network. This can allow large groups of people to select into "irrational" or suboptimal outcomes (Banerjee, 1992; Bikhchandani, Hirshleifer, & Welch, 1992).

Acemoglu, Dahleh, Lobel, and Ozdaglar (2011) provide a more general treatment of herding models, characterizing conditions in which networks will and will not produce herding and information cascades onto suboptimal choices. In increasingly large networks, the influence that preceding choices has on later ones shrinks, eliminating herding outcomes. But in social networks where a small number of nodes exert considerable influence, then their choices can lead to inefficient outcomes. Although this class of model is attractive as a means of explaining either rapid changes in observed choices within a network, or the collective arrival at an inefficient-seeming outcome, herd models rely on a specific ordering of choices. Agents within a network must follow a strict order with which choices are made and then observed, and choices must be permanent once made. This is potentially applicable to instances where choices are infrequent and time-limited as to when agents can make them. However, if agents are repeatedly making the same choice or can vary the timing of when they make it, then a herd model may be less appropriate.

2.4 Aggregation

Aggregation problems are those which involve more complex beliefs and information transmission. Unlike in contagion and diffusion models, where agents' states or beliefs are represented using a small number of discrete states, aggregation and learning problems typically deal with a continuous state space. A successful diffusion process ends with the network being saturated, failure indicates a process that does not spread beyond its initial seeds or subgraph. An aggregation process is similar in that success is defined as the entire graph arriving at a consensus, sharing the same information set. Failure in an aggregation model is when the network does not

end up sharing the same information set. This type of model is useful to consider how more complex consensuses and beliefs form, which in turn can be mapped onto discrete actions and outcomes. Nuanced beliefs require more complex representation, and the relevant question may not be whether an individual has heard of a new technology, or even whether they have adopted it, but rather what they believe its actual contribution to their seasonal profit to be. The payoff to this approach is that it allows the consideration of a distribution of beliefs and perspectives by agents at the cost of increased complexity.

2.4.1 DeGroot learning

DeGroot learning is an analytically tractable model of how an individual sequentially updates their beliefs using information derived from people that they are connected to. This approach allows us to model how a network might arrive at a consensus, as information is aggregated between agents. Modeling DeGroot learning involves a sequence of time periods, nodes, edges, beliefs, and weights. Each node begins with an initial prior in the first period, t_0. In each subsequent period, $\{t_1, t_2, \ldots\}$ each node updates their beliefs as a weighted average of the beliefs of nodes that they are directly connected to. The weights correspond to how much a node trusts or values the information that they receive from each other node that they are connected to, where a weight of 0 denotes that they completely disregard information from that particular node. DeGroot learning is a heuristic, as nodes do not account for the interdependence of beliefs between each of the people they are connected to, which may be similarly interconnected (DeGroot, 1974; DeMarzo, Vayanos, & Zwiebel, 2003).

An analysis of DeGroot learning focuses on the network derived from the weights that each node applies to the information that they receive from others, as this defines the relevant social graph for how the network aggregates information. By iterating over the initial vector of priors and weight matrix, we can compute whether the network arrives at a consensus. The influence of each nodes initial beliefs on the final consensus is a function of their centrality within the network and the weights that others afford them. Beyond the question of whether a consensus can be achieved, it is also important whether or not that consensus is correct. The concept of the "wisdom of the crowd" is the idea that individual node level beliefs may be arbitrarily biased, but society as a whole may hold accurate beliefs on average. Consider the case where all individuals receive a noisy signals on the true state of the world and aggregate information over time using a DeGroot learning process. As the influence of a small set of nodes increases within this example, such that their information is increasingly weighted by others, the network can converge on a consensus. That said, it may not be accurate in the sense of revealing the true state of the world, instead representing the potentially biased priors of the most influential nodes within the graph. However, if the total weight conferred to the beliefs of a fixed set of nodes becomes arbitrarily small as the size of a network grows, and if groups are not too insular, then the consensus that the network arrives at will be accurate. The conclusion is that insular groups, in which the opinions of a few nodes are overweighted, can maintain beliefs that are disjoint with the rest of the network. These insular groups can also potentially

exert a significant effect on the broader social consensus that the entire network arrives at (Golub & Jackson, 2010).

2.4.2 Bayesian learning

Bayesian learning exchanges the weighted average update rule of a DeGroot model for a proper prior-to-posterior update rule. Nodes must account for interdependence in the information they receive from others in the network. Because all nodes are simultaneously updating based on information they receive from everyone they are connected with, who are also interconnected, not all of the information they receive is actually new. A node may be repeatedly counting the information from one person transmitted to them through multiple sources. The Bayesian learner will consider the correlation in signals and adjust their updated belief accordingly.

In order to adjust for the correlation in information they receive, a Bayesian learner faces of a problem of significant complexity: if each node knows the entire structure of the network—who the nodes they are connected to *are connected to*, and then who the nodes that they are connected to are connected to are connected to—then solving the sequential updating problem is possible, if likely computationally infeasible. It may not be a good description of reality, however, allowing for nodes in a network to communicate both their information as well as the complete provenance of that information. For a problem with more than a small state space, the extensively detailed conversation explaining where each new observation or idea originated from is difficult to imagine occurring naturally. If on the other hand, nodes are not provided with complete information both over network structure and the interdependence of signals they receive, which is increasingly likely as we iteratively move away from a node's initial neighborhood, then each node requires prior beliefs over the likelihood of each node they are connected to being connected to other nodes, and so on. For anything other than a very small-world, a model of Bayesian learning is likely intractable.

2.4.3 DeGroot vs Bayesian

The simplicity of DeGroot learning makes it computationally tractable; it is also a relatively accurate description of how people learn. A number of studies have used experiments to examine the applicability of either a Bayesian or DeGroot learning model at describing the patterns with which people update their beliefs and make decisions.

There are a number of ways to distinguish DeGroot from Bayesian learners: DeGroot learners can become "stuck" in inward-facing clans, unable to learn from the rest of their network; DeGroot learners are oblivious to others' position within the network, which should not affect how they weight the information from those nodes. Experimental work focusing on small networks, where participants receive private information and make observable decisions, found Bayesian updating to be a more accurate description of individual behavior (Choi, Gale, & Kariv, 2005, 2012). Grimm and Mengel (2018) find that varying the amount of information provided to participants about the overall network structure does have an impact on how they

weight information from more central nodes, a Bayesian approach. However, the majority of individual outcomes in their experiment were consistent with both a Bayesian or DeGroot learning model. Testing similar learning rules, Chandrasekhar, Larreguy, and Xandri (2020) suggest that a fraction of nodes within a network may update their beliefs as DeGroot or Bayesian learners, and find that a nonconstant fraction of experimental participants in villages in India and a university in Mexico exhibit either DeGroot or Bayesian learning behavior.

2.4.4 Behavioral learning models

Where DeGroot and Bayesian learning vary by the update rule and information set that each node possesses, there are other models of attention and learning that may be useful within the context of network analysis. In a behavioral model of learning, an individual has only a limited capacity to attend to all of the multiple dimensions of a new idea or technology, as well as strong prior beliefs over which of these attributes are important and which are not (Gabaix, 2014; Schwartzstein, 2014). Hanna, Mullainathan, and Schwartzstein (2014) test a model of selective attention with data on Indonesian seaweed farmers from seven villages in Nusa Penida, Indonesia. There are numerous important dimensions to seaweed farming: some of which the farmers are aware are important, and have learned how to optimize, and another dimension which the farmers were unaware of its importance, and therefore have not optimized. Hannah et al. propose the existence of a feedback loop, where ignorance over the importance of an overlooked dimension can persist over time as the lack of experimentation and observation means that its true importance is never revealed. Such behavioral models of learning, in which attention is inherently limited, can be useful in helping explain the time it takes to fully learn complex ideas, beliefs, or technologies. Applying the concept of a sparse consideration set to models of social learning may involve the selective attention about which nodes in the network to learn from, which attributes to learn about, and possibly a combination of the two.

3 Network formation, structure, and directionality

Social networks emerge under a specific set of preconditions. Within a community the repeated interaction of people is necessary for social connections to form, and these interactions may occur randomly or strategically in a variety of contexts. Other types of interactions are far more strategic. Who we choose to work with, borrow from, and where we choose to migrate may depend on a person possessing specific attributes or capabilities. These connections may only emerge through significant effort and design. Lending networks depend on demand just as borrowing networks depend on supply. Connections form around those with greater wealth, and an individual may exert considerable effort to maintaining a connection with someone they can borrow from in times of need. The sharing of information on farming practices likewise may emerge through geographic proximity, but also on other

observable choices, such as a preference for riskier or safer crops, or across distances between farmers who own land that has similar characteristics.

Random networks form randomly, or as-if random. The connections between people emerge from a context in which neither individual would have necessarily exerted significant effort or planning to encounter each other. A strategic network is the opposite, where an individual forges a relationship with another for a specific purpose. The creation and maintenance of strategic edges may be costly, to the extent that an individual would have to pick and choose which people they wish to frequently associate with.

