NAIVE LEARNING WITH UNINFORMED AGENTS

ABHIJIT BANERJEE[†], EMILY BREZA[§], ARUN G. CHANDRASEKHAR[‡], AND MARKUS MOBIUS[⋆]

Abstract. The DeGroot model has emerged as a credible alternative to the standard Bayesian model for studying learning on networks, offering a natural way to model naive learning in a complex setting. One unattractive aspect of this model is the assumption that the process starts with every node in the network having a signal. We study a natural extension of the DeGroot model that can deal with sparse initial signals. We show that an agent's social influence in this *generalized DeGroot model* is essentially proportional to the number of uninformed nodes who will hear about an event for the first time via this agent. This characterization result then allows us to relate network geometry to information aggregation. We show information aggregation preserves "wisdom" in the sense that initial signals are weighed approximately equally in a model of network formation that captures the sparsity, clustering, and small-worlds properties of real-world networks. We also identify an example of a network structure where essentially only the signal of a single agent is aggregated, which helps us pinpoint a condition on the network structure necessary for almost full aggregation. Simulating the modeled learning process on a set of real world networks, we find that there is on average 21.6% information loss in these networks. We also explore how correlation in the location of seeds can exacerbate aggregation failure. Simulations with real world network data show that with clustered seeding, information loss climbs to 35%.

Keywords: Social Networks, Social Learning, DeGroot Learning

Date: January 25, 2021.
We are grateful for financial support from NSF SES-1326661 and IRiSS at Stanford. We also thank Nageeb Ali, Gabriel Carroll, Drew Fudenberg, Ben Golub, Matt Jackson, Jacob Leshno, Adam Szeidl, Alireza Tahbaz-Salehi, Juuso Toikka, Alex Wolitzky, Muhamet Yildiz, Jeff Zwiebel and participants at the MSR Economics Workshop, the Harvard Information Transmission in Networks, Social Identity and Social Interactions in Economics (Université Laval) and seminar participants at MIT, Caltech, and Vienna for helpful discussions. Bobby Kleinberg provided the key ideas for the proof of Theorem 2.
[†]MIT, Department of Economics; BREAD, JPAL and NBER.
[§]Harvard, Department of Economics; BREAD, JPAL and NBER.
[†]Stanford, Department of Economics; BREAD, JPAL and NBER.
[⋆]Microsoft Research New England; University of Michigan; NBER.

> ...[A]s we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns – the ones we don't know we don't know... [I]t is the latter category that tend to be the difficult ones.
> *– Donald Henry Rumsfeld, US Secretary of Defense, (2002)*

1. Introduction

Learning from friends and neighbors is one of the most common ways in which new ideas and opinions about new products get disseminated. There are really two distinct pieces to most real world processes of social learning. One part of it is the exchange of views between two (or more) people who each have an opinion on the issue ("Lyft is better than Uber" or the other way around). The other piece is the spread of new information from an (at least partially) informed person to an uninformed person ("there is now an alternative to Uber called Lyft which is actually better"). *Information aggregation models* (Bala and Goyal, 1998; DeMarzo et al., 2003; Eyster and Rabin, 2014) emphasize the first while models of *diffusion* (Calvo-Armengol and Jackson, 2004; Jackson and Yariv, 2007; Banerjee et al., 2013) emphasize the second.

In reality both processes occur at the same time. For example, in the lead up to the financial crisis of 2007-2008, if popular accounts are to be believed, most investors were not tracking news on subprime lending, despite its central role in what ultimately happened. After all, *ex ante* there is a whole host of other factors that are potentially important to keep an eye on – this was also a period when world commodity prices were changing rapidly, and China seemed poised to take over the world economy. For most individual investors, information about the sheer volume and nature of subprime lending was new information, an *unknown unknown* when they heard about it from someone. After that of course many of them started tracking the state of the subprime market and started to form and share their own opinions about where it was going.

Microcredit programs provide another example. Most microfinance borrowers did not know that the product existed before a branch opened in their neighborhood.[1] Indeed we know from Banerjee et al. (2013) that the MFI studied in that paper has an explicit strategy of making its case to the opinion leaders in the village and then assuming that the information will flow from them to the rest of the village. However,

[1]Marketing materials of microfinance institutions (MFIs) often feature quotes from their beneficiaries to the effect that they never imagined that they could ever be clients of a formal financial institution.

once people hear about the product, they may seek out the opinions of others before deciding whether to take the plunge.

In this paper, we develop a generalization of the DeGroot model (DeGroot, 1974; DeMarzo et al., 2003) that accommodates both these aspects of social learning. We feel that this is important because the DeGroot model has a number of attractive properties that has made it perhaps the canonical model of boundedly rational information aggregation in network settings. First, the rule itself is simple and intuitive, whereas the correct Bayesian information aggregation rule in network settings can be so complex that it is hard to believe that anyone would actually use it. Indeed the experimental evidence supports the view that most people's information aggregation rules are better approximated by the DeGroot (or DeGroot-like) model than the Bayesian alternative (Corazzini et al., 2012; Mueller-Frank and Neri, 2013; Mengel and Grimm, 2015; Chandrasekhar et al., 2020).[2] Second the model has the attractive long-run property that under some relatively weak assumptions beliefs in a large population converge to the belief that would result from full aggregation of everyone's signal. This is the "wisdom of crowds" result (Golub and Jackson, 2010). Finally Molavi et al. (2017) provide axiomatic foundations of DeGroot social learning for a setting where agents have imperfect recall.

However, the DeGroot model makes the somewhat unrealistic assumption that everyone is informed about the issue at hand to start with (Golub and Jackson, 2010) – no one needs to be told that Lyft or microcredit or widespread subprime lending exists. The current paper relaxes that assumption and allows the initial signals to be sparse relative to the number of eventual participants in the information exchange. In other words, we allow for the possibility that many or even most network members may start by having absolutely no views on a particular issue, and only start having an opinion after someone else shares their opinion with them. While in the standard DeGroot model, agents average the opinions of their neighbors (including themselves) in every period, under our *Generalized DeGroot* (GDG) updating rule agents only average the opinion of their *informed* neighbors while ignoring uninformed neighbors (this make sense, for example, if those who have no signal are silent). Hence, an agent who has a seed signal and is surrounded by uninformed neighbors will stick to her initial opinion and only will start averaging once her neighbors become informed. An uninformed agent who has an informed neighbor will adopt that opinion. Our model

[2]In particular, (Corazzini et al., 2012) conduct a lab experiment which closely mimics the standard DeGroot setting–individuals pass on real-numbers (as needed)–and find that for the most part results are consistent with DeGroot learning.

reduces to the standard DeGroot model when all agents are initially informed and a standard diffusion model if informed agents all start with the same seed. Just like the standard DeGroot rule, the GDG rule can also be thought as a form of naive static Bayesian updating where uninformed agents have diffuse signals that are ignored in the aggregation while the stronger signals of informed agents are averaged. It turns out that the social learning dynamics under these assumptions can be thought of as the result of two separate processes: signals first *diffuse* through the social network such that uninformed direct and indirect neighbors of the initially informed agents adopt the opinion of the socially closest informed agent. But second, as soon as there are at least one pair of informed neighbors, they start exchanging opinions and engage in DeGroot averaging. This roughly corresponds to the two stages of social learning that we highlighted in our examples; what is an *unknown unknown* for some people at a point of time is a *known unknown* for others.[3] We show that what determines the long-run outcomes is the partition of the set of nodes into those that got their initial opinion from the same seed – the so-called *Voronoi tessellation* of the social network induced by the set of initially informed agents. The Voronoi tessellation therefore describes the seeded agents' social influence in our model, unlike in the standard DeGroot model where social influence is proportional to an agent's popularity (in the symmetric DeGroot version). Being popular isn't enough to be influential in our generalized model: agents might have to surround themselves with other popular but uninformed neighbors in order to enlarge their Voronoi set and make their opinion heard. Each element of this partition effectively plays the role of a single node in the standard DeGroot process; the (common) signal associated with all the nodes in that element gets averaged with the signals associated with the other elements of the partition over and over again, exactly as in the standard DeGroot model. The one difference is that the weight given to a particular signal is (essentially) the degree-weighted share of the nodes in the element of the partition associated with that signal. The topology of the social network embodied in the structure of the Voronoi partition therefore interacts with the ability of the DeGroot process to aggregate the signals of informed agents to generate the ultimate outcome.

An important consequence of this insight is that networks that would generate asymptotic full aggregation of all available signals in the standard DeGroot case (the "wisdom of crowds" effect analyzed by Golub and Jackson (2010)), may not do so

[3]Of course, in reality it is likely that both these processes of pure opinion aggregation intersects with another process of acquiring information by direct observation (for example by taking a Lyft), but this of course was also true in the original DeGroot model.

in the Generalized DeGroot case.[4] In other words, the long-run outcome may reflect only a fraction of the initially available signals. To demonstrate a worst-case version of this, we construct a class of networks which, for most initial sparse seed sets, "aggregates" only the signal of a single agent in the Generalized DeGroot case; this is what we call a *belief dictatorship*. With the same set of networks, there would be no dictatorships in the standard DeGroot case where all agents receive signals initially, since no agent in these networks has a particularly high degree.

However, we can also characterize large classes of networks that model real-world structure — the resultant graphs are sparse, clustered, and have small worlds (Watts and Strogatz, 1998; Newman, 2003; Jackson, 2008; Newman, 2010) —and demonstrate that they do not suffer from belief dictatorships. On the contrary, they aggregate signals almost perfectly.

The quality of the signal aggregation is therefore a function of the structure of the network. To get some empirical insight into whether the average real world network is closer to the belief dictatorship case or to the full aggregation case, we simulate the Generalized DeGroot process on a set of 75 village networks where we had previously collected complete network data (Banerjee et al., 2019) by injecting signals at a number of randomly chosen nodes. The variance of the long-run outcome of our simulated process across multiple rounds of injections gives us a measure of information loss. In our simulations we find that the average amount of information loss is 21.6%. We also find that there is substantial heterogeneity in how much information is lost/preserved with the 25th percentile losing about 33% of information and the 75th percentile losing 13% of information.

Our model also has direct implications for how network structure impacts the quality of aggregation. For a given realization of the location of the seeds, asymmetry in the degree-weighted Voronoi shares drives information loss in the model. Using the same data we show by simulations that the expectation of the Herfindahl index of the interior degree-weighted Voronoi shares is predictive of information loss across different networks. Moreover, we also show that other graph statistics relevant for the standard DeGroot model have no predictive power in the same simulations. This

[4]Importantly here we are not asking whether the Generalized DeGroot process leads to the same long-run outcome as the standard DeGroot process; because there are potentially many more initial signals in the standard DeGroot case, that would be an unfair comparison. The claim here is about the extent to which the long-run opinion reflects all the available signals, taking into account the fact that there are more signals in the standard DeGroot case.

suggests that the Voronoi Herfindahl index has relevant empirical content beyond other widely-used network measures, under the model.

In much of the paper, we analyze cases where the initial signals are distributed uniformly at random in the network. However, there are many real-world situations that might lead to information being clustered in a small number of sub-communities. We next show that for a class of networks, such clustered seeding can dramatically exacerbate information loss. Finally, we simulate the model using a clustered seeding protocol in the 75 Indian village networks and show that on average, correlation in the location of signals does indeed lead to a higher variance in limit beliefs, holding the number of signals fixed. We find that under clustered seeding, the average information loss climbs to 35%.

The remainder of the paper is organized as follows. Section 2 sets up the formal model. Section 3 shows how the limit belief can be thought of as a Voronoi-weighted average of the initial signals. In Section 4 describe how the network's geometry affects information loss. We explore how correlation in the location of initial signals can influence information loss in Section 5. Both in the theoretical illustration and in our data, such correlation exacerbates information loss. Section 6 concludes and introduces some questions for future research, inspired by our model.

2. A Model of DeGroot Learning with Uninformed Agents

2.1. **Setup.** Our model builds on the standard DeGroot model as introduced by DeMarzo et al. (2003) but adds uninformed agents. We consider a finite set of agents who each may observe a signal about the state of the world $\theta \in \mathbb{R}$.

