NAIVE LEARNING WITH UNINFORMED AGENTS

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

1. Introduction

> ...[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, Secretary of Defense, (2002)*

Learning from friends and neighbors is one of the most common ways in which new ideas and ideas about new products get disseminated. There are two aspects of social learning. First of all, *diffusion models* describe how information spreads from a small set of agents to the wider population of uninformed agents (Calvo-Armengol and Jackson, 2004; Jackson and Yariv, 2007; Banerjee et al., 2013). Informed agents in these models typically all have the same information and the analysis is focused on describing the diffusion patterns as uninformed agents gradually learn about the existence of the new product or service. Second, *information aggregation models* describe how rational or boundedly rational agents aggregate different signals (Bala and Goyal, 2000; DeMarzo et al., 2003). In these models, every agent is typically informed and the analysis is focused on how quickly agents converge to a limit opinion (if any).

Date: February 29, 2016.
We are grateful for financial support from NSF SES-1326661 and IRiSS at Stanford. We also thank Gabriel Carroll, Ben Golub, Matt Jackson, Bobby Kleinberg, Jacob Leshno, Jeff Zwiebel and participants at the MSR Economics Workshop and the Harvard Information Transmission in Networks for helpful discussions.
[†]MIT, Department of Economics; NBER.
[§]Columbia Business School.
[†]Stanford University, Department of Economics; NBER.
[⋆]Microsoft Research New England; NBER.

In reality both processes occur at the same time. For example, during the financial crisis of 2007-2008, most investors might not have tracked news on subprime lending. After all, there are many states of the world to keep track of such as the price of commodities or changes in tax policy. Until investors first encounter information about subprime lending, it is an *unknown unknown*. However, once they hear at least once about problems with subprime lending, they might start to track the state of the market and try to aggregate different opinions about the severity of problems with subprime lending. Microcredit programs provide another example: in the early phase of microcredit, marketing materials of microfinance institutions (MFIs) often featured quotes from their beneficiaries to the effect that they never imagined that they could ever be clients of a formal financial institution. It is unlikely that these women were exchanging signals about the costs and benefits of microcredit before an MFI showed up in the landscape. Even after such an organization opens up nearby, information about its existence does not immediately spread to everyone. 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 letting word of mouth take care of the rest of the communication. Information aggregation only starts after agents have become informed and start to hear about the experiences of different early adopters.

We show that both aspects of social learning can be easily captured by a simple modification of the DeGroot model (DeMarzo et al., 2003) that has emerged as one one of the standard ways to model information aggregation. The standard DeGroot model describes how boundedly rational agents aggregation information within a social network setting: given the sheer complexity of full Bayesian learning on networks resulting from the multiplicity of networks pathways through which the information could have reached someone, analysts have tended to view boundedly rational learning rules as behaviorally more realistic. The experimental evidence also supports this view (Chandrasekhar et al., 2015; Mengel and Grimm, 2015; Mueller-Frank and Neri, 2013). Among boundedly rational learning rules the DeGroot model has the dual advantage of both being intuitively attractive as a description of information aggregation behavior and, under some mild assumptions, displaying attractive long run properties. It is a model of back and forth exchange of opinions within a fixed group, with each agent linearly averaging the opinions he encountered in the previous period with his own to generate his current opinion.

However the analysis of this model rests on one rather restrictive assumption: it requires that everyone in the population starts with signal and the entire learning process evolves as an exchange of opinions among an already informed population. The current paper relaxes that assumption and allows signals to be sparse. 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. More specifically the *Generalized DeGroot* (GDG) updating rule that we introduce has three defining features: for network members who get one of the initial signals (seeds), their starting opinion is their signal; the rest start with no opinion and they remain in that state until one or several agents with an opinion pass this opinion on to them. From then on, agents start to average the opinions of their informed neighbors exactly as in the standard DeGroot model.

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 two informed neighbors, they start exchanging opinions and engage in DeGroot averaging.