Networks can further be characterized as directed or undirected. An undirected network is one in which relationships are reciprocal. In the case of a family or caste-based social network, edges are inherently reciprocal. A cousin is the cousin to their own cousin, just as a member of a caste or a social group within a community shares a reciprocal relationship with others in that group. Note that reflexivity is not about hierarchies within the network—a patriarch may rank higher than their son—but that familial connection is necessarily reflexive: there are no sons without corresponding fathers, and vice versa. This is different from the case of *directed* networks, where connections are not inherently reciprocal. In the case of a borrowing-lending network, a borrower is not necessarily a lender, loans may flow in only one direction. Again, for people who prefer to learn about new agricultural technologies from the most educated and proficient farmers in their community, the flow of information may be directed (De Paula, 2020; Jackson, 2003).

3.1 Random formation

A workhorse model of network formation is one in which edges between nodes are randomly assigned. Although not a strictly accurate representation of network formation in a number of contexts, this approach may still be useful. The properties of random networks have been extensively studied, and may sufficient to provide inference, or at least serve as a benchmark for more complex network formation models. In a graph consisting of a fixed number of nodes, allow for the probability that a link forms between any two nodes to be distributed Bernoulli with probability p. With a sufficiently large graph and a constant likelihood of edge formation, the distribution of the *degree* of nodes approximates a Poisson distribution, which is why this model can be called a "Poisson random graph."

A Poisson random graph exhibits a number of attributes. When the probability of edges forming is small, the graph is sparse, exhibiting tree-like structures with limited cross-connections between nodes. As the average degree of each node increases, Poisson random graphs undergo a phase transition: without loss of generality, above a critical threshold the likelihood of certain properties existing within a network move sharply from 0 to 1. For example, when the probability of link formation p exceeds $\frac{\log(N)}{N}$ the probability of the network becoming interconnected converges to 1. Random networks are easy to simulate: start with an empty upper adjacency

matrix, and create an edge between nodes $\{i, k\}$ if $\epsilon > p : \epsilon \sim U(0, 1)$. To target a network with a given mean degree value of c, the probability of link formation is set to $p = \frac{c}{n-1}$ (Jackson, 2010).

3.2 Strategic formation

Real-world social networks often result from deliberate interaction. This is not to suggest a purely instrumental view of social connectedness, but rather that long-term friendships and ties are built around mutual trust and enjoyment. People seek out the company of those who make their lives better, and may fail to maintain relationships with individuals who they find unpleasant or bring little to the relationship. This may be understood in terms of strictly economic payoffs, but it also adheres in the case of advice or gossip networks: we may derive more from connecting and interacting with some certain people who can tell us more than those who have less to share. As a result our interactions with others are mediated by *their* interactions with others. The amount of new information a person provides can be a function of the people they are connected to, just as a mutual support network relies on the entirety of its members to maintain its function.

Edges between nodes in a strategically formed must be *pairwise-stable*. Both members of the dyad must agree to establish and maintain the connection, no one has the ability to unilaterally forge a connection with someone who does not want to be connected with. It is difficult to get advice from someone who does not want to give it, or to borrow money from someone who does not want to lend it. Formally, a pairwise-stable network is one in which no node would be better off cutting ties with someone they are connected with, and no two unconnected nodes should be able to improve their outcomes by forming a link (Jackson & Wolinsky, 1996).

Since the value of any connection is further determined by each node's position within the network, who their friends are, link formation needs to consider each node's degree. The likelihood of a stable link forming can be given by a utility function, taking each nodes observed and unobserved characteristics and its location within the graph. Of course, since the position of each node is determined by this exact link-formation process, one approach has been to focus on subsets of the network which are computable. Heuristics are necessary to insure that the problem is tractable, such as agents choosing attachments based on the current and not evolving network structure as a way of avoiding instances of multiple equilibria (Jackson & Rogers, 2007; Jackson, Rogers, & Zenou, 2017).

3.3 Homophily

If one of the goals of studying network effects is to uncover causal relationships between the actions and information that one node possesses, and how this is in turn transmitted to others connected to them, *homophily* complicates inference and can have significant repercussions for the diffusion of information across a graph. Often termed the "reflection problem," homophily is a phenomenon by which people with a

similar background, whether age, ethnicity, caste, religion—or any number of less-observable characteristics—tend to more frequently associate and form ties (Manski, 1993). The similarity of people that form interconnected groups makes inference on network effects more difficult: if a group of similar people begins to modify its behavior, is this due to peer influence or is it the result of a common unobserved factor that affects people like them? The discussion of diffusion and aggregation models implied causality, as social ties allowed for the direct transmission of a state. But absent a well-defined mechanism of transmission, it is possible that no real information is being transmitted across the social graph, and instead members of a community are reacting to a change in their environment.

Consider a researcher interested in studying the migration decisions in a village. Households in this village have begun sending members to a nearby urban center as migrants, so there is a defined order to which migration occurs. Interconnected households may have a causal effect on each other, where the action of sending a migrant, the money they remit, or the sharing of information about their experiences with other households may induce that connected household to likewise send a migrant. However, it is also possible that no information is being transmitted, and that this increase in migration is the result of a change in urban working conditions, culture, or laws that affect members of this group (Bramoullé, Djebbari, & Fortin, 2009; Manski, 1993; McPherson, Smith-Lovin, & Cook, 2001).

Homophily has implications on the rate at which nodes can aggregate information. Extreme homophily suggests small, inward-looking subgroups within a graph, making it difficult to transmit information between the subgroups or for members to arrive at a broader consensus. Golub and Jackson (2012) examine the effect of homophily on the rate at which consensus emerges within and between groups. If nodes weight information from others like them more than nodes not like them, then within-group convergence on a consensus is more rapid, as the rate at which the entire graph arrives at a consensus becomes more slow. Essentially, the way in which information is weighted between nodes can be more important than the degree distribution of the network. This has natural implications for the rate of social learning possible in agrarian communities that exhibit high levels of within-group segregation, and therefore may not be able to effectively aggregate information on questions like how to access employment in urban areas, or whether the adoption of a new technology will be profitable. An increase in the effectiveness of communication technologies, like the introduction of cellphones, can even further reduce the rate at which an entire community arrives at a consensus if it leads to more inward-looking subgroups.

Conducting inference when homophily introduces correlations in outcomes between similar units requires either eliminating or recovering the presence of unobserved correlated effects. This can be accomplished through the use of local fixed effects or through a network formation model that accounts for the likelihood of individuals with similar unobservable characteristics forming connections (Bramoullé et al., 2009). Goldsmith-Pinkham and Imbens (2013) detail how to use models of both exogenous and endogenous network formation, focusing the likelihood of any two nodes sharing similar unobserved traits as proportional to their observed

traits and whether they are connected. For example, a dyad with dissimilar observed traits is likely more similar across unobservable traits, based on the assumption that homophily induces clustering among similar individuals. Similarly, the lack of an edge between two similar-appearing nodes suggests a stronger dissimilarity in unobservable traits. Using an estimate of these latent parameters in the analysis of network effects can attenuate the bias that the correlation in unobservable traits introduces (Aronow, Samii, et al., 2017; Jackson et al., 2017).

3.4 Directionality

Links between nodes are not always reciprocal, regardless of whether they form in a random or strategic context. Consider the case of a village moneylender who provides loans to households with less wealth, but does not receive loans from them. Similarly, a person might seek advice from a more experienced and educated farmer, while the more experienced farmer does not seek their advice in return. These people are still connected, but information and goods do not flow both ways—the edges connecting nodes possess a defined direction. This contrasts with the undirected network, a case in which edges are reciprocal between nodes, as friendships or group memberships are mutual by definition. While undirected networks may exhibit a directedness at some points in time, individuals sharing information (a gossip network) may alternate who provides and who received information. But there is an expected eventual reciprocity in the way that many mutual aid networks involve the repeated borrowing and lending of small amount of food, oil, or money. Separate from the question of petty borrowing would be who a person would rely on to acquire larger amounts of credit; likely the local moneylender, who may or may not engage in the petty lending with the same person that characterizes.

3.5 Centrality measures

One premise involved in the study of networks is the assumption that certain nodes play a more important role in the economic process being studied than others. It is intuitively tractable that certain individuals or households may be more trusted than others in their community, potentially seen as bellweathers, or else simply be more connected, and therefore more able to transmit information across the relevant social graph. Centrality measures are a way of formalizing this notion of importance or connectedness within a social graph. An individual node may be key in connecting disparate groups, or otherwise be the crucial point in a graph that allows information to be aggregated more quickly. If the diffusion or aggregation of information is markedly different across communities, heterogeneity both at the network structure, but also at the node level may be key.

Formalizing the notion of importance within a network can be done in a number of ways, each with a slightly different emphasis on what constitutes importance. *Degree centrality* is the simplest measure of importance, counting the number of edges each node posses. While this measure speaks to the relative connectedness

of each node, it is not informative about that node's position within the network. Rather than count the raw number of connections a node possess, we might consider *eigenvector centrality*, a recursive measure of importance where the importance of an individual is measured through the importance of the nodes that they are connected to. A related measure to degree and eigenvector centrality is *diffusion centrality*, which is based on the importance of each node in communicating information over time (Banerjee, Chandrasekhar, Duflo, & Jackson, 2013). *Betweenness centrality* is a measure of the importance of each node in connecting disparate nodes within a network. This may be a key consideration for communities in which there exist relatively close-knit groups with many ties, but few individuals or households with cross-cutting connections. This list of measures of centrality is not exhaustive, with a more complete survey provided in Jackson (2010) and Bloch, Jackson, and Tebaldi (2021). It is important to note that these measures of centrality are all correlated and nonexclusive. They all can be computed using the same data, and certain measures such as diffusion centrality represent a weighted average of other centrality measures.