There are a finite number n of agents who are embedded in a fully connected, undirected, unweighted graph g such that $(i,j) \in g$ implies $(j,i) \in g$ for any two agents i and j. Denote the set of agents/nodes in the network by N. All agents to whom a node is linked are called *neighbors*: this will be the group of people an agent listens to. We also assume that g includes self-loops (i,i) implying that an agent also listens to herself. We denote the degree of a node in the graph with d_i (including the self-loop). Our results also hold for undirected weighted graphs and we will indicate this when presenting our results.

At any point in time t an agent is either *informed* or *uninformed*. An informed agent at time t holds belief $x_i^t \in \mathbb{R}$. An uninformed agent holds the empty belief $x_i^t = \emptyset$. Following DeMarzo et al. (2003) we assume that the initial opinions of informed agents are an unbiased signal with finite variance about the true state drawn from the same

absolutely continuous distribution F (e.g., a normal distribution):

$$x_i^0 = \theta + \epsilon_i \quad \text{where} \quad \epsilon_i \sim F(0, \sigma^2). \tag{2.1}$$

At time $t = 0$ a set S of size $k = |S|$ nodes are initially seeded with signals x_i^0. The remaining $n - k$ nodes receive no signal at period 0. Note that if $k = n$, this is the standard DeGroot case, where signals are dense rather than sparse.

2.2. **Learning.** Agents observe their neighbors' opinions in every period and update their own beliefs. We denote the set of informed neighbors of agent i at time t with J_i^t and note that this set can include the agent herself. We then specify the *generalized DeGroot* (GDG) updating process as follows:[5]

$$x_i^{t+1} = \begin{cases} \emptyset & \text{if } J_i^t = \emptyset \\ \frac{\sum_{j \in J_i^t} x_j^t}{|J_i^t|} & \text{if } J_i^t \neq \emptyset. \end{cases}$$

Our updating rule implies that uninformed agents remain uninformed as long as all their neighbors are uninformed. If just one of her neighbors becomes informed, the uninformed agent will adopt the opinion of that neighbor. If there is disagreement the agent will use simple averaging of all its neighbors to derive a new opinion.[6] Note, that our updating rule reduces to the standard DeGroot model once every agent is informed. Section 6 spells out a potential foundation for this rule: it can be seen as a naive dynamic extension of the static optimum Bayesian learning rule.

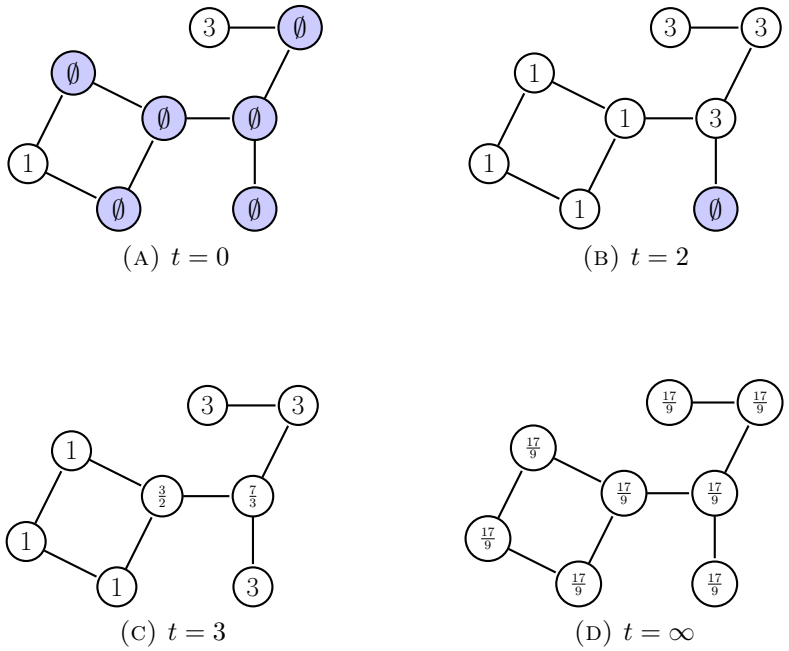
To gain intuition about the learning dynamics, consider the belief dynamic for the social network shown in Figure 1. At time $t = 0$ only two agents are informed and have distinct signals. During the next two periods the seeds' information *diffuses* and the direct neighbors and the neighbors of neighbors adopt the opinion of the seed closest to them. In period 3, an agent first encounters two distinct opinions and averaging starts and continues until all agents have converged to limit belief $\frac{17}{9}$. This example illustrates that the belief dynamics can be broadly described as a diffusion process followed by an averaging phase–while it is generally not possible to cleanly separate these two phases in time, they are helpful for characterizing the long-run behavior of our updating process. The diffusion part determines just how many nodes in the

[5] In the case of a weighted graph we have

$$x_i^{t+1} = \begin{cases} \emptyset & \text{if } J_i^t = \emptyset \\ \frac{\sum_{j \in J^t} x_j^t g_{ij}}{\sum_{j \in J^t} g_{ij}} & \text{if } J_i^t \neq \emptyset. \end{cases}$$

where $g_{ij} > 0$ are the edge weights. For the unweighted graph the weights are simply $g_{ij} = 1$.

[6] In the case of weighted graphs the agent will use a weighted average of her neighbors.



(A) $t = 0$ (B) $t = 2$

(C) $t = 3$ (D) $t = \infty$

FIGURE 1. Evolving beliefs in a sample social network

averaging process share the information from a single seed. In the (standard DeGroot) averaging process, which ensues after they have been informed, these shared signals are treated no differently from the case where they are the original signals these nodes started with, which, from the familiar logic of DeGroot aggregation, means that a seed that feeds more nodes to start with will tend to have greater influence on the long-run outcome.

3. How network geometry affects limit beliefs

3.1. **Different types of limit beliefs.** We next characterize limit beliefs in our model starting from the initial seed set $S = \{i_1, .., i_k\}$ of $k > 0$ informed agents. Note, that beliefs in our model always converge to some uniform limit belief x^∞ because all agents will become eventually informed and our model then reduces to the standard DeGroot model.[7]

Proposition 1. *The limit belief x^∞ is a weighted average $\sum_{i \in S} w_i(S) x_i^0$ of the initial signals of the seeds, where the weight given to the signal of seed i, $w_i(S)$, only depends on the position of the seeds in the network and $\sum w_i(S) = 1$.*

[7]This applies immediately by the same proof to the undirected, unweighted case since it only requires the irreducibility and aperiodicity of the stochastic weighting matrix.

The key intuition for this result is that we apply a linear operator to each agent's beliefs at each time. Proposition 1 also implies that the limit belief – for a fixed seed set S – is an unbiased estimator of the state of the world, where we take the expectation over the possible realizations of the initial signals.

We will call a seed's weight $w_i(S)$ in the limit opinion the seed's *social influence*. It will be convenient to assume from now on that the weights $w_i(S)$ are monotonic in the index i – this can always be accomplished by re-labeling the seeds, and therefore this assumption can be made without loss of generality.

Clearly the most efficient estimator attaches equal weight to each seed's signal since they are equally precise. We are particularly interested in the variability of the limiting social opinion x^∞:

$$\text{var}(x^\infty) = \sum_{i \in S} w_i(S)^2 \sigma^2. \tag{3.1}$$

We can bound this variance above and below:

$$\frac{\sigma^2}{k} \leq \text{var}(x^\infty) \leq \sigma^2. \tag{3.2}$$

Notice that the upper bound is the variance of a single signal, and this says that society effectively pays attention to one node's initial piece of information and has "forgotten" the $k-1$ other pieces of information. The lower bound is just the variance of the sample mean of k independent draws, which corresponds to the most efficient estimator.

Loosely speaking, we say that the generalized DeGroot process exhibits "wisdom" if the variance of the limit belief is close to the lower bound, which is precisely achieved by the optimal estimator. On the contrary, if the variance in the limit belief is close to the upper bound we say the process exhibits "dictatorship" because it only puts weight on the signal of one single agent.

3.2. **What affects the precision of long-term beliefs?** In order to understand the conditions under which wisdom or dictatorship arises we have to understand the weights $w_i(S)$. To study these weights we define the *Voronoi tessellation* of the social network induced by seed set S as a partition of the nodes of the social network into k almost disjoint sets. Each Voronoi set V_i is associated with a seed i and contains all the nodes that are *weakly* closer to seed i than any other seed in terms of network distance.

In the case of a weighted graph, we consider paths comprised of outward links and simply calculate network distance using whether a link exists or not (that is, $g_{ij} = 0$

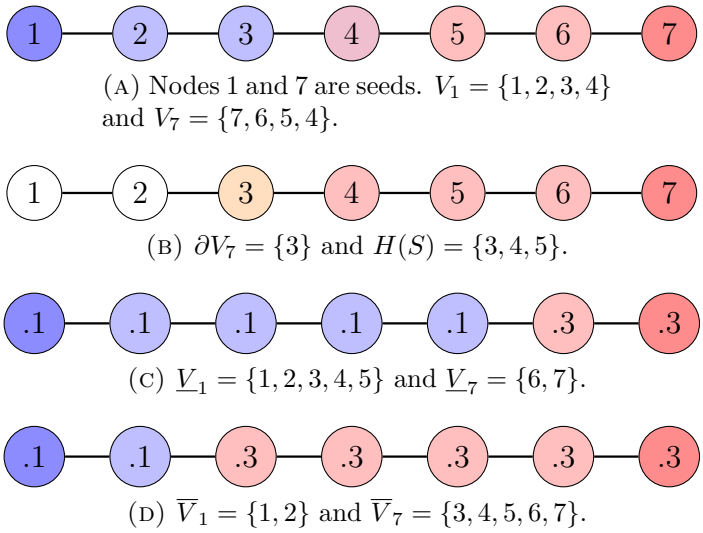
or $g_{ij} > 0$). In other words, the way we calculate network distance does not depend on the actual positive weight itself. That is to say, we define Voronoi sets as if there were no weighting on the link, but we calculate membership of j to the Voronoi set of seed i if its distance calculated by the minimal outward path from i to j is shorter than the minimal such paths to all other seeds.

These sets do not quite form a partition since nodes can be equidistant from two (or more) seeds in which case the nodes are assigned to multiple Voronoi sets. Panel A of Figure 2 provides an example of such a Voronoi tessalation on a line network with 7 agents where agents 1 and 7 are informed. Note, that agent 4 belongs to both Voronoi sets V_1 and V_7.

For each Voronoi set V_i define the boundary of the set to be ∂V_i which is the set of nodes that are not in V_i but are directly connected to an element of V_i (i.e., at distance 1). Panel B of Figure 2 illustrates this boundary for V_7. Next, for each node i' define the closest seed as $c_{i'}$ and the set of *associated seeds* $A(i', S)$ as those seeds whose shortest distance to i' differs from $c_{i'}$ by at most two. The set of associated seeds always includes at least the closest seed itself. We then define the *boundary region* $H(S)$ of seed set S as the set of nodes i' whose set of associated seeds has at least size 2. The boundary region includes equidistant nodes that are shared between two Voronoi sets but also nodes immediately next to the boundary between Voronoi sets. For each seed i we also define the *minimal* Voronoi set $V_i^{\min} = V_i \setminus H(S)$ and the *maximal* Voronoi set $V_i^{\max} = V_i \cup \partial V_i$.

Nodes within the minimal Voronoi sets will start averaging conflicting opinions only after all their neighbors have become informed. Intuitively, information aggregation will therefore occur exactly like in the standard DeGroot model. However, nodes in the boundary region $H(S)$ might enter the averaging phase while their set of informed neighbors is still evolving: based on the rules of GDG updating their initial opinion (once becoming informed) can therefore vary between the lowest and highest signal among their associated seeds. In order to bound these two extremes, we construct the *lower* Voronoi sets $\underline{V}_i$ and the *upper* Voronoi sets $\overline{V}_i$ for a particular signal realization x_i^0 on the seeds as follows.

Let us start with the lower Voronoi sets $\underline{V}_i$ first. All nodes in the minimal set V_i^{min} are assigned to $\underline{V}_i$. Moreover, any node $i' \in H(S)$ is assigned to the associated seed with the *lowest* signal realization. Note that the lower Voronoi sets form a partition. We define the upper Voronoi sets $\overline{V}_i$ analogously by assigning nodes in the boundary region to the *highest* associated seed. We can bound the sets in this lower and upper



(A) Nodes 1 and 7 are seeds. $V_1 = \{1, 2, 3, 4\}$ and $V_7 = \{7, 6, 5, 4\}$.

(B) $\partial V_7 = \{3\}$ and $H(S) = \{3, 4, 5\}$.