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 tessalation* of the social network induced by the set of initially informed agents. 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 get 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 geometry 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 crowd" effect analyzed by Golub and Jackson (2010)), may not do so in the

Generalized DeGroot case.[1] In other words, the long-run outcome may reflect only a fraction of the initially available signals. To demonstrate the 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 particularly high degree. However we can also characterize large classes of other networks where this issue does not arise and there is nearly full signal aggregation even in the sparse case; for example, social networks on rewired lattice graphs as introduced by Watts and Strogatz (1998) do not suffer from belief dictatorships but on the contrary aggregate the initial 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 by injecting signals at a number of randomly chosen nodes (Banerjee et al., 2015). The variance of the long run outcome of our simulated process across multiple rounds of injections gives us a measure of information loss. Our results show that over a range of levels of sparsity at least for these villages, we end up reasonably close to full aggregation; in our simulations we find that the average amount of information loss is only 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 only 13% of information.

We then explore the reasons why there is variation in the degree of information aggregation. Our results point to the relative size of the Voronoi sets as a likely explanatory variable, though it is silent about the relative quantitative importance of that effect versus that of other aspects of the network structure. To get at some magnitudes, we construct the Voronoi sets for each of the 75 networks in our data set for each realization of the signal distribution. Then, using those Voronoi sets, we compute the Herfindahl index of their size distribution, to capture the idea that

[1]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.

inequality in the size of the Voronoi sets leads to overweighting of certain signals and therefore information loss. Finally we regress our measure of information loss on the Herfindahl index, which turns out to be strongly significant and negative, capturing the idea that more inequality in the size of the Voronoi sets is associated with greater information loss. Our calculations suggest that variation in the size distribution of the Voronoi sets explains approximately one fifth of the variation in the extent of information loss.

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 a Voronoi-weighted average of the initial signals. In Section 4 we look at how the geometry of the network influences the extent of information preservation or information loss. Section 5 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 observe a signal about the state $\theta \in \mathbb{R}$ of the "world". The reason the word "world" is in quotes is because, as already noted, DeGroot models are about opinions and not necessarily anything objective about the world; the opinions could be about whether it is true that the Chinese have recreated unicorns.[2]

There are a finite number n of agents who are embedded in a fully connected and symmetric graph g such that $(i, j) \in g$ implies $(j, i) \in g$ for any two agents i and j. We denote the degree of a node in the graph with $d_i - 1$, so d_i represents a self-loop. We maintain throughout that the stochastized matrix corresponding to g is an irreducible, aperiodic stochastic matrix.[3]

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 some distribution F:

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

[2] http://christwire.org/2011/03/evil-chinese-scientists-recreate-real-life-unicorn-baby/

[3] Therefore, once everyone is informed, standard DeGroot results apply.

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 (dense) DeGroot case.

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 this set can include the agent herself. We then specify the *generalized DeGroot* (GDG) updating process as follows:

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

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 to derive a new opinion.[4] Note, that our updating rule reduces to the standard DeGroot model if every agent is initially informed. Also Section 5 spells out a potential foundation for this rule: it can be seen as a naive dynamic extension of the static optimum Bayesian learning rule.