4 Data collection and measurement

High quality survey and experimental data on agricultural economies have become increasing standardized across a variety of outlets. The survey methods and the types of questions found in publicly available microdata, such as those published by the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) or the World Bank's Living Standards Measurement Surveys (LSMS), can provide researchers with a baseline for illustrating and testing a wide set of theories on the form and function of agricultural economies. Unfortunately, the vast majority of these existing agricultural surveys are unsuited to the study of social networks. This unsuitability is not only due to a lack of social network questions on the ties between individuals and households within a community, which are clearly a necessary precondition to study social networks. It is also the sampling methodology used to create locally and nationally representative surveys that makes these data products unsuitable for the study of networks.

Fortifying a typical agricultural household survey into one that can provide relevant social network data is not straightforward. There are a multiplicity of social networks within any community, not all of which are immediately relevant to a particular topic of study. Geographical proximity, the people that are literally within a neighborhood of each other, comprises one social network and correlates to many other types of social networks and interconnectedness. However, in agrarian societies, geographical entities can be noncompact. Consider a farmer who owns multiple, nonadjacent plots of land, which are also disconnected from their physical homestead. Are only neighboring homes relevant to their network, or all the other farmers who own land adjacent to the farmer's land? This question becomes more complex when you consider nonnuclear households where multiple household members can

ARTICLE IN PRESS

own and farm different plots of land within an area. Homophily can also lead to geographic clustering, producing a relatively homogenous group of people within a given area. Kinship ties are also more diffuse in nonnuclear households, where multiple generations living under one roof may connect to a large set of other households in the community; not all of these connections will be relevant.

Yet the question of which social network is most relevant to the topic being studied complicates the addition of social network questions: family ties may prove important for understanding financial flows and risk-sharing agreements, farmers may prefer to learn from those they consider the most technologically proficient, more complex relationships may mediate others types of information flows. Gathering even baseline information on social networks then requires defining both the outcome of interest as well as the multiplicity of potential network connections along which goods and information flow to affect it. Social network questions may constitute real or hypothetical linkages, such as who a person engages with on various activities, and who they *might* go to if they were in need of advice or aid. Such questions can be used to construct directed or undirected graphs.

The most significant problem with using agricultural surveys to estimate network characteristics is their sampling frame. Most surveys are designed to provide a representative sample of an area, with individuals and households randomly sampled according to some stratification design. Only a fraction of a population is sampled and then contacted for the survey, typically never more than half, and usually a much smaller percentage. With proper randomized sampling methodology however, inference is valid on this sample, and through the use of survey weights a researcher can employ this survey data to generate relevant statistics about the broader population. This sampling approach does not allow for the unbiased estimation of network statistics. Partial sampling introduces nonclassical measurement error that creates biases of an unknown sign and magnitude into regressions that rely on the estimation of fundamental network concepts, such as centrality (Chandrasekhar & Lewis, 2016).

4.1 Unit of observation and measurement

The relevant social network is determined by what is being studied. The relevant social network for farming practices may be entirely distinct from the one underpinning migration patterns. Person-to-person linkages may serve to connect households; kinship networks may be relevant in some societies, less powerful in others; proximity and location may also determine diffusion of certain processes. Given the multiplicity of networks in any agrarian society, it is up to the researcher to determine which are the relevant ones, and this is typically why information on multiple networks is collected.

Beyond establishing which networks are most relevant to the topic at hand, the researcher also must consider the interaction between individuals, households, and higher level units of social aggregation. Social connections are inherently interpersonal, but the relevant economic unit within an agrarian society may be a household, a clan, or a caste. Does a relationship between two people in different households link

the two households? For households with significant hierarchies and sex segregation, a connection between two individuals may not transmit information to the head of each household who is responsible for making the relevant economic choices.

Network data is typically gathered through extensive household surveys. These can involve the usual set of questions relating to the household's constituents' socioeconomic attributes, as well as questions regarding their historical, present, and hypothetical interactions. Questions like who a person has recently lent to or borrowed from can be informative of the financial network within a community. Yet recent interactions may not capture the set of historical financial relationships, as well as speculative ones—who *would* the respondent borrow from if the need presented? And would that change depending on whether the amount in question was small or large? A straightforward seeming question about financial transactions can spawn an increasing number of complementary questions. Answering hypotheticals is also tricky, inviting considerable speculation by the respondent as to what circumstance would require that type of interaction.

4.2 Censuses

A census is a survey in which each and every member of the relevant population is contacted and counted. Census surveys are essential to the study of social networks, as random sampling methods are unsuitable for estimating network statistics such as centrality. The amount of data required separates network analysis from most other research in microeconomics, which typically rely heavily on sampled surveys in which enumerators only contact a small fraction of the relevant population. While proper survey design and sampling methodology is a field in of itself, with the proper randomization of survey takers and the use of survey weights, the data that sampled survey produces can be used to conduct valid inference both over the sample and the population of interest. Sample surveys are cheap in that a relatively small sample can be informative about a much larger population, allowing for longer or more varied surveys to be conducted across a variety of contexts. Many economic surveys are also relatively impersonal, inasmuch that they do not necessarily ask about specific friendships, contacts, and social ties, depending on the topic being studied.

Unfortunately, sample surveys are also insufficient for conducting research on networks. Chandrasekhar and Lewis (2016) consider the problem where an economic outcome is a function of either a local or global network statistic, the former pertaining to a node's neighborhood, the latter to the entire graph. They consider the case where a subset of nodes are surveyed out of a population. The authors demonstrate that random sampling of network nodes does not prevent nonclassical measurement error, and leads to attenuation bias, expansion bias, and even sign switching in the estimated effect of a given network statistic on the outcome in question.

A network census requires a full listing of each and every member of every household, then an elicitation of every relevant social tie to the subject being studied. There is likely more than one potentially relevant social network, and so the number of elicited social ties grows for every social network dimension that the researcher wishes to collect data on. In some cases this work is done in two steps: first,

collecting a complete set of names and sometimes images of everyone in the study, then using those to prompt and further narrow down participant responses so that the researcher can be certain that they can identify each individual a survey participant lists as a member of their network. Identifying the right person can be particularly challenging in close-knit agricultural communities where intermarriage is common and last names are not a useful source of unique identification.

Network questions typically ask a person to enumerate the number of people they engage with in a particular behavior. These questions can be about directed or undirected connections. Framing these questions to recover the existence of a linkage can be difficult. A researcher may wish to know from whom a farmer learns about agricultural technologies in general. Writing a survey question asking the farmer to list all of their edges is too abstract to be useful. Instead, questions are commonly written in regard to specific, everyday actions: "who are the five people you are most likely to go to borrow a small amount of rice or cooking oil?" "Who do you spend the most time with socializing or drinking tea?" "Which five people do you discuss farming practices with most commonly?" A survey protocol may include eliciting names until the question (and potentially the subject) is exhausted. The survey protocol may include eliciting names up to a fixed number, usually between 5 and 10, or asking the farmer to list a fixed number of names. Each of these approaches has its drawbacks. Asking for a potentially unlimited number of names allows the collection of both strong and weak ties, but can introduce node level heterogeneity in terms of each farmer's interest in participating in the survey, their own threshold for which names to include, and other sources of classical and nonclassical measurement error. Eliciting a fixed number of links eliminates this heterogeneity, although some respondents may also be unable to provide the full list of requested names. This, however, creates a truncation bias whereby weaker links are ignored and peer effects are under-estimated due to missing edges (Griffith, 2021). Framing the question appropriately is important: does the researcher want to elicit a list of nodes of which the respondent is merely aware? Or the nodes that the respondent can directly observe versus those with whom they discuss the relevant topic? The researcher may be interested in drilling down into eliciting connections with whom the respondent *regularly* discusses the relevant topic. Each of these different framings can introduce variation into the answers that the respondent will provide (Maertens & Barrett, 2013). It is also worth noting that it can be difficult to effectively train enumerators how to best ask these questions. Even with proper translation, in many agrarian societies these questions may not make immediate sense to respondents. As a result the survey can take a significant amount of time to complete, driving up research costs. Therefore, extensively piloting a network is highly recommended to ensure that it can generate high quality data.

4.3 Aggregated relational data

Census surveys are expensive to conduct; censuses containing appropriate data on network connections are even more so. An alternative approach to a full network census survey is to use *aggregated relational data* (ARD), a hybrid of a sampled survey and full network census that admits valid inference over network structure at a

reduced cost relative to a complete network census. ARD is an attempt to reconstruct the latent social space in which a broader social network forms. By taking relationships between nodes that share similar traits or activities, the results can be used to reconstruct a number of network statistics.

An ARD question elicits from a respondent how many people or households they engage with in a given activity share a particular trait. Which trait and which activity to ask about are a matter of context. A good ARD question is one that is informative over the latent social space, in that it is common among people who are socially clustered, and uncommon everywhere else. Traits or behaviors that are evenly distributed throughout a population, or that are uninformative of social or economic standing, do not make good ARD questions. Obviously, there is no one set of traits that will fit every research context, but questions relating practices informative of socioeconomic status, relating to religious or culturally specific practices, or risk-seeking or potentially marginal viewpoints or behaviors, may prove useful at illuminating the latent social space.[2]

Generating ARD data relies on both a brief census, recording some demographic data as well as the responses to each ARD question, and a sampled survey, getting the full response to each ARD question over how many people within the respondent's network share the ARD trait. Combining the two surveys allows the researcher to train a network formation model on the ARD sample, and then predict connections between the households in the census survey. Several statistical methods exist for generating network statistics using ARD data, including Bayesian (Breza et al., 2020) as well as penalized rank approaches (Alidaee, Auerbach, & Leung, 2020).