(C) $\underline{V}_1 = \{1, 2, 3, 4, 5\}$ and $\underline{V}_7 = \{6, 7\}$.

(D) $\overline{V}_1 = \{1, 2\}$ and $\overline{V}_7 = \{3, 4, 5, 6, 7\}$.

FIGURE 2. Nodes 1 and 7 are seeds, with signals 0.1 and 0.3, respectively. The panels describe the Voronoi sets as well as the upper and lower Voronoi sets.

partition as follows:

$$\begin{aligned} V_i^{\min} &\subseteq \underline{V}_i \subseteq V_i^{\max} \\ V_i^{\min} &\subseteq \overline{V}_i \subseteq V_i^{\max} \end{aligned} \tag{3.3}$$

Panels C and D in Figure 2 illustrate this construction of lower and upper Voronoi sets.

We can now bound the limit belief x^∞ only based on the lower and upper Voronoi partition. To state the result we denote the share of nodes in a network that is part of the lower Voronoi set $\underline{V}_i$ with $\underline{v}_i = \frac{|\underline{V}_i|}{n}$ and define the link-weighted share:

$$\underline{v}_i^* = \frac{\sum_{i \in \underline{V}_i} d_i}{\sum_{i=1}^n d_i}. \tag{3.4}$$

Analogously, we define the link-weighted share of agents in the upper Voronoi set $\overline{V}_i$. Note, that for regular graphs such as the circle we have $\underline{v}_i = \underline{v}_i^*$.

Our first theorem (proved in the Appendix) then says:

Theorem 1. *Assume a social network with seed set S. The limit belief is bounded below and above as follows:*[8]

$$\sum_i \underline{v}_i^* x_i^0 \leq x^\infty = \sum_{i \in S} w_i(S) x_i^0 \leq \sum_i \overline{v}_i^* x_i^0. \tag{3.5}$$

The proof of Theorem 1 proceeds by induction: we show that we can sandwich the link-weighted opinion of all *informed* agents in each time period $t = 0, 1, ..$ by the link-weighted average seed opinions that are assigned to these agents by the respective lower and upper Voronoi partition. This is easy to show at time $t = 0$. The inductive argument exploits the fact that the standard DeGroot averaging rule preserves the link-weighted average opinion of agents between time t and $t + 1$ (proved in the Appendix). However, for agents in the boundary region the set of informed neighbors with conflicting opinion tends to increase: the lower and upper Voronoi sets provide the appropriate bounds to bound the evolution of these agents' beliefs until all their neighbors are informed.

Theorem 1 allows us to characterize the limit belief by studying a static problem and relates the *geometry* of the social network to an agent's social influence $w_i(S)$. We define the link-weighted share of the minimal Voronoi set as

$$v_i^{*,min} = \frac{\sum_{i \in V_i^{min}} d_i}{\sum_{i=1}^n d_i}. \tag{3.6}$$

and, analogously, we define $v_i^{*,max}$ as the link-weighted share of the maximal Voronoi set.

Corollary 1. *The social influence $w_i(S)$ of seed i satifies*

$$v_i^{*,\min} \leq w_j(S) \leq v_i^{*,\max}.$$

The proof of Corollary 1 follows immediately from inequality (3.3) and Theorem 1 by setting all signals except for seed i to 0.

Intuitively, an agent's social influence is (approximately) proportional to the size of her Voronoi set which determines how many agents she manages to convince of her opinion before information aggregation commences.

It is instructive to note when the result of Theorem 1 collapses to an equality. We can clearly see this happens in the case of thin boundaries. Recall that the boundary region $H(S)$ is comprised of the set of nodes with at least two associated seeds. This includes both nodes that are equidistant from multiple seeds but also

[8]These bounds are tight if we focus on general networks. For specific classes of geometries we can improve the bounds.

nodes immediately next to these boundaries. Observe that if the boundary region is thin—$|H(S)|$ is small when compared to the total number of links per seed in the network—then Theorem 1 can be treated as an equality.

To get an intuition as to when this could happen, consider any sort of lattice such as a line, a planar lattice, or a lattice in three dimensions. In each of these cases, the volume of any ball about a seed is an order of magnitude larger than its perimeter (which corresponds to the dimension of the possible boundary). So one can easily calculate that the boundary will be thin and therefore ignorable in the limit. Thus, to the extent that real world networks look lattice-like or exhibit neighborhoods around any seed nodes that have small perimeters relative to their volumes, we can treat the limit opinion as behaving as in Theorem 1 but with an equality. This may be the case if real-world networks are well-approximated by geographic networks (Ambrus et al., 2014) or latent-space networks (Hoff et al., 2002; Penrose, 2003; McCormick and Zheng, 2015; Chandrasekhar et al., 2020).

4. How network geometry affects wisdom

In this section we explore how the geometry of the network influences how much information gets aggregated into the final opinion. This is particularly important in the sparse case because, even with large n the actual number of signals, k, can be a small number and therefore we cannot assume that there is a law of large numbers in operation.

We begin by studying a class of random network models that are designed to model the features of real-world social networks. In the models, nodes reside in a latent space, in our case a hypercube, and then closer nodes are more likely to be linked as described below. The resulting networks are sparse, clustered, and exhibit small worlds (Penrose, 2003; Watts and Strogatz, 1998; Newman, 2003; Jackson, 2008). The models we study, where proximity modulates linking propensity, are both intuitive and related to those often used in statistics, sociology, and econometrics (Hoff et al., 2002; Hoff, 2008; Graham, 2017; McCormick and Zheng, 2015; Smith et al., 2019; Breza et al., 2020). We demonstrate that the Voronoi sets in these models tend to be of roughly equal size and, therefore, wisdom is preserved.

We then study a stylized example to contrast the logic of our setting with sparse signals with the widely-studied setting where everyone gets a signal. The goal of this example is to demonstrate that the failure to learn with sparse signals does not have to come from the presence of hyper-connected individuals, unlike in the case when

everyone gets a signal (which we call the dense case).[9] Failure can happen even when all the nodes have a small number of links but the network exhibits specific kinds of asymmetries that make seeds at certain location become much more influential than others.

Before we begin our study, it is instructive to start by comparing the case of sparse signals to the dense case. Golub and Jackson (2010) characterize when crowds will be wise in the dense case and show that, for a setting like ours, the degree distribution is a sufficient statistic for characterizing asymptotic learning. Other network statistics, such as average path length are irrelevant. Formally, Golub and Jackson (2010) show that a sequence of graphs $(g_j)_{j\in\mathbb{N}}$ is wise only if

$$\max_{1\leq i\leq n_j} \frac{d_i(g_j)}{\sum_{i'=1}^{n_j} d_{i'}(g_n)} \to 0.$$

where $d_i(g_j)$ denotes the degree of node i in graph g_j and the size of graph g_j is equal to n_j.[10]

In the sparse case, however, the above condition no longer guarantees wisdom.

In fact, we can construct a sequence of networks, all satisfying the Golub and Jackson (2010) condition, where one of the k signals comes to fully dominate everybody's opinion. That is, the society's converged opinion reflects just one signal and therefore is arbitrarily close to having the maximal possible variance. More generally, this suggests that networks with important asymmetries may destroy a considerable part of the available information in sparse learning environments.

From an empirical perspective, our analysis points to contexts in which networks exhibit such asymmetries are vulnerable to failures of wisdom

4.1. **Wisdom in "realistic" models of social networks.** In this section we study classes of networks which capture many of the standard features of real-world social networks and show that for a typical seeding, the Voronoi sets of the seeds are essentially all of the same order of magnitude. In this case, the wisdom of crowds result continues to hold in the sense that the final opinion almost always reflects equally weighted information from all k seeds.

For this exercise, we first look at random geometric graphs (RGG), which are simple spatial networks where nodes are randomly distributed over an m-dimensional metric space (Penrose, 2003). For example, for $m = 1$ such a network looks similar to a

[9] Golub and Jackson (2010) highlight the importance of hyper-connected individuals as a source of the failure of information aggregation in the dense case.

[10] Recall, that our definition of degree includes a self-loop.

line (but allows for local rewiring) and for $m = 2$ it resembles a plane. These graphs resemble social networks in the sense that they are typically characterized by sparsity, community structure and clustering which are fundamental properties of real-world social networks.

After showing that RGGs exhibit wisdom, we consider the possibility of adding a small number of "shortcuts" to these graphs to capture the idea that while most of the links are between nearby nodes there is some chance that distant nodes link as well. RGGs with shortcuts preserve the fundamental "small-world" property of small average path length between any two nodes (see Watts and Strogatz (1998)) and describe real world networks better. We show that making this change induces only small changes in the variance of limit beliefs and therefore preserves wisdom.

4.1.1. *Random Geometric Graphs.* We consider an m-dimensional hyper-cube $[0, L)^m$ where L is a positive integer. The volume of this hypercube is therefore L^m. For example, on a line this hyper-cube is simply the interval $[0, L]$ and on a plane it is a square with side length L. We then draw n nodes on this this cube and define $\delta = \frac{n}{L^m}$ as the *node density*. Any two nodes are said to be neighbors if their Euclidean distance is at most 1.

In a standard RGG these nodes are randomly placed on the hyper-cube with uniform probability. This poses a technical problem because any instance of such a RGG is not guaranteed to be connected: the node density has to be above a threshold to make sure that two random nodes belong to the same connected component with high probability even if the RGG is very large. While the exact thresholds are not known they can be bounded: for example, for the plane the node density has to be at least 1.44.[11] In order to not have to deal with partially connected graphs and multiple components we assume that nodes inside the hyper-graph whose coordinates are integers always belong to the RGG (we call these *base nodes*). Effectively, this implies that the density of the network always satisfies $\delta \geq 1$ because base nodes alone contribute 1 unit of density. The remaining density $(\delta - 1)L^m$ nodes are distributed uniformly and

[11]The threshold is called the *percolation threshold*. If the node density is smaller then the probability that any two random nodes are connected through the same *giant component* converges to 0 as L grows. Otherwise, it remains bounded away from 0. For the plane the threshold node density on the plane lies in the interval $[1.43, 1.44]$ (Balister et al., 2005): in other words, every unit cube has to contain at least 1.44 nodes on average for a giant component to emerge.

independently over the hyper-cube just like in the standard model (Penrose, 2003). Thanks to these base nodes all our RGGs are fully connected.[12]

A second technical problem arises due to the fact that the maximum degree of our RGG increases linearly with n because there are network realizations where all non-base nodes are selected within a small ball of radius 1. These events are rare: the expected number of neighbors for any node as well the standard deviation in the degree is proportional to the node density (for example the expected number of neighbors is $\delta\pi$ on the plane). Random graphs with unbounded degree nevertheless pose a technical difficulty for us because the social influence of a seed is proportional to the degree-weighted size of the associated Voronoi set. To simplify our analysis we impose a maximum degree d^* by *pruning* any regions of an instance of a random graph with excess degree as follows: (i) we take the union of all edges of agents with degree exceeding d^* except edges involving base nodes; (ii) we then randomly delete one of these excess edges; (iii) we repeat this algorithm until the degree of every node is at most d^*. Since we do not delete base edges the graph remains connected. Moreover, by choosing the maximum degree d^* large enough we can minimize the need for pruning.

We define the resulting class of random geometric graphs as $RGG(L|m, \delta, d^*)$. Such a network consists of $n = \delta L^m$ nodes. For this class of RGGs the following "approximate wisdom" result holds.

Theorem 2. *Consider random geometric graphs in the class $RGG(L|m, \delta, d^*)$ with $n = \delta L^m$ nodes and assume that $k \leq n$ seeds are randomly chosen on the network. Then there is a constant $C(m, d^*)$, independent of n, such that we can bound the expected variance in the limit opinion*

$$\mathrm{E}_{S,RGG(L|m,\delta,d^*)}\left[\mathrm{var}(x^\infty)\right] \leq \frac{C\sigma^2}{k}, \tag{4.1}$$

where on the left-hand side we take the expectation over all seed sets S of size k and all realizations of RGGs in the class $RGG(L|m, \delta, d^*)$.

To understand the significance of this result recall the basic inequality (3.2) that bounds the variance of the limit belief:

$$\frac{\sigma^2}{k} \leq \mathrm{var}(x^\infty) \leq \sigma^2.$$

[12]This excludes node densities below 1 – however, these densities are below the percolation threshold of standard RGGs and hence it would not make sense to study them for the purpose of this paper where we focus on connected graphs where all agents' beliefs converge.