The advantages to ARD is that it is potentially cheaper to conduct, potentially 70%–80% less expensive than a full network survey (Breza et al., 2020). However, ARD still relies on a census of the relevant population, and although the survey length can be short, transportation costs may be significant in a variety of rural contexts. ARD questions also requiring identifying distinct social subgroups, and therefore may require a number of questions that respondents find difficult or uncomfortable to answer. For example, asking questions about criminality is likely to identify an important space in the social topography, yet is a difficult question to pose to an entire population. Finally, ARD only provides data on local network statistics, but not the edges between nodes, which are often a subject of microeconomic interest.

5 Applications

Research on network effects in agricultural economies comprises how the flow of information and goods along the social ties that bind households within a community, and the longer-distance ties between rural and urban households, affects local

[2]The online appendix to Breza, Chandrasekhar, McCormick, and Pan (2020) provides a longer discussion over how to create ARD questions and conduct an ARD census and survey.

economic behavior and decision-making. Network theory has produced insights into a number of topics of significant economic and policy-related interest, a nonexhaustive list including the diffusion of technology and migration patterns, which continue to be active areas of research. There are naturally other areas in the study of agricultural economies where social learning and information flows play an important role in mediating behavior, and as such network theory can be fruitfully applied.

If consistent technological progress has enabled near-perpetual growth in agricultural productivity in high-income countries, the persistent gulf between the productivity level of farmers in lower-income countries has been attributed to the slow diffusion of new technologies. Technology diffusion is a process that has inherent network effects: learning how best to deploy and use new agricultural technologies has significant social spillovers, and there are local complementarities as more people adopt a given technology (Maertens & Barrett, 2013). The flow of credit within agricultural communities has long been understood as inherently dependent on social ties and standing. However, the limitations of microcredit to produce transformative effects on entrepreneurship and incomes in agrarian societies has prompted researchers to look into the way in which communities exchange credit and self-insurance in lieu of formal credit markets. Networks have long been understood to direct migration flows from rural to urban areas, as well as across the world. Networks affect not just where a person chooses to migrate, but also can play a role in determining whether they choose to migrate or not. Research on network effects can be divided into network theoretic and reduced-form approaches. Network theoretic studies often specify a model of network formation, and utilizer full network data on linkages between nodes. Reduced-form studies typically take networks as given, and lack full network data.

5.1 Information flows and network structure

Underpinning the domain-specific applications of network theory to the study of financial flows, technology adoption, and migration, a body of research has focused on the dynamics of information diffusion and aggregation within agrarian societies and how network structure affects this. Key topics include how heterogeneity within a community at the level of each node affects information aggregation, methods for identifying more central nodes, and how interventions can change the underlying network structure.

In information aggregation models, nodes use the experiences of other agents in their neighborhood to form their beliefs. Not all linkages are equal, however, as an agent will differentially weight information from nodes that it may consider to be more trustworthy or better informed. What characteristics then make information from a node more or less salient? BenYishay, Jones, Kondylis, and Mobarak (2020) look at the role that gender plays in how information is aggregated within a network. Studying the transmission of information on maize farming practices in 143 villages in Mali, the authors find a significant distrust of information coming from female farmers, even when conditioned on their knowledge of the subject and

their teaching ability. Male farmers distrust information coming from female farmers, and as a result, underweight the signals that they receive from them. Despite the presence of valuable information in the network, discriminatory practices limit the extent to which it is effectively aggregated. Beaman and Dillon (2018) conduct an experiment to study the diffusion of agricultural information by gendered social networks. In the context of farmers with incomplete information over new technology, targeting central nodes may exacerbate information asymmetries in the presence of gender segregation in social networks. The authors use partially complete network data from fifty-two villages in Mali to seed information, choosing nodes either at random, according to betweenness centrality, or degree centrality. They find that information diffusion declines with social distance, suggesting frictions in the transmission of new information. Importantly, given limited diffusion of information, the targeting of more central nodes led to lower transmission of information to women, suggesting significant gender effects in who receives as well as sends information. Given significant node level heterogeneity in both information set and returns to technology adoption, it is unclear whether individual adoption choices are optimal or not.

Discriminatory communication preferences present barriers to information diffusion, even when social linkages nominally connect households. Two conceptually similar studies on information flows investigate the degree to which existing network connections can be used to properly aggregate information, and find considerable interpersonal frictions. Banerjee, Breza, Chandrasekhar, and Golub (2018) study the diffusion of information on India's rushed 2016 demonetization policy, finding counterintuitively that widely broadcasting information may lead to the less effective aggregation of information than deliberately seeding it. Conducting an experiment in 225 villages in Odisha, India, the authors compare the efficacy of seeding information to people chosen for their propensity to spread information ("gossipy-ness," a concept developed in Banerjee, Chandrasekhar, Duflo, & Jackson, 2014) versus widely broadcasting the information in terms of engagement in social learning, policy knowledge, and an financially incentivized choice. The authors find that when individuals are expected to have common knowledge of a policy, they are less willing to seek advice and information from others in their network. Widely broadcasting information means that the correct information exists in dispersed form across the network, but individuals are reluctant to ask others out of embarrassment. This exact friction is investigated further in Chandrasekhar, Golub, and Yang (2018), who looks at how shame can constitute a serious barrier to communication, even across established network ties.

The concept of what makes a node central in terms of diffusing information is further explored in the question of whether social networks can also be used to provide information on their own structure. People inherently possess significant information about their neighbors, which can be used to identify nodes with particular characteristics. For example, appropriately measuring and targeting poor households is a persistent problem for aid and relief organizations. Alatas, Banerjee, Hanna, Olken, and Tobias (2012) and Alatas, Banerjee, Chandrasekhar, Hanna, and

Olken (2016) study the degree to which nodes in a social network have information about how poor their neighbors are. The authors find that networks in 640 villages in Indonesia are efficient aggregators of hard-to-predict information on poverty. The common knowledge of network structure held within the network itself is exploited by Banerjee, Chandrasekhar, Duflo, and Jackson (2019) to identify central individuals. Optimally seeding information in a diffusion model offers the potential for significant gains, where a highly central seed may quickly disperse information, while a seed on the periphery may lead to information failing to saturate a network. Identifying central seeds, however, requires extensive network survey data, and so the authors investigate the capacity for people within a community to identify central seeds. They have members of 213 villages in Karnataka, India, rank who they considered to be the best at diffusing information, their "gossipy-ness." The authors use an experiment in the villages in Karnataka as well as an additional 521 villages in Haryana, India, to show that not only are these community-identified seeds significantly better at diffusing information, they also meet the criteria of diffusion and eigenvector centrality (Jackson, 2014). This correspondence between community-selected seeds and their estimated centrality suggests the possibility of simply asking local communities for the ideal seeds, allowing for better-targeted interventions.[3]

5.2 Credit

One of the motivating reasons for the development of microfinance institutions was the belief in an unmet demand for formal credit products. Researchers observed complex, sometimes exploitative-seeming financial arrangements in rural and agricultural settings, where access to credit was both expensive and determined by an individual's connections and location within a social network. Households in prominent social positions had greater access to credit, while less well-connected households had less. The introduction of formal credit markets could provide these less-connected members of a community with access to reasonably priced credit on demand. Despite its theoretical tractability and decades of research and development into microfinance, in many contexts these lending institutions have suffered a failure to diffuse or displace informal lending institutions.

To try to understand how information about the availability of microcredit diffuses through a community, Banerjee et al. (2013) conduct a network study of 43 villages in Karnataka, India. The microcredit institution in the study relied on word-of-mouth communication to reach potential borrowers, people who might be interested in acquiring loans. In remote areas such as rural Karnataka, there are not many other feasible methods for informing a community about a new source of credit. The microcredit institution attempted to disseminate information by

[3]Beaman, Keleher, Magruder, and Trachtman (2021) conduct a similar community-led targeting scheme in an urban community in Monrovia, Liberia, but find only modest gains over random targeting.

identifying influential people within a community and then asking them to spread the word. The authors find that the people identified as influential varied considerably in their efficacy at diffusing information about microcredit within their community, highly correlated to their centrality measure within their social network.

The work by Banerjee et al. (2013) is notable in several ways. First, it explicitly makes the case for the value of network-based targeting as important for facilitating a diffusion process. Second, the authors provide a full network theoretic study of a diffusion process in an agricultural context. While there already existed a sizable literature dealing with networks in agricultural economics, data on actual network structure had been scarce. Banerjee et al. (2013) were among the first to employ measures of node and network level characteristics in an applied context. The study further employs different measures of centrality in its estimation of the diffusion of information, showing how certain measures (i.e., diffusion centrality) better correlate with the spread of information than other measures of centrality, such as degree. Finally, the authors created a rich data set of social networks measured across twelve dimensions, which has led to the use and reuse of this data by other studies, which we discuss further in Section 6.2.

The provision of microcredit and the introduction of formal financial institutions can also impact the very structure of a social network. Heß, Jaimovich, and Schündeln (2020) study the impact of development projects on fifty-six rural villages in the Gambia, and find that these interventions reduced overall social connectivity. The authors argue that the reduction in social connectivity over time is partially the result of households becoming wealthier, and therefore less reliant on interpersonal relationships to substitute for missing markets or backstop them in times of crisis. They also argue that the returns from the development projects were unevenly distributed, increasing the centrality of specific interventions within a network. Elite capture and favoritism might therefore focus social ties on a smaller set of individuals, leading to fewer overall interactions. Similarly Banerjee, Chandrasekhar, Duflo, and Jackson (2021) find in Karnataka, India, that the introduction of a microfinance institution crowds out social network relations, creating a network-level externality. The study finds that even people who were not receiving microcredit ended up with fewer social connections; the outside option provided by microfinance loans decreased the value to social interactions everywhere.

5.3 Risk sharing

The diffusion of microcredit lending throughout an agrarian community can be seen as an alternative technology to informal risk-sharing networks. Agricultural households face numerous risks from inclement weather, pests and disease, and volatile local markets. For households with only limited savings and without easy access to insurance and credit markets, a single bad harvest can leave them below subsistence levels. To compensate for this extant risk, households often create mutual commitments to pool risk, offering money, food, labor, and other forms of assistance to a

member who receives a negative shock. While the flow of goods and services from better- to worse-off households is tractable, or equal interest is the information—trust and social capital—that sustain these relations. These risk-sharing networks require that households commit to transfers to one another with only the expectation of reciprocal aid at some point of the future. Given the lack of any contractual-type enforcement mechanisms, the network structures and information that allow these risk-sharing networks to persist and function has been a major topic of study (De Weerdt & Dercon, 2006; Fafchamps & Gubert, 2007; Fafchamps & Lund, 2003).

Jackson, Rodriguez-Barraquer, and Tan (2012) study favor exchanges networks in 75 villages in Karnataka, India, where they argue that the network structure is crucial to sustaining the promise of reciprocation of favors rendered. A risk-sharing agreement between any two households may be difficult to sustain over a long period of time, as the frequency with which the favor may be repaid is possibly too low to incentive cooperation. Why provide assistance today when the need for reciprocation may be years away, with plenty of opportunities for the member receiving the transfer in the present to shirk in the future? It is through embedding the exchange of favors between two parties within a network that cooperation becomes sustainable, allowing for the network to collectively sever ties with nonparticipants and shirkers. The result, which the authors illustrate using data on favor networks in Karnataka is that edges between nodes are more frequently supported by a connection to at least a third node to increase the stability of the arrangement. The use of social connections and close ties is necessary to create the social capital that allows these risk-sharing networks to function (Ambrus, Mobius, & Szeidl, 2014; Karlan, Mobius, Rosenblat, & Szeidl, 2009).

Strong social ties are a precondition for effective risk-sharing networks. Parties must be able to monitor each other to confirm whether someone asking for money or goods has an actual need of a transfer, or that someone who says that they cannot make a transfer due to receiving a negative shock, has in fact received that alleged shock. As a result, geographic proximity and kinship relations strongly predict participation in risk-sharing networks (Fafchamps & Gubert, 2007). Attanasio, Barr, Cardenas, Genicot, and Meghir (2012) use an experiment to study the formation of risk-sharing groups in 70 communities in Colombia, finding that trust and familiarity between members of a network, often through familial relations, is the primary determinant of the durability of a risk-sharing network.

The geographic locality of risk-sharing networks is an unattractive quality, as covariate risks are more likely to occur within a small region, and therefore affect all members of the networks. For a group of households in close geographic proximity who are engaged in a risk-sharing network, an outbreak of crop disease in the region is likely to affect all of them, leading to a case where none of the farmers will have a surplus which they can transfer to the others. Jack and Suri (2014) study a mobile money technology (M-PESA in Kenya), which reduces the transaction costs associated with remitting income, allowing for a larger geographic spread, and thus more efficient, risk-sharing network.

5.4 Migration

There is a long history of migratory specificity, where people from one region or community send migrants to another geographically specific region. This specificity in migration flows occurs even over large distances, and despite the large number of notionally similar locations. What appears to drive this pattern of movement is that people prefer moving to areas where they can benefit from social networks already in place. Migration networks provide information and support to new arrivals: they help people find jobs and housing at their destination, and they provide a community and similar cultural traditions for those far from home. What distinguishes the study of migration networks from other types of networks is that they are geographically disparate by definition, with edges between nodes stretching across countries and even continents. While the study of networks in agrarian societies is often demarcated by village and region, migration networks connect the hinterlands with urban areas around the world, facilitating the movement of people, money, and goods.

Some of the earliest work on migration networks emerged from a sociological tradition studying migration from Mexico to the United States. Despite a static wage differential between the two countries, migration steadily increased over time. A "network effect" was used to explain this pattern, where the increasing density of networks of migrants in the United States lowered the cost of moving for future migrants, leading to an increase in flows (Durand, Massey, & Charvet, 2000; Massey & Espana, 1987). These studies build on earlier sociological research on the "strength of weak ties" in networks, where information, aid, and other goods flow across what appear to be relatively weak connections between individuals (Granovetter, 1978). Networks of earlier migrants provide information on how to best cross borders, as well as find jobs. Job referral networks connect new migrants with access to employment upon arrival. Social networks further provided material support necessary to establishing oneself in a new country or region. It is important to distinguish between two of the theorized roles that social networks play in determining migration choices and outcomes. The first function of networks is informational: they can provide crucial information about how to make the necessary journey, how to obtain the proper documentation or else get by without it, how to find a job, where to go, as well as the types of work and living conditions that are available—ultimately whether a temporary or permanent move would even be worthwhile. The second function that networks provide is a material one, offering credit and loans to potential migrants, as well as places to stay upon arrival, and access to jobs (McKenzie & Rapoport, 2010; Munshi, 2003).

There are frictions in the transmission of information across migration networks. One potential explanation for low rates of rural-to-urban migration in low-income countries, despite the existence of significant spatial productivity and wage gaps, is that many migrants do not accurately signal their returns to working in urban areas. Baseler (2021) conducts a randomized control trial in fifteen rural villages in western Kenya, studying how much households knew about the wage distribution in the three nearest cities. Baseler finds that urban migrants strategically underreport their

earnings in the city to their family and friends back home so as to reduce expectations over the amount of remittances they send back. What is notable about this finding is that strategic underreporting is both widespread, but also undetected by family members back home. Despite the likely strength of the social ties between family members, accurate information on wages is not being transmitted across these networks.

While social networks at the destination may serve to increase migration rates by reducing the start-up costs of moving to that location, social networks at the origin may conversely inhibit rural-to-urban migration. Munshi and Rosenzweig (2016) point to risk-sharing networks in rural areas as a primary impediment to greater mobility. The authors argue that village-based risk-sharing networks require locality, members of a risk-sharing network must be able to credibly punish defectors from failing to uphold their end of the risk-sharing agreement. As households gain access to alternative sources of income and their outside option improves, defection is no longer so easily punished. Second, remittances are more easily hidden than locally derived income, and so a household sending migrants can more easily pretend not to have sufficient cash on hand to help mitigate others' losses. Because sending migrants to gain access to wages means that a household effective reduces its ability to participate in risk-sharing networks, this leads to lower overall mobility, particularly in villages with more effective risk-sharing networks.

Meghir, Mobarak, Mommaerts, and Morten (2021) build on an earlier RCT that subsidized temporary labor migration in villages in Bangladesh to study its spillover effects on local risk sharing (Bryan, Chowdhury, & Mobarak, 2014). The authors find that the experimentally induced increase in migration improved risk sharing within each village, leading to increased transfers between households and a greater willingness to help others, as well as reduced correlation between household consumption and income. They develop a model in which migration is a risky endeavor, and local risk sharing at the village allows households to invest in it. As the risk of migration goes down and the relative payoff goes up, the importance of local risk-sharing declines. In this way a short-term migration subsidy can increase risk sharing, while a longer-term shift to more migration would lead to less risk sharing (Morten, 2019).

5.5 Technology

The role that social networks play in the diffusion and adoption of new technologies has been an object of study due to the relatively slow adoption of what are considered potentially promising agricultural technologies, such as fertilizers, high yield varietal seeds (HYVs), and innovative farming practices that have been essential to the sustained increase in agricultural productivity in higher-income countries. Despite the widespread availability of these technologies, their adoption exists in a state of arrested development, with only a fraction of farmers in lower-income countries adopting these technologies and inputs. A lack of widespread adoption may not be altogether surprising, given the significant investment made into agricultural

extension services in the United States and Europe, and the establishment of large land-grant universities dedicated to both the development and dissemination of new agricultural technologies. Disseminating existing technologies remains a challenge in poorer countries, where access to cellphones and other communication technologies are often more limited, and the farmers most in need can be located in relatively inaccessible areas. Disseminating a technology in an agricultural community necessarily then relies on experimentation, word-of-mouth, and the use of others' experiences to decide whether to adopt or not. Technology is also studied in the context of network analysis because of the role of social learning. New technologies can be novel and complex, requiring that people first learn how to use them, then how to properly adapt them to their own context. Groups of people can learn from each other, transmitting information about the proper use or advantage of a new technology, and producing a consensus that will spread to the entire community.