The theorem shows that on average the variance in the limit belief is at most a constant factor larger than the first-best case where all signals are equally weighted. In particular, the variance of the limit belief scales inversely proportional with k just like the optimal bound. We can therefore view Theorem 2 as an approximate "wisdom of crowds" result similar to Golub and Jackson (2010) for this class of networks.

4.1.2. *Adding Shortcuts to RGGs.* In addition to being sparse and clustered, real-world networks have low average path length. Watts and Strogatz (1998) note that this cannot be generated from spatial graphs by local rewiring only. Instead, we have to allow for limited *long-range* rewiring that creates shortcuts in the social network. Indeed, typical empirical models of network structure in the statistics, sociology, and econometrics literatures allow for such shortcuts.

Formally, we define a rewiring of the class $RGG(L|m,\delta,d^*)$ of RGG graphs that we introduced in the previous section: we randomly pair all agents in the giant component with a random partner and with probability η we add a link between these two nodes for each of the $n/2$ pairs.[13] By construction, the degree of every node increases by η in expectation. We denote the resulting class of "small-world" RGG networks as $RGG(L|m,\delta,d^*,\eta)$.

Theorem 3. *Consider the class $RGG(L|m,\delta,d^*,\eta)$ of "small-world" RGG networks with $n=\delta L^m$ nodes and assume that $k=o(\sqrt{n})$ seeds are randomly chosen on the network. Then there is a constant $C(m,d^*,\eta)$, independent of n, such that we can bound the expected variance in the limit opinion*

$$\mathrm{E}_{S,RGG(L|m,\delta,d^*,\eta)}\left[\mathrm{var}(x^\infty)\right] \leq \frac{C\sigma^2}{k}, \tag{4.2}$$

where on the left-hand side we take the expectation over all seed sets S of size k and all realizations of RGGs in the class $RGG(L|m,\delta,d^*,\eta)$.

This result implies that typical small world RGG exhibits wisdom for sparse seed sets. The intuition behind this result is as follows. In "small-world" RGG networks a agent's neighborhood at distance r increases at an exponential rather than at a polynomial rate as in standard RGGs. This has a dramatic on average path length (as shown in Watts and Strogatz (1998)). However, it does not affect the relative ability of every seed to diffuse information. Therefore, it does not exacerbate the imbalance of the Voronoi set size distribution.

[13] If n is odd we leave out one randomly selected agent.

4.1.3. *Sparse Erdős-Renyi Graphs.* The same intuition holds for the class of random Erdős-Renyi graphs $R(n|d^*, p)$ with n where every possible pairwise edge exists with probability p such that the expected number of edges $np > 0$. This ensures that the graph has a giant connected component. As for RGGs we again prune edges for nodes whose degree exceeds d^*.

Proposition 2. *Consider the class $R(n|d^*, p)$ of random Erdős-Renyi graphs with $np > 1$ and assume that $k = o(\sqrt{n})$ seeds are randomly chosen on the network. Then there is a constant $C(p)$, independent of n, such that we can bound the expected variance in the limit opinion*

$$\mathrm{E}_{S,R(n|d^*,p)}\left[\mathrm{var}(x^\infty)\right] \leq \frac{C\sigma^2}{k}, \tag{4.3}$$

where on the left-hand side we take the expectation over all seed sets S of size k and all realizations of random graphs in the class $R(n|d^*, p)$.

4.2. **Belief Dictators.** To highlight the fact that wisdom is not guaranteed, we construct a class of networks such that the generalized DeGroot process selects an opinion dictator with probability close to 1 in the sparse case despite being wise in the dense case. In particular, we highlight a case where despite the fact that no node has high degree, we typically end up with a single Voronoi set that envelops the entire network.

For each integer r we define a graph $G_T(r)$ that – intuitively – consists of a central tree graph surrounded by a "wheel". We construct the tree by starting with a root agent who is connected to 3 neighbors. Each of these neighbors in turn is connected to 2 neighbors, and we let this tree grow outward up to radius r. We can calculate the number of agents in this tree network as:

$$n = 1 + 3 + 3 \times 2 + 3 \times 2^2 + ... + 3 \times 2^{r-1} = 1 + 3\left(2^r - 1\right). \tag{4.4}$$

Agents at the perimeter of this tree have 3×2^r unassigned links. We surround the tree by a circle of size 3^{r+1} and connect the tree's unassigned links like spokes on a wheel to this circle such that spokes connect to an equidistant set of nodes on the circle. All agents in this network have degree 2 or 3: agents who are connected to any agent in the central tree have degree 3 and all other agents have degree 2.

Proposition 3. *Consider the class $G_T(r)$ of social networks and assume that k seeds are randomly chosen on the network with $k = o\left((c \cdot 3/2)^r\right)$ for $0 < c < 1$. The expected*

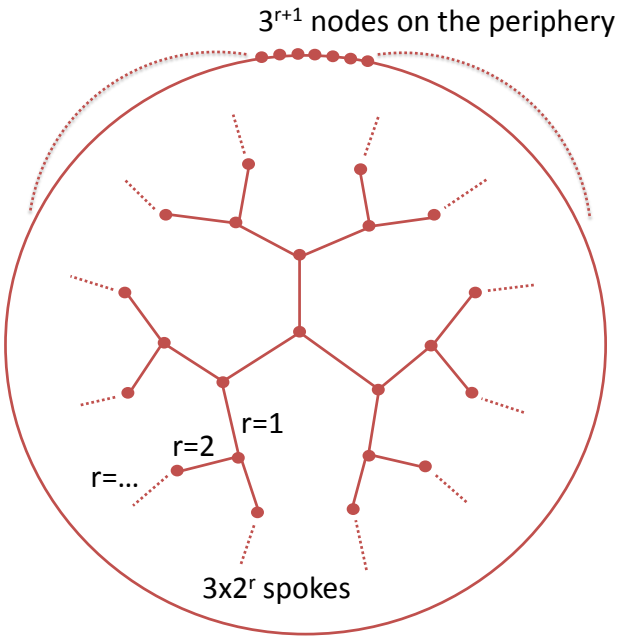


FIGURE 3. Belief Dictators Example

value of the largest weight $\mathrm{E}_S(w_k(S))$ *(taken over all the seeds sets) converges to* 1 *as* $r \to \infty$ *while the expectation of all lower-ranked weights converges to* 0.

In other words, one of the seeds becomes, with high probability, a belief dictator. The intuition is simple: the share of agents in the center is $o(\left(\frac{2}{3}\right)^r)$ and therefore converges to 0 as r increases. Hence, for large enough r it becomes highly unlikely that any of the seeds are located in the center. Now consider the seed that happens to be closest to a spoke. It is easy to see that this distance is uniformly distributed since seeds are drawn randomly. Moreover, the distance between two spokes on the wheel is $O(\left(\frac{3}{2}\right)^r)$. Therefore, the distance between the closest and second closest seed increases exponentially in r. However, the closest seed needs only $O(r)$ time periods to infect the central tree and spread out to the all the other spokes. Hence, the opinion of the closest seed will take over almost the entire network. Put differently, the Voronoi set of the closest seed encompasses almost the entire network.

Further, observe that the result holds even if the number of seeds grows in n provided that the growth is not too fast (this encompasses, for example, polylogarithmic growth rates and any sub-exponential growth rate relative to the size of the network itself.

It is instructive to contrast this observation to the "wisdom of crowds" result in Golub and Jackson (2010). Each $G_T(r)$ network has bounded degree and therefore, based on the result in Golub and Jackson (2010), aggregates final opinions (almost) perfectly efficiently in the standard (dense) DeGroot model. However, since our process adds a diffusion stage to the social learning process, properties of the social network – such as expansiveness, which is the number of links outgoing from a given set of nodes relative to the number of links among member of that set –, also matter for learning.

4.2.1. *Discussion: Uniform at Random Seeding.* It is important to remember that our prior analysis all assumes *uniform at random* seeding. By studying both the RGG with shortcuts model and the tree and wheel graph under uniform at random seeding, we observe stark results.

In the model of more realistic networks, RGG with shortcuts, the following happens. In the typical random seeding, because of the relative homogeneity of the network, the partition into Voronoi sets is somewhat balanced across the graph on the hypercube. As a consequence, wisdom prevails. Of course, this does depend on the distribution of seeding: one can consider an adversarial seeding process which precisely eliminates some of the seeds; we explore this in the case of clustered seeding below.

Turning to our belief dictatorship example, the tree and wheel graph has the property that for a typical seed configuration drawn under random seeding, one seed is uniquely situated to spread information, through the key central apparatus, to almost the entire network. That means that typically one Voronoi set will be of outsized importance due to this asymmetry.

We consider departures from uniform at random seeding, specifically the case of clustered seeding, in Section 5, below.

4.3. **Simulations in Indian Village Networks.** We have explored network geometries in cases where there is wisdom (i.e., all k units of information are preserved) as well as cases where belief dictatorships arise (i.e., where $k-1$ units of information are destroyed). However, whether GDG dynamics in real-world networks tend more toward belief dictatorship or wisdom is ultimately an empirical question. To investigate this, we simulate our model using network data collected from 75 independent villages in India and analyze the resulting variance of each community's beliefs across simulation draws. Next, following the logic suggested by our theoretical analysis, we

TABLE 1. Summary Statistics

	(1) Mean	(2) Standard Deviation
Village Size	216.37	70.65
Fraction in Giant Component	0.96	0.02
Average Degree	10.18	2.50
Variance of the Degree Distribution	33.41	20.17
Average Clustering Coefficient	0.26	0.05
Average Path Length	2.81	0.35
Village Diameter (Longest Shortest Path)	5.93	1.07
First Eigenvalue	13.79	3.47

empirically explore the extent to which asymmetries in the expected size in Voronoi sets predict heterogeneity in information loss.

4.3.1. *Data Description.* For this exercise, we use the household network data collected by Banerjee et al. (2019). The data set captures twelve dimensions of interactions between almost all households in 75 villages located in the Indian state of Karnataka. Surveys were completed with household heads in 89.14% of the 16,476 households across these villages. Thus the data respresents a near-complete snapshot of each village's network.

For simplicity, in this analysis we assume two households to be linked if in the surveys, either household indicated that they exchange information or advice with the other.[14] Thus, our resulting empirical networks are undirected,[15] which means that we effectively have link data on 98.8% of pairs of nodes.[16] For this exercise, we further restrict our analysis to only the giant connected component of each graph.

Table 1 contains descriptive statistics across all 75 of the empirical networks. The average village in the sample contains approximately 216 households, 96% of whom are typically contained in the village's giant component. Restricting only to those nodes in the giant component, the average degree in the sample is 10.18, but exhibits a large amount of dispersion with an average variance of 33.41. Average path lengths

[14]Specifically, the questions ask about which households come to the respondent seeking medical advice or help in making decisions. Symmetrically, the questions also ask to whom the respondent goes for medical advice or for help in making decisions.

[15]See Banerjee et al. (2013) and Banerjee et al. (2019) for a detailed description of the data collection methodology and for a general discussion of the data.

[16]This follows from $1 - (1 - 0.8914)^2 = 0.988$.

in these networks are quite short, with a minimum distance of 2.81 between two arbitrarily-chosen households in the sample. Moreover, the average diameter (i.e., the longest shortest path) of the 75 villages in the sample is 5.93. We also observe that the average clustering coefficient in 0.26, which implies that any pair of common links for a household are themselves linked with 26% probability.

4.3.2. *Signal Structure.* For our simulations, we take the world to be $\theta = \frac{1}{2}$. Further, we assume signals to be distributed $\mathcal{N}(\theta, \sigma^2)$ with $\sigma^2 = 1$. We conduct simulations for varying levels of sparsity: $k \in \{2, 4, 6, 8, 10, 14, 18, 22, 26, 30\}$. For each village, for each k and for each simulation run, we randomly seed k out of the n total nodes with a signal and calculate the limit opinion under GDG. We simulate the model 50 times for each village, for each k.

We are interested in measuring the variance of these limit opinions in the simulations, which we denote as $\sigma^2_{x^\infty}$. We can then compare this variance to the natural benchmark that would arise if each individual could observe all k signals simultaneously. In that case, the limit belief would simply be the sample mean over the realizations of each of the k signals. This sample mean has variance $\frac{\sigma^2}{k} = \frac{1}{k}$.