The application of social network theory to the adoption of a new technology typically models the decision as a diffusion process, with the most salient questions being who comprises a farmer's learning network, and the learning dynamics relevant for a specific technology. Two canonical studies of social networks on the use and adoption of agricultural technologies are Foster and Rosenzweig (1995) and Conley and Udry (2010), which focus on the substitution between what an individual farmer learns experientially about a new technology and what they can learn from others' experiences. Foster and Rosenzweig highlight the role that limited information and uncertainty play in the slow adoption of HYVs. The farmers that they study in India are uncertain about the best way to deploy HYVs, the right combination of complementary inputs to use with them, and the expected increase in profit from adopting this technology. Farmers also use the experiences of those around them to increase how profitably they themselves can use HYVs. The combination of learning-by-doing and learning-from-others creates a free-riding problem, where later adoption is more profitable than earlier adoption, with earlier costly adoption by some farmers a necessary precondition for adoption of these seeds by later farmers.

Conley and Udry (2010) employ detailed network data, looking for evidence of social learning in the adoption of a new technology. Specifically, the authors study how farmers in Ghana calibrate their use of inputs when growing a new, potentially lucrative crop (pineapples). One of their concerns is that local adoption and input decisions are determined by local factors, unrelated to the exchange of information between farmers. To address this "reflection problem" the authors use the timing of staggered planting of crops to isolate which farmers could have learned from which. They find evidence that farmers respond to the input choices of their neighbors, increasing and decreasing the use of fertilizer in response to the outcomes of their neighbors, and with a greater responsiveness to their neighbors' outcomes if they were more recent adoptees. Other studies corroborate these learning spillover dynamics, and test how they work differentially for different types of crops (Munshi, 2004). In the Zambezia region of Mozambique Bandiera and Rasul (2006) study the initial diffusion of a technology rather than social learning over

its characteristics, and recover an inverse U-shaped pattern to adoption. They argue that initial adoption is increasing in the number of adopters, then decreasing as the number of adopters grows large.

Maertens (2017) studies the adoption decision of a genetically modified (GM) crop, Bt Cotton, as function of how farmers learn both about the profitability of the new seeds, as well as how others view the technology. GM crops are often viewed suspiciously, as potential hazards to animal and human health. Farmers are concerned that if they adopt this new technology they may be blamed for any illness in their village. Within the context of three villages in India, Maertens finds two related learning networks, a group of farmers considered the best or most technologically proficient, whom others in their community learn from about the potential profits of adopting the GM crop, and then the broader social network which farmer learn about the overall social acceptability of GM crops. There are competing effects in this study: adoption by peers signals that the technology is acceptable to them, leading to higher rates of adoption, while adoption by the more technologically proficient farmers leads to lower rates of adoption, as others delay their decision to better observe their outcomes.

Social network effects from peers can play a significant role in the adoption of other types of agricultural technologies. Cai, De Janvry, and Sadoulet (2015) study how information over weather insurance diffuses through farmers' social networks in China, finding that information over how the insurance product actually works is transmitted between peers, regardless of whether each member the network chooses to adopt. They also find, unsurprisingly, that more central individuals in each network are more effective at diffusing information. Beaman, BenYishay, Magruder, and Mobarak (2021) use information on the centrality of farmers in Malawi to test the optimal method of seeding information on agricultural technologies. The authors compare the value of seeding technology to farmers chosen by extension agents, with those chosen by network theory as optimal seeds under a contagion or complex contagion model. The complex contagion model, described in Section 2.3.1, requires that the number of edges exceed 1 in order to induce adoption in connected nodes. This threshold model is useful in explaining the case of failed adoption. For example, the practice of pit planting entirely failed to diffuse in a number of the villages. The authors find that the targeting of more central nodes significantly outperforms the selection of nodes by extension agents, but the selection of nodes according to complex contagion does not significantly outperform nodes selected according to a simple contagion model.

Rather than allow for the transmission and aggregation of information within a network to procedure organically, BenYishay and Mobarak (2019) conduct an RCT to test the efficacy of incentives in spreading information about new planting technologies in Malawi.[4] The authors select either a government-employed extension

[4]Maertens, Michelson, and Nourani (2021) highlight the difficulty that farmers have in learning new technologies from extension agents.

worker, a "lead farmer," or peer farmer as the primary communicator to help diffuse a technology, and then provide performance-based incentives based on their outcomes. The authors find the incentives crucial to inducing initial adoption among the chosen lead and peer farmers, but also that incentivized peer farmers outperform lead farmers. Shikuku, Pieters, Bulte, and Läderach (2019) likewise corroborate the importance of incentives in inducing seeds to diffuse information about the technology, and use both material and social rewards to motivate the dispersal of information.

6 Caveats and future directions

There are a number of caveats that researchers studying networks in agricultural economies should consider in pursuing research in this field. The value of network information in promoting practical policy objectives has not been extensively tested in real-world contexts. If social networks are particularly salient to economic behavior and decision-making in agricultural communities, these are also the areas in which generating accurate information on the relevant social graph may be the most difficult, time consuming, and expensive. Whether generating this data is ultimately a cost-effective strategy to promote a variety of policy outcomes remains a point of contention. In a similar vein, there is a concern as to the generalizability of the effects of social networks across geographic and cultural areas. A goal central to much of the work in economics is finding behaviors and effects that can be transposed between contexts, holding true, at least on average, for individuals, households, or firms. As the study of social networks in agricultural economies abuts work done in sociology and anthropology, the degree to which learning about localized network properties and effects will translate to other cultural contexts is a concern. Given the relative paucity of data on networks in agricultural economies, some caution is necessary in assuming that the findings presented so far will hold everywhere.

Predictions are notoriously difficult to make, especially about the future, and potential directions for the study of networks in agrarian economies are no exception. We have presented work on credit, risk-sharing, technology adoption, and migration as they have historically been important areas of research on the topic of network studies, and are likely to continue to be so. This chapter cannot provide a comprehensive overview of all the promising areas of work in the study of networks in agricultural economies, and as such there are many topics that will be valuable to this field of study which have not been included in this chapter. Two avenues of research that appear promising include those that aim to reduce the burden of network data collection, and those that increase the spread or efficacy of peer-effect interventions. The latter are particularly suited to digital communication technologies, particularly those surrounding cellphone communications, which allow both the rapid dissemination of information into a network, and allow members of agricultural communities new methods for extending and maintaining their social networks.

6.1 Optimal seeding strategies

The idea of seeding information or goods, providing them to a subset of people within a community, is an intuitively tractable method of instigating their diffusion or adoption. Seeding, rather than providing the information or good to everyone in a community, may be necessary due to inherent capacity or cost constraints. The information or goods may be complex, and therefore require time and study for people to fully understand and utilize them in their own context. There may also be complex learning dynamics and threshold effects in the underlying diffusion process, making the choice of initial seeds consequential to whether the diffusion process is ultimately successful. In theory, a researcher or a policy-maker could produce a more complete diffusion process by selecting a more central seed than a randomly chosen one. When the idea or good can be seeded to multiple individuals, the optimal choice of seeds is difficult (NP-hard), and solutions will vary by the underlying diffusion or cascade model (Kempe, Kleinberg, & Tardos, 2003, 2005).

If the difference in potential outcomes between an optimal seeding strategy and a less-optimal one is large, there are still significant attendant costs that come with generating the necessary data to optimally select seeds. Censuses and network surveys are expensive and difficult to undertake, and even the use of ARD methods or using the network itself to identify central nodes (e.g., "gossips") still comes with punishing data requirements. Rather than test whether more central seeds outperform randomly selected seeds in instigating a diffusion process, Akbarpour, Malladi, and Saberi (2020) suggest that the question should be whether we should select more seeds. In their paper, the authors consider instigating a simple diffusion process under a standard SIR model. Over time the diffusion process will either end in success or failure; it either saturates the network or it dies out. In instances of successful diffusion, optimal seeding strategies would increase the speed at which the process saturates the network, but not fundamentally change the outcome. The value of this increased speed in diffusion is highly context dependent, perhaps high in the case of a pandemic, perhaps less so in the case of minor agricultural technologies and innovations. If a diffusion process is going to be ultimately unsuccessful with a fixed number of seeds, then either a random or optimally targeted seeding strategy will perform equally dismally. The only parameter space in which optimal seeding matters is the narrow window in which the centrality and connectedness of a seed makes the difference between success or failure, such as the case described by a complex contagion model. Akbarpour et al. argue that in the context of the diffusion models used in Banerjee et al. (2013) and Cai et al. (2015), the addition of one to three additional random seeds would perform as well at instigating the diffusion process as using network-informed targeting.

The finding that a strategy of more seeds can outperform a strategy of selecting more central seeds in Akbarpour et al. (2020) is a theoretical one. In a diffusion model of complex contagion or one with complex learning dynamics (e.g., where widely broadcasting information inhibits its spread), a few more seeds may lead to an incomplete diffusion process, while optimally selected ones will saturate the network. Indeed, the incomplete diffusion of new technologies and information

related to credit and lending, agricultural productivity, and migration in agrarian economies—one of the motivating reasons for the study of network dynamics—suggests the insufficiency of a simple model of diffusion. Yet while the costs of network data collection vary by context, the information and goods studied in agrarian economies often share the attribute of being inexpensive. If the addition of a few additional, randomly chosen seeds produces a diffusion outcome that is approximately as effective as the case when the network structure has been entirely mapped and understood, it is difficult to find a use-case where the study of network structure is practically viable. Moreover, the increasing complexity of adding layers of strategic learning and behavioral network dynamics to diffusion models may end up describing a "piranha problem," where the sheer number of network effects would interfere with each other in unpredictable ways (Tosh et al., 2021).

6.2 Data reuse

Data on agricultural communities are expensive to come by and network data are even more so. Where many subfields have a set of data that provides the basis for much of the research on that topic, the data that support much of the literature on network effects in agricultural economies is taken from under 100 villages in Karnataka, southwestern India. The network data collected for the 2012 paper "The diffusion of microfinance" has become a repository that has been revisited to test many new network theories.[5] Over the course of a decade, approximately 25 groups of researchers have used this census data to write over 50 papers on network theory and its applications. This data have been used in the study of public health (Shakya, Christakis, & Fowler, 2015; Yang, McKhann, Chen, Harling, & Onnela, 2019), social monitoring (Banerjee, Chandrasekhar, et al., 2019; Breza & Chandrasekhar, 2019; Chandrasekhar, Kinnan, & Larreguy, 2018), individual knowledge of network structure (Breza, Chandrasekhar, & Tahbaz-Salehi, 2018), learning patterns (Chandrasekhar et al., 2020), network theory (Awan et al., 2020; Banerjee, Chandrasekhar, et al., 2019; De Paula, 2020; Dutta, Mira, & Onnela, 2018; Montes et al., 2020; Montes, Jimenez, & Onnela, 2018; Omodei & Arenas, 2016; Salter-Townshend & McCormick, 2017; Song, 2018), and other applications (Davidson & Sanyal, 2017; Shakya, Christakis, & Fowler, 2017); this list is not exhaustive. Other research on networks that does not visit these data have been drawn nearby villages, only 200 km distant (Johny, Wichmann, & Swallow, 2017).

The social network structures present in Karnataka from the mid-2000s onwards may be representative of social network structures elsewhere in agricultural economies. Certainly the data painstakingly gathered over many years in this geographic area are both valuable and unique. The specificity of this data however, in terms of geography, culture, and time at which it was gathered, may limit the representativeness of the results on network structures to other agricultural economies and new contexts.

[5]The network data gathered by "Learning about a new technology: Pineapple in Ghana" has spawned a similar, if smaller literature.

6.3 Future directions

Potential future directions on the study of networks in agricultural economies appear to be those that leverage new methods of data collection to recover network statistics, as well as those that elide the structure of networks entirely, and instead look for peer effects in the widespread dissemination of information or goods. Continued development of network surveys and the use of ARD methods will allow researchers greater access to larger, more varied datasets. This in turn will allow a greater correspondence between many of the theoretical developments made on the study of network effects and learning models with real-world examples. This may be particularly relevant for models of diffusion and aggregation, which make sharp predictions over the value of seeding strategies, and the learning dynamics that allow the spread and aggregation of information. More psychologically-grounded theories too are being developed to consider the uniquely social elements of these networks, wherein the behavioral and emotional state of nodes, as well as the history of their relationship to others, can play a role in the efficacy of transmitting information along edges connecting them.

New work in the field is looking at the peer effects of widely broadcasted messages, and the regional impact of choosing different types of people with whom to seed information and technologies. The increasing rate at which individuals in agricultural economies have access to digital communication technologies, from cellphones with SMS, to internet connected phones with more sophisticated communication and financial applications (e.g., WhatsApp) allows for the widespread dissemination, and even individual targeting of information. This research relies on theories of network dynamics and social learning, even if it does not produce data on the structure of the network.

Acknowledgments

The chapter has benefited from comments from Hossein Alidaee, Chris Barrett, Annemie Maertens, and two anonymous reviewers. Elizabeth Dolan provided research assistance.

References

Acemoglu, D., Dahleh, M. A., Lobel, I., & Ozdaglar, A. (2011). Bayesian learning in social networks. *The Review of Economic Studies*, *78*(4), 1201–1236.

Acemoglu, D., Ozdaglar, A., & Yildiz, E. (2011). Diffusion of innovations in social networks. In *2011 50th IEEE conference on decision and control and European control conference* (pp. 2329–2334).

Akbarpour, M., Malladi, S., & Saberi, A. (2020). *Just a few seeds more: Value of network information for diffusion*. Available from SSRN. https://dx.doi.org/10.2139/ssrn.3062830.

Alatas, V., Banerjee, A., Chandrasekhar, A. G., Hanna, R., & Olken, B. A. (2016). Network structure and the aggregation of information: Theory and evidence from Indonesia. *American Economic Review*, *106*(7), 1663–1704.

Alatas, V., Banerjee, A., Hanna, R., Olken, B. A., & Tobias, J. (2012). Targeting the poor: Evidence from a field experiment in Indonesia. *American Economic Review*, *102*(4), 1206–1240.

Alidaee, H., Auerbach, E., & Leung, M. P. (2020). *Recovering network structure from aggregated relational data using penalized regression*. arXiv. https://arxiv.org/abs/2001.06052.

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–676.

Ambrus, A., Mobius, M., & Szeidl, A. (2014). Consumption risk-sharing in social networks. *American Economic Review*, *104*(1), 149–182.

Aronow, P. M., & Samii, C. (2017). Estimating average causal effects under general interference, with application to a social network experiment. *The Annals of Applied Statistics*, *11*(4), 1912–1947.

Attanasio, O., Barr, A., Cardenas, J. C., Genicot, G., & Meghir, C. (2012). Risk pooling, risk preferences, and social networks. *American Economic Journal: Applied Economics*, *4*(2), 134–167.

Awan, U., Morucci, M., Orlandi, V., Roy, S., Rudin, C., & Volfovsky, A. (2020). Almost-matching-exactly for treatment effect estimation under network interference. In S. Chiappa, & R. Calandra (Eds.), *Proceedings of the twenty third international conference on artificial intelligence and statistics, August: Vol. 108* (pp. 3252–3262). PMLR. http://proceedings.mlr.press/v108/awan20a.html.

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

Banerjee, A., Breza, E., Chandrasekhar, A. G., & Golub, B. (2018). *When less is more: Experimental evidence on information delivery during India's demonetization*.

Banerjee, A., Breza, E., Chandrasekhar, A. G., & Mobius, M. (2019). *Naive learning with uninformed agents*.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). *Science*, *341*(6144), 1236498.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2014). *Gossip: Identifying central individuals in a social network*.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, *86*(6), 2453–2490.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2021). *Changes in social network structure in response to exposure to formal credit markets*.

Banerjee, A. V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, *107*(3), 797–817.

Baseler, T. (2021). *Hidden income and the perceived returns to migration: Experimental evidence from Kenya*. Available from SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3534715.

Bayer, P., Ross, S. L., & Topa, G. (2008). Place of work and place of residence: Informal hiring networks and labor market outcomes. *Journal of Political Economy*, *116*(6), 1150–1196.

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021). Can network theory-based targeting increase technology adoption? *American Economic Review*, *111*(6), 1918–1943. https://doi.org/10.1257/aer.20200295.

Beaman, L., & Dillon, A. (2018). Diffusion of agricultural information within social networks: Evidence on gender inequalities from Mali. *Journal of Development Economics*, *133*, 147–161.

ARTICLE IN PRESS

Beaman, L., Keleher, N., Magruder, J., & Trachtman, C. (2021). Urban networks and targeting: Evidence from liberia. *AEA Papers and Proceedings*, *111*, 572–576. https://doi.org/10.1257/pandp.20211061.

Beaman, L. A. (2012). Social networks and the dynamics of labour market outcomes: Evidence from refugees resettled in the US. *The Review of Economic Studies*, *79*(1), 128–161.

Beck, F., Burch, M., Diehl, S., & Weiskopf, D. (2017). A taxonomy and survey of dynamic graph visualization. In *Computer graphics forum: Vol. 36* (pp. 133–159).

BenYishay, A., Jones, M., Kondylis, F., & Mobarak, A. M. (2020). Gender gaps in technology diffusion. *Journal of Development Economics*, *143*, 102380.

BenYishay, A., & Mobarak, A. M. (2019). Social learning and incentives for experimentation and communication. *The Review of Economic Studies*, *86*(3), 976–1009.

Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, *100*(5), 992–1026.

Bloch, F., Jackson, M. O., & Tebaldi, P. (2021). Centrality measures in networks. *arXiv*. https://arxiv.org/abs/1608.05845.

Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of Peer effects through social networks. *Journal of Econometrics*, *150*(1), 41–55.

Breza, E., Chandrasekhar, A., Golub, B., & Parvathaneni, A. (2019). Networks in economic development. *Oxford Review of Economic Policy*, *35*(4), 678–721.

Breza, E., & Chandrasekhar, A. G. (2019). Social networks, reputation, and commitment: Evidence from a savings monitors experiment. *Econometrica*, *87*(1), 175–216.

Breza, E., Chandrasekhar, A. G., McCormick, T. H., & Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, *110*(8), 2454–2484. https://doi.org/10.1257/aer.20170861.

Breza, E., Chandrasekhar, A. G., & Tahbaz-Salehi, A. (2018). *Seeing the forest for the trees? An investigation of network knowledge*.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in Bangladesh. *Econometrica*, *82*(5), 1671–1748.

Cai, J., De Janvry, A., & Sadoulet, E. (2015). Social networks and the decision to insure. *American Economic Journal: Applied Economics*, *7*(2), 81–108.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, *113*(3), 702–734.