Given that some network geometries destroy information (belief dictatorships), while others preserve all k signals, we use the simulation exercise to quantify how much information is destroyed in the village networks. To do this, we define the effective number of signals as

$$k^{effective} := \frac{\sigma^2}{\sigma^2_{x^\infty}}.$$

Given that $\sigma^2_{x^\infty} \leq \frac{\sigma^2}{k}$, $k^{effective}$ (which must be less than or equal to k) measures the number of signals that would generate a variance equivalent to $\sigma^2_{x^\infty}$ if all of those signals could be observed simultaneously by an individual. The extent of information preservation is given by $\frac{k^{effective}}{k}$.

4.3.3. *Results.* Figure 4, we plot the the mean $k^{effective}$ against the true k, averaging across all 75 villages. We do find some evidence of information loss across the different values of k; note that each point falls below the 45-degree line. On average, adding one additional signal improves $k^{effective}$ by 0.775 signals.

In addition, we find substantial heterogeneity in the degree of information loss across the 75 networks. We plot the interquartile range of average village outcomes for each k. That is, we calculate the 25th percentile and the 75th percentile in the distribution of $k^{effective}$ across the 75 villages. We find that on average, the 25th

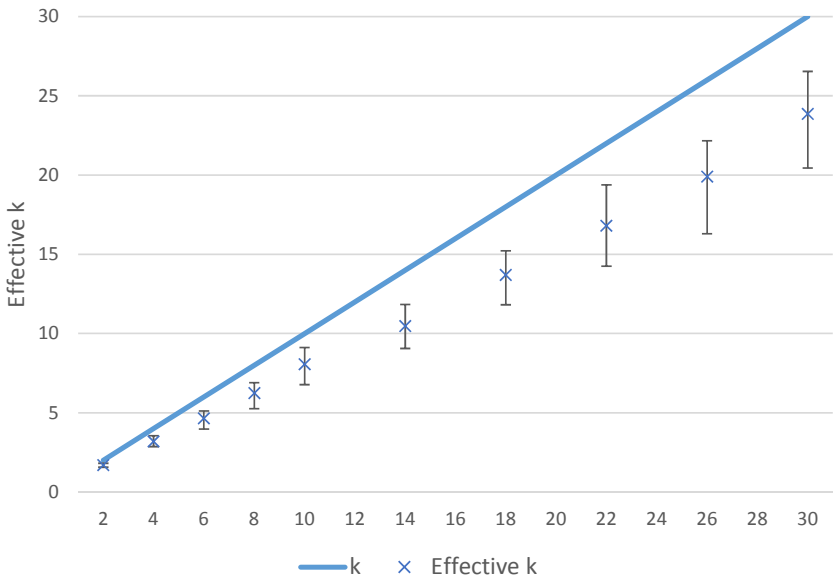
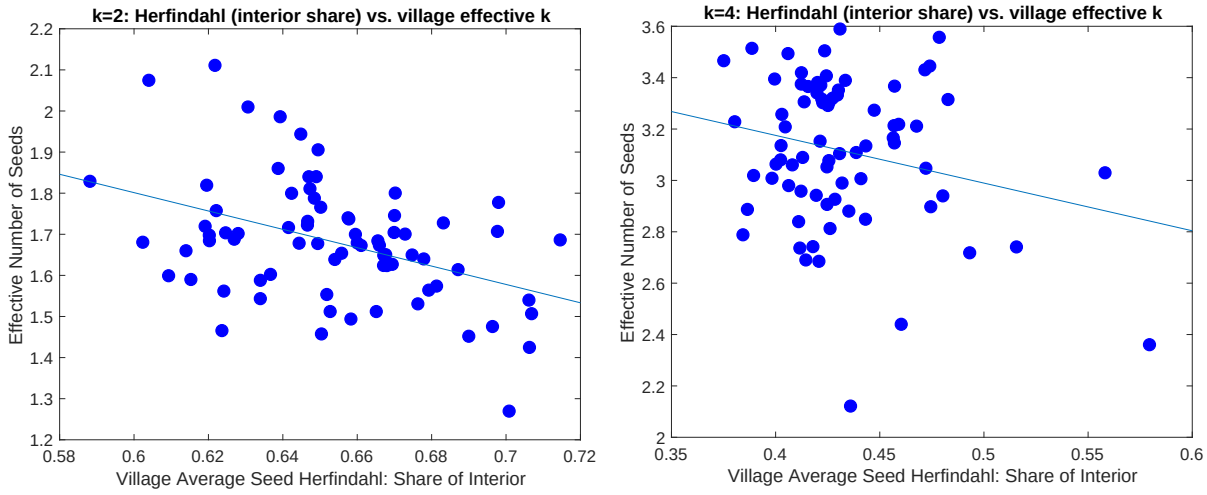


FIGURE 4. Simulations on 75 Indian village networks. The average is taken over all simulations and all networks. The bars represent the interquartile range across networks, for each k.

percentile village experiences 33% information loss, while the 75th percentile village experiences 13% information loss.

While the village networks are inconsistent with a stark belief dictatorship, it is difficult to assess whether the 22.5% average information loss is consistent with wisdom given the asymptotic nature of our theoretical results.[17] However, we can use the model to better understand the heterogeneity across networks and to better diagnose which network stuctures produce more information loss. Clearly under our model, the degree-weighted shares of the interior Voronoi cells and the signal variance are key determinants of the limiting variance. In our empirical simulations we therefore explore the extent to which the variance in the size of the interior Voronoi cells drives information loss in practice. To operationalize this, we calculate the degree-weighted share of nodes in the interior Voronoi cell associated with seed node i, $\tilde{w}_i(S) = \frac{\sum_{j \in V_i^{min}} d_j}{\sum_{i=1}^{k} \sum_{j \in V_i^{min}} d_j}$. In our simulations, for each k, in each village, we can calculate a Herfindahl index for these $\tilde{w}_i(S)$ shares which captures the extent of inhomogeneity of influence across seeds.

[17]After all, the wisdom results impose bounds on the limit variance of the form $\frac{C\sigma^2}{k}$. In finite networks, the constants may be important.



(A) $k = 2$ (B) $k = 4$

FIGURE 5. Interior Voronoi Share Herfindahl and Village Effective k. Simulations on 75 Indian village networks. The solid line represents the OLS regression line. The slope of the regression line can be rejected at the 1% level in panel (A) and the 10% level in panel (B).

Figure 5 plots Effective k on the y-axis, against this Herfinadahl index on the x-axis, averaging across all random signal placement realizations within a village. Panel a) considers the case when $k = 2$. A value of 0.5 indicates that averaging across all 50 simulated signal placements, the two Voronoi cells are of equal size, while values approaching 1 indicate that one Voronoi cell completely dominates the other in degree-weighted share. Panel b) considers the case when $k = 4$. Here, a Herfindahl index of 0.25 would be consistent with equal shares. For both $k = 2$ and $k = 4$, we indeed observe that information loss is greater when the degree-weighted Voronoi shares are more asymmetric in size. The results show that in a finite and real-world network, the core intuition from our result carries through–the imbalance in cell sizes greatly affects the effective amount of information, highlighting the empirical content of our model and a direction for further empirical exploration.

Next, we ask whether other network-level statistics are predictive of information loss or if information loss is only well-captured by our Voronoi Herfindahl. We consider two network statistics related to the standard DeGroot model. To capture the potential for outsized influence by some nodes, for each network we calculate the ratio of the 80th percentile to the 20th percentile of the degree distribution. For completeness, we

TABLE 2. Relationship Between Information Loss and Graph-Level Statistics

Information Loss	(1) Voronoi Herfindahl	(2) 80/20 Degree Ratio	(3) Second Eigenvalue
Graph-Level Statistic	-2.234	-0.051	-0.317
	(0.630)	(0.059)	(0.300)
	[0.000]	[0.382]	[0.290]
Number of Villages (Networks)	75	75	75

Notes: In each column, average information loss from the simulations by village is the dependent variable. Column names denote the graph-level statistic regressor. Robust standard errors in parentheses. P-values in brackets.

also include the second eigenvalue of the stochastic adjacency matrix.[18] This value is related to the speed of convergence in the DeGroot model.[19]

Table 2 shows the strength of the relationship between each network-level measure and information loss in our simulations for $k = 2$. Each column reports the OLS regression coefficient for one of the three statistics. We observe that the relationship between the interior Voronoi Herfindal index (also graphed in Figure 5, Panel A) is statistically significant at the 1% level. However, we cannot reject that the relationship is 0 for the other measures.

Moreover, the results have implications for the placement of seeds. To the extent possible, selecting seeds to minimize asymmetries in the Voronoi shares should limit information loss.[20] A related targeting question is explored below.

4.3.4. *Discussion.* In sum, that real world social networks do quite well at preserving information both in theory (as our "small worlds" results show) and in practice (using Indian village data) have 22.5% information loss in our simulations. An interesting avenue to explore in future research is to look at which sorts of economic environments give rise to equilibrium networks that are more likely to generate wisdom.

[18]Specifically, we divide each entry of the adjacency matrix with binary entries by the row sum and then calculate the second largest eigenvalue.

[19]Because our model discusses limit beliefs, which is reflected in our choice of $T = 20$ for the simulations, this measure is likely less relevant than the 80/20 degree ratio.

[20]We note that selecting the seeds to exactly minimize the Voronoi Herfindahl is an NP-hard problem, computationally. However, an approximation may be quite useful.

5. Clustered Seeding

We have thus far focused our analysis on situations where the set of initially-informed agents is drawn uniformly at random from the population. However, in many real-world settings, opinion-leaders tend to be clustered in a small number of locations. Firms often offer promotions to those they perceive as opinion leaders (e.g., on Twitter) and agricultural extension workers target new technology to those who they perceive to be "model farmers". And these targeted people often tend to be clustered just because the same kind of people tend to be connected to each other. Here, we explore the consequences of clustered seeding on the variance of the limit beliefs using an illustrative example.

5.1. **An Illustrative Example.** We consider a circle network of n nodes in which each individual has two friends, $d_i = 2$ for each i, one friend to the right and one friend to the left. Assume that there are R intervals or "regions" that collectively contain all of the k opinion leaders in the network. These regions are distributed randomly over the circle and together comprise a small number of nodes relative to n. In other words, if the rth interval has b_r nodes, $b = \sum_{r \in R} b_r \ll n$. To capture the idea that opinion leaders are often the first to learn about new technologies or opportunities, we assume that seeds are drawn randomly from these b nodes only. Note, that we abstract away from any difference in network structure within and outside regions (the network structure is the same).

Given this structure, the variance of the limit beliefs is constrained by the number of regions and not just the number of seeds. If there are few regions then the limit opinion is less predictable even if there are many seeds.

To see this, we begin with a simplification of the above setup. Assume that the R regions each have $\frac{b}{R}$ nodes and the regions are equally spaced in the network of size n. Let z_n denote the distance between two adjacent regions in the circle measured by the closest members of each region, recognizing that this distance is the same for any pair of adjacent regions. Finally let $k = b$, so every node in every region receives an initial signal x_i^0.

With this setup, notice that the Voronoi set for every seed node is either 1 (for all interior nodes within each region) or $\frac{z_n}{2}$ for each boundary node, of which there are $2R$. Since the number of regions R and the number of nodes per region b/R is held constant as $n \to \infty$, it follows from the arguments of Proposition 1 and Theorem 1,

$$\text{var}(x^\infty) = \frac{\sigma^2}{2R} + o(1) \text{ as } n \to \infty.$$

The logic behind this is that $k - 2R$ of the seed nodes all have a Voronoi set of size 1 which vanishes relative to the $2R$ seed nodes that have growing Voronoi sets, all of equal size, $\frac{z_n}{2}$.

This means that even though there are $k > 2R$ nodes that serve as initial seeds, because they are divided into R regions, the boundaries of these regions drive the limit opinion. Therefore the limit opinion under a clustered allocation of seeds mighthave far more variance than under a more dispersed allocation.