Chandrasekhar, A. G., Golub, B., & Yang, H. (2018). *Signaling, shame, and silence in social learning*.

Chandrasekhar, A. G., Kinnan, C., & Larreguy, H. (2018). Social networks as contract enforcement: Evidence from a lab experiment in the field. *American Economic Journal: Applied Economics*, *10*(4), 43–78.

Chandrasekhar, A. G., Larreguy, H., & Xandri, J. P. (2020). Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, *88*(1), 1–32.

Chandrasekhar, A. G., & Lewis, R. (2016). *Econometrics of sampled networks*. Accessed October 20, 2020, from http://stanford.edu/arungc/CL.pdf.

Choi, S., Gale, D., & Kariv, S. (2005). Behavioral aspects of learning in social networks: An experimental study. *Advances in Applied Microeconomics*, *13*, 25–61.

Choi, S., Gale, D., & Kariv, S. (2012). Social learning in networks: A quantal response equilibrium analysis of experimental data. *Review of Economic Design*, *16*(2), 135–157.

Chuang, Y., & Schechter, L. (2015). Social networks in developing countries. *Annual Review of Resource Economics*, *7*(1), 451–472.

Conley, T. G., & Udry, C. R. (2010). Learning about a new technology: Pineapple in Ghana. *American Economic Review*, *100*(1), 35–69.

Davidson, T., & Sanyal, P. (2017). Associational participation and network expansion: Microcredit self-help groups and poor women's social ties in rural India. *Social Forces*, *95*(4), 1695–1724.

DeGroot, M. H. (1974). Reaching a consensus. *Journal of the American Statistical Association*, *69*(345), 118–121.

DeMarzo, P. M., Vayanos, D., & Zwiebel, J. (2003). Persuasion bias, social influence, and unidimensional opinions. *Quarterly Journal of Economics*, *118*(3), 909–968.

De Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics*, *12*, 775–799.

De Weerdt, J., & Dercon, S. (2006). Risk-sharing networks and insurance against illness. *Journal of Development Economics*, *81*(2), 337–356.

Durand, J., Massey, D. S., & Charvet, F. (2000). The changing geography of Mexican immigration to the United States: 1910–1996. *Social Science Quarterly*, *81*, 1–15.

Dutta, R., Mira, A., & Onnela, J.-P. (2018). Bayesian inference of spreading processes on networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *474*(2215), 20180129.

Fafchamps, M., & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, *83*(2), 326–350.

Fafchamps, M., & Lund, S. (2003). Risk-sharing networks in rural Philippines. *Journal of Development Economics*, *71*(2), 261–287.

Foster, A. D., & Rosenzweig, M. R. (1995). Learning by doing and learning from others: Human capital and technical change in agriculture. *Journal of Political Economy*, *103*(6), 1176–1209.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *The Quarterly Journal of Economics*, *129*(4), 1661–1710.

Goldsmith-Pinkham, P., & Imbens, G. W. (2013). Social networks and the identification of Peer effects. *Journal of Business & Economic Statistics*, *31*(3), 253–264.

Golub, B., & Jackson, M. O. (2010). Naive learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, *2*(1), 112–149.

Golub, B., & Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics*, *127*(3), 1287–1338.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology*, *83*(6), 1420–1443.

Griffith, A. (2021). Name your friends, but only five? The importance of censoring in Peer effects estimates using social network data. *Journal of Labor Economics (accepted)*. https://www.journals.uchicago.edu/doi/abs/10.1086/717935.

Grimm, V., & Mengel, F. (2018). Experiments on belief formation in networks. *Journal of the European Economic Association*, *18*(1), 49–82. https://doi.org/10.1093/jeea/jvy038.

Hanna, R., Mullainathan, S., & Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics*, *129*(3), 1311–1353.

Herman, I., Melancon, G., & Marshall, M. S. (2000). Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, *6*(1), 24–43. https://doi.org/10.1109/2945.841119.

Heß, S., Jaimovich, D., & Schündeln, M. (2020). *The Review of Economic Studies*, *88*(3), 1347–1384. https://doi.org/10.1093/restud/rdaa033.

Jack, W., & Suri, T. (2014). Risk sharing and transactions costs: Evidence from kenya's mobile money revolution. *American Economic Review*, *104*(1), 183–223.

Jackson, M. O. (2003). *A survey of models of network formation: Stability and efficiency*.

Jackson, M. O. (2010). *Social and economic networks*. Princeton University Press.

Jackson, M. O. (2014). Centrality, gossip, and diffusion of information in networks. In *Proceedings of the fifteenth ACM conference on economics and computation* (p. p. 1).

Jackson, M. O., Rodriguez-Barraquer, T., & Tan, X. (2012). Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, *102*(5), 1857–1897.

Jackson, M. O., & Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, *97*(3), 890–915.

Jackson, M. O., Rogers, B. W., & Zenou, Y. (2017). The economic consequences of social-network structure. *Journal of Economic Literature*, *55*(1), 49–95.

Jackson, M. O., & Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, *71*(1), 44–74.

Johny, J., Wichmann, B., & Swallow, B. M. (2017). Characterizing social networks and their effects on income diversification in rural Kerala, India. *World Development*, *94*, 375–392.

Karlan, D., Mobius, M., Rosenblat, T., & Szeidl, A. (2009). Trust and social collateral. *The Quarterly Journal of Economics*, *124*(3), 1307–1361.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 137–146).

Kempe, D., Kleinberg, J., & Tardos, É. (2005). Influential nodes in a diffusion model for social networks. In *International colloquium on automata, languages, and programming* (pp. 1127–1138).

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American Economic Review*, *111*(3), 831–870.

Maertens, A. (2017). Who cares what others think (or do)? Social learning and social pressures in cotton farming in India. *American Journal of Agricultural Economics*, *99*(4), 988–1007.

Maertens, A., & Barrett, C. B. (2013). Measuring social networks' effects on agricultural technology adoption. *American Journal of Agricultural Economics*, *95*(2), 353–359.

Maertens, A., Michelson, H., & Nourani, V. (2021). How do farmers learn from extension services? Evidence from Malawi. *American Journal of Agricultural Economics*, *103*(2), 569–595.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Massey, D. S., & Espana, F. G. (1987). The social process of international migration. *Science*, *237*(4816), 733–738.

McKenzie, D., & Rapoport, H. (2010). Self-selection patterns in Mexico-US migration: The role of migration networks. *The Review of Economics and Statistics*, *92*(4), 811–821.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Meghir, C., Mobarak, A. M., Mommaerts, C. D., & Morten, M. (2021). *Migration and informal insurance*.

Montes, F., Jaramillo, A. M., Meisel, J. D., Diaz-Guilera, A., Valdivia, J. A., Sarmiento, O. L., & Zarama, R. (2020). Benchmarking seeding strategies for spreading processes in social networks: An interplay between influencers, topologies and sizes. *Scientific Reports*, *10*, 3666.

Montes, F., Jimenez, R. C., & Onnela, J.-P. (2018). Connected but segregated: Social networks in rural villages. *Journal of Complex Networks*, *6*(5), 693–705.

Morten, M. (2019). Temporary migration and endogenous risk sharing in village India. *Journal of Political Economy*, *127*(1), 1–46.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics*, *118*(2), 549–599.

Munshi, K. (2004). Social learning in a heterogeneous population: Technology diffusion in the Indian Green Revolution. *Journal of Development Economics*, *73*(1), 185–213.

Munshi, K. (2014). Community networks and the process of development. *Journal of Economic Perspectives*, *28*(4), 49–76.

Munshi, K., & Rosenzweig, M. (2016). Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, *106*(1), 46–98.

Omodei, E., & Arenas, A. (2016). Untangling the role of diverse social dimensions in the diffusion of microfinance. *Applied Network Science*, *1*(1), 1–10.

Peterson, E., & Kagalwala, A. (2021). When unfamiliarity breeds contempt: How partisan selective exposure sustains oppositional media hostility. *American Political Science Review*, *115*(2), 585–598.

Salter-Townshend, M., & McCormick, T. H. (2017). Latent space models for multiview network data. *The Annals of Applied Statistics*, *11*(3), 1217.

Schwartzstein, J. (2014). Selective attention and learning. *Journal of the European Economic Association*, *12*(6), 1423–1452.

Shakya, H. B., Christakis, N. A., & Fowler, J. H. (2015). Social network predictors of latrine ownership. *Social Science & Medicine*, *125*, 129–138.

Shakya, H. B., Christakis, N. A., & Fowler, J. H. (2017). An exploratory comparison of name generator content: Data from rural India. *Social Networks*, *48*, 157–168.

Shikuku, K. M., Pieters, J., Bulte, E., & Läderach, P. (2019). Incentives and the diffusion of agricultural knowledge: Experimental evidence from Northern Uganda. *American Journal of Agricultural Economics*, *101*(4), 1164–1180.

Song, K. (2018). Measuring the graph concordance of locally dependent observations. *Review of Economics and Statistics*, *100*(3), 535–549.

Tosh, C., Greengard, P., Goodrich, B., Gelman, A., Vehtari, A., & Hsu, D. (2021). The piranha problem: Large effects swimming in a small pond. *arXiv*. https://arxiv.org/abs/2105.13445.

Yang, Y., McKhann, A., Chen, S., Harling, G., & Onnela, J.-P. (2019). Efficient vaccination strategies for epidemic control using network information. *Epidemics*, *27*, 115–122.