Returning to the more general setup, where now b can differ from k, the regions can be distributed randomly over the network, and seeds are drawn randomly from the $b \geq k \geq R$ nodes. In this case, the reader can check, there are constants C_1 and C_2 that do not depend on R, k, n, or b such that we can bound the variance in the limit opinion as follows:

$$\mathrm{E}_S[\mathrm{var}(x^\infty)] \leq \left(\frac{C_1 \frac{b}{n}}{k} + \frac{C_2(1 - \frac{b}{n})}{\min(R, k)} \right) \sigma^2.$$

Note that the variance of the limiting belief is bounded above by a function that is decreasing in the number of regions R. The intuition here is that the Voronoi sets are determined by the seeds that are closest to the boundary of each region. Any seeds that are sandwiched between boundary seeds essentially won't matter because the combined size of their Voronoi sets is bounded by the sum of the R intervals, which is small by assumption. Thus clustered seeding can result in much more information loss than random seeding.

Figure 6 presents a simple illustration of the phenomenon. Here we have a circle with $n = 300$, $k = 8$, $R = 2$, and $b_r = 6$ for each region. The large balls indicate the initial seeds. The darkly shaded nodes indicate members of the two regions.

The example plots the limit beliefs following a specific realization of the eight signals. The large, solid balls indicate a signal realization of 1, while the empty balls indicate a signal realization of 0. Note that in both examples, the average signal is 0.375. However, the signal configurations and signal realizations have been chosen to show how the interior signals are basically ignored. In the left panel, both the left-most and right-most signal in each region are essentially preserved, resulting in a limit belief close to 0.75. In the right panel the regions are close together, and therefore the signals that are closest to each other in the different regions also do not influence the limit opinion by much. In this case, the limit belief, 0.97, is close to 1, which makes sense because it largely reflects only the outer two signals, that is the two signals with realization 1.

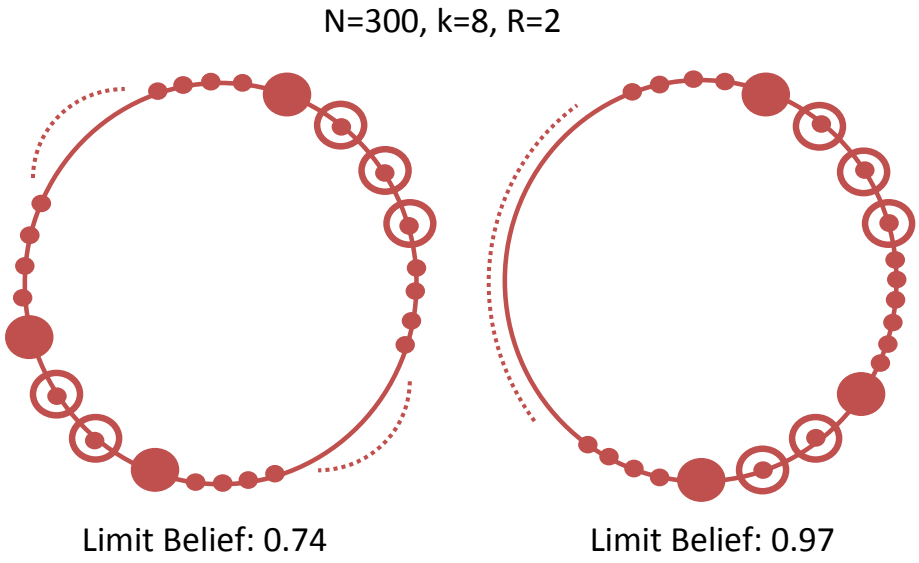


FIGURE 6. Example: Clustered Seeding

5.2. **Simulations in Indian Village Networks.** We now repeat the exercise of Section 4.3, where we simulate the GDG process on or Indian village network data. In all of our simulations we fix $k = 20$ and we vary the number of regions seeds can come from, from three to ten. These regions are located randomly throughout the network.

Figure 7 shows the results, repeating 50 simulations per network for each of the 75 networks.

We show that the effective number of signals ranges from 13 to 15, depending on the number of regions, which range from three to ten here. If there are only three regions, for instance, this corresponds to a loss of 35% of the information. When we compare this to the case where signals are distributed i.i.d., in Figure 7, we see that this represents a 12.5pp further decline in effective number of signals on top of the 22.5pp loss just due to the GDG process. This shows that in empirical networks, when information is not distributed uniformly at random, the loss can be sizable.

These results again highlight a role for thoughtful seed selection. Clustered seeding creates excess asymmetry in the Voronoi set size distribution and drives further information loss. However, if seed locations could be selected to instead reduce such asymmetries, even more information could be preserved, relative to the case of random seed selection.

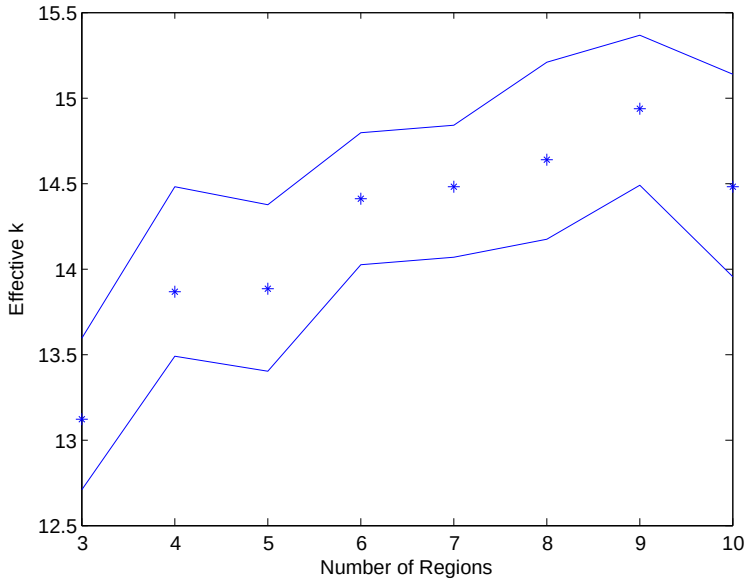


FIGURE 7. Plots of the mean $k^{effective}$ against R, where the average is taken over all simulations and all networks. The solid lines represent the 5th and 95th percentiles of $k^{effective}$, bootstrapped across the simulation draws.

These implications are related, more broadly, to the literature on how to select seeds when starting a social learning process.[21] This literature has emphasized seeding central nodes, for example. But this could lead to clustered seeding because central nodes, by construction, are homophilistic. Similarly, agricultural extension agents are likely to select model farmers with similar characteristics. Moreover, there is an operating intuition that is used in a variety of contexts– from model farmers to influencers–that if information is provided in a clustered setting, then it mixes early, and a collection of influential individuals may all obtain the right beliefs–before it diffuses. In our model, while such mixing between these central or clustered nodes may happen, at the same time, there is the simultaneous diffusion process which begins from each of these seeds. However, the diffusion originating from each cluster is asymmetric–some signals start to spread to previously-uninformed nodes right away, while some mix and become trapped. Rather than focusing on getting groups of nodes

[21]See, for example, Katz and Lazarsfeld (1955); Kempe et al. (2003, 2005); Aral and Walker (2012); Banerjee et al. (2013); Beaman et al. (2018); Banerjee et al. (2019). Also related is the literature on model farmers in agricultural extension, such as Emerick and Dar (2020); BenYishay and Mobarak (2019).

to buy into the right opinion by giving them a collection of signals, a policy-maker could potentially reduce information loss by spreading the signals apart.

6. Discussion and Conclusions

There is a continuum of possible naive learning rules – for example, one can think of rules that aggregate signals in some non-linear way or that incorporate the presence of uninformed neighbors (other than ignoring them as GDG does). In this concluding section, we argue that GDG has a number of desirable properties which make it a focal choice for naive learning in the presence of uninformed agents.

6.1. **GDG as One-Step Bayesian Updating.** In the standard DeGroot model with Gaussian signals the linear learning rule is the optimal Bayesian rule in period $t = 1$ which the agent then "naively" applies in all subsequent periods when it is no longer optimal (DeMarzo et al., 2003). The following argument shows that the GDG rule is the obvious analogue to that rule in the presence of uninformed agents.

To see this, assume that the signals are drawn normally for informed agents: $F(\theta, \sigma^2) = \mathcal{N}(\theta, \sigma^2)$. In order to perform Bayesian learning with uninformed agents we assume that an uninformed agent i has also a normally distributed but highly imprecise signal $\tilde{x}_i$:

$$\tilde{x}_i^0 = \theta + \tilde{\epsilon}_i \quad \text{where} \quad \tilde{\epsilon}_i \sim \mathcal{N}(0, \tilde{\sigma}^2) \tag{6.1}$$

We assume that the variance $\tilde{\sigma}^2$ is very large and we will implicitly consider the limit case as $\tilde{\sigma}^2 \to \infty$.

It is now easy to see that a Bayesian learner who has at least one informed neighbor would exactly apply GDG as $\tilde{\sigma}^2 \to \infty$. Moreover, a Bayesian learner who has no informed neighbors (including herself) would arrive at a low-precision posterior which we can interpret as "staying uninformed". Hence, the GDG model can be interpreted as the naive application of one-step Bayesian updating in every period: in both the original and our generalized DeGroot model agents behave like "naive Bayesians".[22]

[22]Note that our results on belief dictatorships do not discontinuously rely on ignoring the uninformed. To see this formally, let $h_I = \frac{1}{\sigma^2}$ and $h_U = \frac{1}{\tilde{\sigma}^2}$ be the respective precisions, which will be used in the weighting formula. Agents average over their informed and uninformed neighbors, weighting by precisions. It is easy to check that the belief dictatorship in $G_T(r)$ described in Section 4.2 persists if h_I is sufficiently large relative to h_U. Formally, allowing the ratio h_I/h_U to depend on r, the result follows if $(3/2)^r = o(h_I/h_U)$ as $r \to \infty$. However, our modeling choice in having the informed not weigh the uninformed is *intentional*: those who have nothing to say about a topic do not contribute to the conversation and are purely consumers of the newly discovered information.

6.2. **GDG and the Loss of Precision.** While, as we show above, there are cases where the generalized DeGroot model allows society to learn the average of all the seeds, it is worth commenting that they do not learn the number of seeds k that make up this average. In other words, they don't learn the precision of what they have learned. In the standard DeGroot model there is no need to learn k because everyone starts informed and therefore if the population is large, the long-run outcome of the DeGroot process is almost always the exact truth – precision of the prediction is not an issue. In contrast, under GDG only a relatively small number of signals get aggregated even for large networks, at least in the interesting case. In such an environment even after many rounds of aggregation, participants in the learning process would want to know if the opinion aggregated 3 or 30 signals.

One way to modify the GDG process to solve this *precision problem* is to require everyone to keep track of the uninformed agents they encounter. For example, agents could keep track of two different numbers: (1) the share of informed agents (with the initial opinion equal to the share of informed neighbors at time $t = 0$) and (2) the average opinion of informed agents (as in GDG). Agents then use the standard DeGroot rule for updating their estimate of the share of informed agents in the population and GDG for learning the average signals of informed agents. By learning the share of informed agents, the naive learner can infer k (assuming she knows n) while, as shown above, GDG allows her to learn the average of these k seed agents in many classes of social networks.

However, to learn k, decision-makers need to keep track of the share of all the uninformed agents they encounter from *the beginning of time in all states of the world*. This may be a plausible assumption when the state of the world is a *known unknown*: for example, agents might have no information about the state of the economy right now but they are probably interested in this outcome from the beginning and know that some people have received signals. Hence, they might keep track of the share of informed agents even before any signal reaches them. However when dealing with *unknown unknowns* (such as a new product or an unanticipated state of the world) it seems implausible that agents will start updating their information before they have talked to at least one informed neighbor.

It turns out however that there are ways to solve the problem of estimating precision without using uninformed agents: for example, agents could "tag" informed seeds and transmit these tags to their neighbors. What this means is that an agent could explicitly tell her neighbors the actual names of the seeds that she knows of, and her

neighbors can do the same, thereby keeping track of exactly which individuals were original seeds (as well as possibly their seed values). If k is not too large then tagging is an excellent way to easily learn k. However, tagging quickly becomes cognitively expensive for larger k.

The examples suggest that learning precision (e.g., k) might be difficult. At the same time, our results in this paper show that learning the average is inexpensive and can be achieved through GDG in many settings. We hope to address the topic of precision in social learning in future work.

6.3. **Concluding remarks.** The DeGroot model is fast becoming a work-horse model for learning on social networks. We relax one key and potentially unrealistic assumption of the model and show that this can completely undermine the full information aggregation result associated with the standard DeGroot model. However, we also characterize a large class of networks where this does not happen. Our simulations using 75 real world social networks from Indian villages suggest that the outcome corresponds to 22.5% information loss on average. Moreover, the heterogeneity in average information loss across villages is related to network structure through asymmetries in the degree-weighted Voronoi shares, and not through conventional measures motivated by the analysis of the standard DeGroot model with dense seeding. Finally, we observe that the extent of information aggregation depends on the clustering of signals on the network. Under clustered seeding, the average information loss is 35% in our simulations using real world social networks.

References

Ambrus, A., M. Mobius, and A. Szeidl (2014): "Consumption Risk-Sharing in Social Networks," *American Economic Review*, 104, 149–82. 3.2

Aral, S. and D. Walker (2012): "Creating social contagion through viral product design: A randomized trial of peer influence in networks," *Management Science*. 21

Bala, V. and S. Goyal (1998): "Learning from neighbours," *Review of Economic Studies*, 65, 595–621. 1

Balister, P., B. Boolobas, and M. Walters (2005): "Continuum percolation with steps in the square of the disc," *Random Structures and Algorithms*, 26, 392–403. 11

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013): "Diffusion of Microfinance," *Science*, 341, DOI: 10.1126/science.1236498, July 26 2013. 1, 15, 21

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2019): "Using gossips to spread information: Theory and evidence from two randomized controlled trials," *The Review of Economic Studies*, 86, 2453–2490. 1, 4.3.1, 15, 21

Beaman, L., A. BenYishay, J. Magruder, and A. M. Mobarak (2018): "Can network theory-based targeting increase technology adoption?" Tech. rep., National Bureau of Economic Research. 21

BenYishay, A. and A. M. Mobarak (2019): "Social learning and incentives for experimentation and communication," *The Review of Economic Studies*, 86, 976–1009. 21

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2020): "Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data," *American Economic Review*, 110, 2454–84. 4

Calvo-Armengol, A. and M. Jackson (2004): "The effects of social networks on employment and inequality," *The American Economic Review*, 94, 426–454. 1

Chandrasekhar, A. G., H. Larreguy, and J. P. Xandri (2020): "Testing models of social learning on networks: Evidence from two experiments," *Econometrica*, 88, 1–32. 1, 3.2

Corazzini, L., F. Pavesi, B. Petrovich, and L. Stanca (2012): "Influential listeners: An experiment on persuasion bias in social networks," *European Economic Review*, 56, 1276–1288. 1, 2

DeGroot, M. (1974): "Reaching a consensus," *Journal of the American Statistical Association*, 69, 118–121. 1

DeMarzo, P., D. Vayanos, and J. Zwiebel (2003): "Persuasion Bias, Social Influence, and Unidimensional Opinions*," *Quarterly journal of economics*, 118, 909–968. 1, 2.1, 6.1

Emerick, K. and M. H. Dar (2020): "Farmer field days and demonstrator selection for increasing technology adoption," *Review of Economics and Statistics*, 1–41. 21

Eyster, E. and M. Rabin (2014): "Extensive Imitation is Irrational and Harmful," *The Quarterly Journal of Economics*, 129, 1861–1898. 1

Golub, B. and M. Jackson (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, 112–149. 1, 4, 9, 4.1.1, 4.2

Graham, B. S. (2017): "An econometric model of network formation with degree heterogeneity," *Econometrica*, 85, 1033–1063. 4

Hoff, P. (2008): "Modeling homophily and stochastic equivalence in symmetric relational data," in *Advances in Neural Information Processing Systems*, 657–664. 4

Hoff, P., A. Raftery, and M. Handcock (2002): "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97:460, 1090–1098. 3.2, 4

Jackson, M. (2008): *Social and Economic Networks*, Princeton: Princeton University Press. 1, 4

Jackson, M. and L. Yariv (2007): "Diffusion of Behavior and Equilibrium Properties in Network Games," *American Economic Review*, 97, 92–98. 1

Katz, E. and P. Lazarsfeld (1955): *Personal influence: The part played by people in the flow of mass communication*, Free Press, Glencoe, IL. 21

Kempe, D., J. Kleinberg, and E. Tardos (2003): "Maximizing the Spread of Influence through a Social Network," *Proc. 9th Intl. Conf. on Knowledge Discovery and Data Mining*, 137 – 146. 21

——— (2005): "Influential Nodes in a Diffusion Model for Social Networks," *In Proc. 32nd Intl. Colloq. on Automata, Languages and Programming*, 1127 – 1138. 21

McCormick, T. H. and T. Zheng (2015): "Latent surface models for networks using Aggregated Relational Data," *Journal of the American Statistical Association*, 110, 1684–1695. 3.2, 4

Mengel, F. and V. Grimm (2015): "An Experiment on Learning in a Multiple Games Environment," . 1

Molavi, P., A. Tahbaz-Salehi, and A. Jadbabaie (2017): "Foundations of Non-Bayesian Social Learning," *Working Paper*. 1

Mueller-Frank, M. and C. Neri (2013): "Social Learning in Networks: Theory and Experiments," . 1

Newman, M. (2010): *Networks: An Introduction*, Oxford University Press. 1

Newman, M. E. (2003): "The structure and function of complex networks," *SIAM review*, 45, 167–256. 1, 4

Penrose, M. (2003): *Random Geometric Graphs*, Oxford University Press. 3.2, 4, 4.1, 4.1.1

Rosenblat, T. S. and M. M. Mobius (2004): "Getting Closer or Drifting Apart?*," *The Quarterly Journal of Economics*, 119, 971–1009. A.4

Smith, A. L., D. M. Asta, C. A. Calder, et al. (2019): "The geometry of continuous latent space models for network data," *Statistical Science*, 34, 428–453. 4

Watts, D. and S. Strogatz (1998): "Collective dynamics of small-world networks," *Nature*, 393, 440–442. 1, 4, 4.1, 4.1.2, 4.1.2

Appendix A. Proofs

A.1. **Proof of Proposition 1.** The limit x^∞ exists since once all agents are informed, standard (dense) DeGroot commences and we have assumed g is such that the corresponding stochastic matrix is irreducible and aperiodic.

Consider $t^*(S)$ as the period where the last uninformed agent becomes informed. Because the generalized DeGroot learning process is a composition of linear operators, it must be the case that $x_i^{t^*}$ for every i is a linear combination of x_j^0 for $j \in S$. And beginning at $t^*(S)$, we can treat the process as standard DeGroot since everyone has a signal, so the limit is just a weighted average of the initial signals, and we denote the weights $w_j(S)$ for $j \in S$. Note that this does not depend on whether g is weighted as long as the corresponding stochastic matrix is irreducible and aperiodic the argument follows.

A.2. **Proof of Theorem 1.** We will make use of a simple auxiliary lemma that characterizes the evolution of beliefs under the standard DeGroot model. To gain some intuition consider a graph where every agent has opinion 0 except agent i who has opinion 1. Denote the set of neighbors of i with $N(i)$ and assume that every agent j has degree d_j where we use the convention that the degree is equal to $|N(j)|+1$. More generally in the case of weighted graph, d_j will equal the sum of weights including the self-loop weight as well. Denote the opinion of each agent j at time t in the network with $x_j^{t,i}$.

It is easy to see that the opinion of agent i at time $t = 1$ will equal $x_i^{1,i} = \frac{1}{d_i}$ and the opinion of neighbor $j \in N(i)$ at time t with $x_j^{1,i} = \frac{1}{d_j}$. Note that we have:

$$\sum_j d_j x_j^{0,i} = \sum_j d_j x_j^{1,i}. \tag{A.1}$$

In this example both sides of this equation are equal to d_i.

We can show that this holds more generally, at every t and for arbitrary initial signal vector x^0.

Lemma 1. *In the standard DeGroot model with undirected links the link-weighted sum of beliefs is preserved:*

$$\sum_j d_j x_j^{t-1} = \sum_j d_j x_j^t. \tag{A.2}$$

Proof of Lemma 1. Denote the (column) vector of opinions at time $t+1$ with $x^{t+1} = \left(x_i^{t+1}\right)$ and the vector of opinions at time t with x^t. Also introduce the degree (row)

vector $D = (d_i)$. Finally, denote the DeGroot transition matrix with M. We then have:

$$x^{t+1} = Mx^t \tag{A.3}$$

Now left-multiply both sides with the row vector D:

$$Dx^{t+1} = D \cdot Mx^t \tag{A.4}$$

It is easy to see that $D \cdot M = D$. This proves the lemma. □

Note that Lemma 1 implies that $\sum_j d_j x_j^0 = x^\infty \sum_j d_j$ for limit belief x^∞ which delivers the well-known limit belief of the DeGoot model with symmetric links.

We next prove Theorem 1.

We assume that the process starts from a seed set S and initial opinions x_i for $i \in S$. We also denote the opinion of each agent at time t in the network with $\tilde{x}_i^t$ such that $\tilde{x}_i^0 = x_i$ for all $i \in S$ and $\tilde{x}_i^0 = \emptyset$ otherwise.

We denote the set of agents who become newly informed at time $t = 0, 1, 2, ..$ with ∂S^t and the agents who are already informed with S^t. Hence the total set of informed agents after time t is $S^t \cup \partial S^t$. We use the convention $S^0 = \emptyset$ and $\partial S^0 = S$ (initial seed set). Note that eventually every agent becomes informed such that $\partial S^t = \emptyset$ for $t \geq T$ and some T that depends on the graph and the seed set.

We denote the opinion of agent i in the lower Voronoi configuration with $\underline{x}_i$ and in the upper Voronoi configuration with $\overline{x}_i$. These opinions are defined for all agents in the network and are equal to the opinion of the closest seed (except in case of ties when the lower and upper configuration differ).

We want to prove the following claim:

Claim 1. *The following inequality holds for all times:*

$$\sum_{j \in S^t} d_j \underline{x}_j \leq \sum_{j \in S^t} d_j x_j^t \leq \sum_{j \in S^t} d_j \overline{x}_j$$

Note, that this claim implies as $t \to \infty$

$$\sum_{j=1}^n d_j \underline{x}_j \leq \sum_{j=1}^n d_j x^\infty \leq \sum_{j=1}^n d_j \overline{x}_j$$

which proves Theorem 1.

We prove the claim by induction on $t = 0, 1, ...$ At time $t = 0$ the claim is trivially true because S^0 is an empty sets. Now assume that the claim holds at time t. We

show that this implies that the claim holds for $t+1$ as well (which completes the inductive argument).

We can think of the evolution of beliefs from time t to $t+1$ as the result of two processes: (a) for all agents in the set $S^t \cup \partial S^t$ the process evolves like a standard DeGroot process on the truncated network that only includes edges of the graph where both nodes are in $S^t \cup \partial S^t$; (b) agents in the set ∂S^{t+1} become informed.

Let's look at the DeGroot process on the truncated network first. We can use Lemma 1 to show

$$\sum_{j \in S^t \cup \partial S^t} \hat{d}_j x_j^t = \sum_{j \in S^t \cup \partial S^t} \hat{d}_j x_j^{t+1} \tag{A.5}$$

where $\hat{d}_j$ is the degree of agent j in the truncated network at time t that only involves agents in the set $S^t \cup \partial S^t$. Next, we note that $\hat{d}_j = d_j$ for all $j \in S_t$ and $\hat{d}_j \leq d_j$ for $j \in \partial S^t$. Since we also have $S^{t+1} = S^t \cup \partial S^t$, we can rewrite equation (A.5) as follows:

$$\sum_{j \in S^t} d_j x_j^t + \sum_{j \in \partial S^t} \left[\hat{d}_j x_j^t + (d_j - \hat{d}_j) x_j^{t+1} \right] = \sum_{j \in S^{t+1}} d_j x_j^{t+1} \tag{A.6}$$

Now we use the definition the upper and lower Voronoi sets to derive the following inequalities for $j \in \partial S^t$:

$$\begin{aligned} \underline{x}_j &\leq x_j^t \leq \overline{x}_j \\ \underline{x}_j &\leq x_j^{t+1} \leq \overline{x}_j \end{aligned} \tag{A.7}$$

Both follow because j lies either on the "fat" boundary between Voronoi sets or completely inside a Voronoi set. In the latter case both x_j^t and x_j^{t+1} equal the value of the closest seed and the inequalities are trivially true. Otherwise, by the way we define ∂S^t, the only seeds that can possibly affect the opinion of j at times t and $t+1$ are the ones that determines $\underline{x}_j$ and $\underline{x}_j$. Since the opinion of j is always a convex linear combination of these seeds the inequalities have to hold.

From A.7 and $\hat{d}_j \leq d_j$ we get:

$$\begin{aligned} \hat{d}_j \underline{x}_j + (d_j - \hat{d}_j)\underline{x}_j &\leq \hat{d}_j x_j^t + (d_j - \hat{d}_j) x_j^{t+1} \leq \hat{d}_j \overline{x}_j + (d_j - \hat{d}_j)\overline{x}_j \\ d_j \underline{x}_j &\leq \hat{d}_j x_j^t + (d_j - \hat{d}_j) x_j^{t+1} \leq d_j \overline{x}_j \end{aligned} \tag{A.8}$$

We sum these inequalities over $j \in \partial S^t$ and add $\sum_{j \in S^t} d_j x_j^t$:

$$\sum_{j \in S^t} d_j x_j^t + \sum_{j \in \partial S^t} d_j \underline{x}_j \leq \sum_{j \in S^t} d_j x_j^t + \sum_{j \in \partial S^t} \left[\hat{d}_j x_j^t + (d_j - \hat{d}_j) x_j^{t+1} \right] \leq \sum_{j \in S^t} d_j x_j^t + \sum_{j \in \partial S^t} d_j \overline{x}_j$$

Now we use A.6 and obtain:

$$\sum_{j\in S^t} d_j x_j^t + \sum_{j\in \partial S^t} d_j \underline{x}_j \leq \sum_{j\in S^{t+1}} d_j x_j^{t+1} \leq \sum_{j\in S^t} d_j x_j^t + \sum_{j\in \partial S^t} d_j \overline{x}_j \tag{A.9}$$

Since the induce claim holds at time t we get:

$$\sum_{j\in S^{t+1}} d_j \underline{x}_j \leq \sum_{j\in S^{t+1}} d_j x_j^{t+1} \leq \sum_{j\in S^{t+1}} d_j \overline{x}_j \tag{A.10}$$

This completes the inductive argument and hence the proof of Theorem 1.

A.3. **Proof of Theorem 2.** Our proof proceeds in two steps. Step 1 shows that we can reduce the problem of bounding the variance of the limit opinion by bounding the probability that two randomly drawn nodes are in the same minimal Voronoi set. Step 2 then uses a geometric argument to find such a bound.

Step 1: For a particular realization G of a random graph in class $RGG(L|m, \delta, d^*)$ fix the seed set S. We know from Corollary 1 that the social influence of a seed i is bounded above by $v_i^{*,max}$, the link-weighted share of its maximal Voronoi set. Therefore, we can bound the variance of the limit belief as follows:

$$\text{var}(x^\infty) \leq (d^*)^2 \sigma^2 \sum_{i\in S} (v_i^{max})^2 \tag{A.11}$$

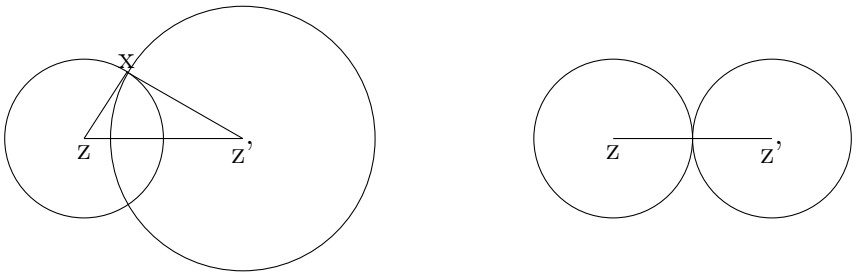
Here we use the simple size share v_i^{max} of i's maximal Voronoi set and note that the link-weighted share can be at most $(d^*)^2$ times as large.

Next, note that the maximal Voronoi differs from the minimal Voronoi set only by the boundary $H(S)$ which can be at most a 3-node thick layer. Hence, the we know that $v_i^{max} <= (1 + d^* + (d^*)^2 + (d^*)^3) v_i^{min}$. Therefore, the problem reduces to finding an upper bound on $\sum_{i\in S} (v_i^{min})^2$. This switched from maximal to minimal Voronoi sets is to ensure that there is no overlap between the Voronoi sets.

Now consider taking two random points on the graph G. The probability that two of them lie in the same minimal Voronoi set for seed i is $(v_i^{min})^2$. We define the random indicator variable $I_{zz'}$ which equals 1 iff z and z' are in the same minimal Voronoi set. Hence, $\text{E}(I_{zz'}) = \sum_{i\in S} (v_i^{min})^2$ and our problem reduced to finding an upper bound on $\text{E}(I_{zz'})$, the probability that two random nodes on G lie within the same Voronoi set.

Step 2: We now bound $\text{E}(I_{zz'})$ as follows. We can fix a point z on the hyper-cube and then draw a random z' at Euclidean distance r from z. When will z and z' be in the same Voronoi set? To gain some intuition we look at Figure 8 which shows z and z' and the associated seed x on the left panel. It has to be the case that the two

FIGURE 8. Bounding $E_S(I_{zz'})$



Left Panel Right Panel

discs do not contain any more seeds. It is easy to see that the area of these disks is always at least as large as the two tangential, equal-sized discs on the right panel with radius $r/2$ each. We have to be careful because network distance in our graph is not necessarily the Euclidean distance: however, thanks to the presence of base nodes the network distance between two point is at most the L_1 or *Manhattan* distance, which is at most $\sqrt{2}$ larger than the Euclidean distance. We therefore consider disks around z and z' with Euclidean radius $\frac{r}{2}\sqrt{2}$: the probability that z and z' are part of the same Voronoi set is bounded above by the probability that these two disks do not contain any seeds.[23]

The probability that a seed is contained in this disk equals $2V(\frac{r}{2\sqrt{2}})/L^m$ where $V(r)$ is the Euclidean volume of the disk (since we are dealing with hypercubes a disk is m-dimensional). The probability that no seed lies inside this disk equals:

$$\left(1 - \frac{2V(\frac{r}{2\sqrt{2}})}{L^m}\right)^k \tag{A.12}$$

Next, note that z' is randomly drawn. Consider the number of nodes that are in a ring of width dr around the surface of a disk with radius r centered at z. The surface area equals $A(r)$ and hence the number of nodes in this area equals $\delta A(r)dr$ (where δ is the node density). We can therefore draw z' randomly from these thin rings and thus draw from all potential z' in the hypercube (effectively, we are using polar coordinates). The probability that any one of the n randomly drawn points lies in the same Voronoi set as z can therefore be expressed as an integral:

$$\mathrm{E}_S(I_{zz'}) \leq \int_0^L \left(1 - \frac{2V(\frac{r}{2\sqrt{2}})}{L^m}\right)^k \frac{\delta A(r)}{n} dr \tag{A.13}$$

[23]We are grateful to Bobby Kleinberg for this insight.

We know that $V(r) = \alpha r^m$ and $A(r) = \beta r^{m-1}$ for some α and β (for example, for the plane we have $\alpha = \pi$ and $\beta = 2\pi$). Hence we can write:

$$\mathrm{E}_S\left(I_{zz'}\right) \leq \int_0^L \left(1 - \frac{2\alpha}{2^{\frac{3m}{2}} L^m} r^m\right)^k \frac{\delta\beta r^{m-1}}{n} dr \tag{A.14}$$

We use a change in variable with $x = r^m$ and $dx = mr^{m-1}dr$ and get:

$$\begin{aligned} \mathrm{E}_S\left(I_{zz'}\right) &\leq \int_0^{L^m} \left(1 - \frac{2\alpha}{2^{\frac{3m}{2}} L^m} x\right)^k \frac{\delta\beta}{mn} dx \\ &= \frac{1}{k+1}\left[1 - \left(1 - \frac{2\alpha}{2^{\frac{3m}{2}}}\right)^{k+1}\right] \frac{2^{\frac{3m}{2}-1}\beta}{m\alpha} \\ &\leq \frac{1}{k+1} \frac{2^{\frac{3m}{2}-1}\beta}{m\alpha} \end{aligned} \tag{A.15}$$

This completes the proof of Theorem 2.

A.4. **Proof of Theorem 3 and Proposition 2.** Our proof follows the same two steps as the proof of Theorem 2. Step 1 is identical: for small-world RGGs the maximal degree d^* increases by at most 1 due to the random rewiring and for random graphs we use prune to bound the maximal degree. Hence, we can again reduce the problem to finding a bound for the probability that two random nodes z and z' lie in the same Voronoi set.

We cannot use the geometric approach from the proof of theorem 2 because a node's neighborhood at distance r increases exponentially due to the random links. In fact we can show that there are constants $C_i > 0$ such that the expected size of any node's neighborhood $B(r)$ at distance r satisfies:

$$C_1 \exp(C_3 r) \leq \mathrm{E}|B_r(z)| \leq C_2 \exp(C_3 r) \tag{A.16}$$

for $r \leq \frac{\ln(n)}{2C_3}$. The proof follows readily from Rosenblat and Mobius (2004). The key intuition is that (a) the expanding neighborhood disk keeps spawning new remote seeds from which Euclidean disks can grow; (b) as the neighborhood grows the overall growth rate of the neighborhood becomes deterministic; (c) we focus on r small enough so that the volume of the r-neighborhood is less than $\sqrt{n}$ and hence the extent to which newly formed disks overlap is small.

Now consider the following thought experiment: consider a seed x as well as one of the initial random nodes (say z). Consider the r-neighborhood around these two nodes and gradually increase r: we now show that the probability that these two

neighborhoods overlap becomes large exactly when they reach size $O(\sqrt{n})$ (which they do simultaneously because the two neighborhoods share the same growth rates).

Consider the single step from r to $r+1$: the number of new nodes around z that have potential connections to the ball around seed x equals $D_z \exp(-C_3 r)$ for some constant $C_1 \leq D_z \leq C_3$. Each such node connects to the outer layer of the ball around x with probability $\eta \frac{D_x \exp(-C_3 r)}{n}$ for some constant $C_1 \leq D_x \leq C_3$. Hence, the probability that none of these new connections connects to the x-ball is equal to:

$$\left(1 - \eta \frac{D_x \exp(C_3 r)}{n}\right)^{D_z \exp(C_3 r)} = \left(1 - \frac{\eta}{\frac{n}{D_x \exp(C_3 r)}}\right)^{\frac{n}{D_x \exp(C_3 r)} \frac{D_x D_z \exp(2C_3 r)}{n}} \tag{A.17}$$

Recall that $D_x \exp(C_3 r) \leq \sqrt{n}$. We therefore express the probability of no new connections as:

$$\exp\left(-\eta \frac{D_x D_z \exp(2C_3 r)}{n}\right) \tag{A.18}$$

For $r = \frac{\ln(n)}{2C_3}$ this probability is bounded below by $\exp(-\eta C_2^2)$ and above by $\exp(-\eta C_1^2)$ and is therefore strictly between 0 and 1.

At the same, since we considered a sparse seed set such that $k = o(\sqrt{n})$ we can ignore the possibility that the r-neighborhood around z contains other seeds. Therefore, we can treat the *connection time* $r^* = T(z, x)$ when the r-disks around z and x start to overlap as independent random variables for different seeds. Since the probability of connecting to any of seeds is bounded away from 0 the node is closest to any specific seed x with probability C/k.

This argument holds for both z and z' independently and hence completes step 2.

A.5. **Proof of Proposition 3.** Observe that the share of agents in the center is $o\left(\left(\frac{2}{3}\right)^r\right) \to 0$ as $r \to \infty$. Therefore, with probability approaching one, all seeds are on the circle since $k = o\left((c \cdot 3/2)^r\right)$.

Condition on an allocation of seeds that are not on the central tree. These are uniformly placed along the outer circle.

We need to compute the distance between the closest seed to a spoke and the second closest seed to a spoke. In order to study this, we need the difference between the first and second order statistics from k draws on a line segment of length $\left(\frac{3}{2}\right)^r$. Note that for a uniform distribution on $[0, 1]$, this order statistic difference is going to be on the order $O\left(\left(\frac{3}{2}\right)^r\right)$.

Next, observe that it takes $O(2r)$ steps for the nearest seed to go up the tree and down the other ends along all other spokes, since the height is r.

This implies that of the $3 \times 2^{r-1}$ nodes at the bottom of the tree, all but $o(1)$ are infected with the signal from the nearest seed to the tree as $r \to \infty$.