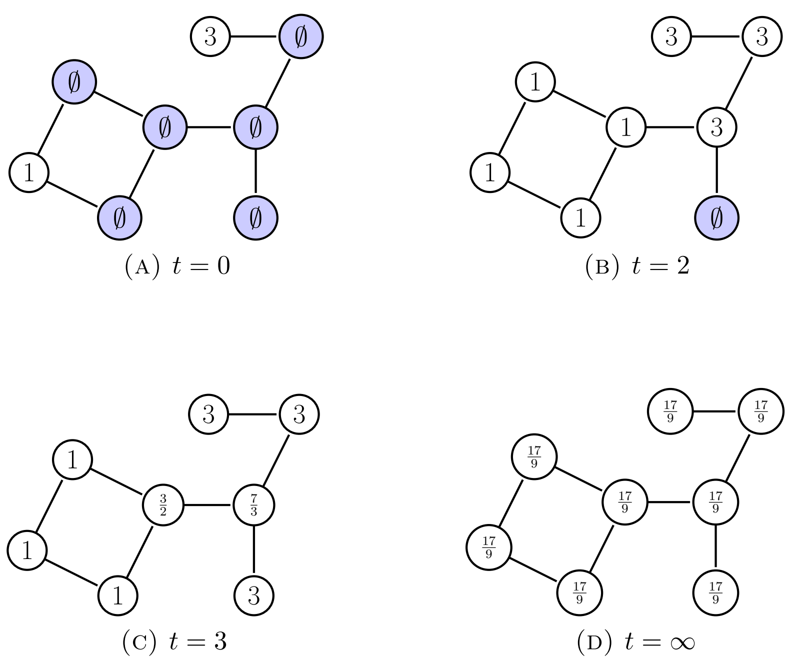


FIGURE 1. Evolving beliefs in a sample social network

[4]Our results generalize to non-uniform weighting, but they are cleaner to present in this way.

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 *diffuse* and the direct and distance 2 neighbors adopt the opinion of the seed closest to them. In period 3, 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.

3. How network geometry affects limit beliefs

We next characterize limit beliefs in our model starting from the initial seed set $S = \{i_j\}$ $(j = 1, ..., k)$ 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.

Proposition 1. *The limit belief x^∞ is a weighted average $\sum w_j(S)x_j^0$ of the initial signals of the seeds, where the weights given to the signal of seed j, $w_j(S)$, only depend on the position of the seeds in the network.*

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.

It will be convenient to assume from now on that the weights $w_j(S)$ are monotonic in the index j – 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^∞:

$$\text{var}(x^\infty) = \sum_j w_j(S)^2\sigma^2. \tag{3.1}$$

We can bound this variance above and below:

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

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 $k-1$ pieces of information. The lower bound is just the variance of the sample mean of k independent draws.

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 (once again the quotes on wisdom represent the fact that we are talking about opinions and not facts). On the contrary, if the variance in the limit belief is close to the upper bound we say the process exhibits "dictatorship" because it puts exclusive weight on the signal of one single agent.

In order to understand the conditions under which wisdom or dictatorship arises we have to understand the weights $w_j(S)$. 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 disjoint sets. Each Voronoi set is associated with a seed i and contains all the nodes that are closer to seed i than any other seed in terms of social distance and *include* ties. These sets are disjoint except at the boundaries which might overlap. For each Voronoi set V_i define the boundary of the set to be ∂V_i which we obtain by extending the Voronoi set V_i one more step outward. We define the lower Voronoi set, $\underline{V}_i$, where we assign values $\underline{x}_i$ to the minimum of any associated Voronoi set *or* the boundary of a Voronoi set. Intuitively, this creates "fat" boundaries on a line of either width 2 (if the distance between adjacent seeds is odd) or width 3 (if the distance is even). We define the upper tessalation, $\overline{V}_i$, in an analogous manner.

Intuitively, the Voronoi tessalation captures the diffusion of information in our model. All agents who reside strictly within a set will first adopt the information of the closest seed before entering an averaging phase. We can now show that it is sufficient to study the Voronoi sets in order to characterize the limit belief x^∞. To state the result we denote the share of nodes in a network that are part of the lower Voronoi set $\underline{V}_i$ with $\underline{v}_i = \frac{|\underline{V}_i|}{n}$ and define the link-weighted share:

$$\underline{v}_i^* = \frac{\sum_{i \in \underline{V}_i} d_i}{\sum_{i=1}^n d_i}. \tag{3.3}$$

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 $\underline{v}_i = \underline{v}_i^*$.

Theorem 1. *Assume a social network with seed set S. The limit belief is bounded below and above as follows:*[5]

$$\sum_i \underline{v}_i^* x_i \leq x^\infty \leq \sum_i \overline{v}_i^* x_i \tag{3.4}$$

The intuition for this Theorem 1 is that once everyone has an opinion, Generalized DeGroot learning is just like the standard DeGroot learnimg (DeMarzo et al., 2003; Golub and Jackson, 2010). However the starting signals for this averaging phase are generated by the diffusion phase and are summarized by the size of each element in the Voronoi partition (measured by $\underline{v}_i^*$ or $\bar{v}_i^*$) and the signals associated with it.

Theorem 1 allows us to characterize the limit belief by studying a static problem and relates the *geometry* of the social network to the opinion weights ($w_i(S)$).

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 society can just lean on a law of large numbers.

It is instructive to start by comparing the case of sparse signals to the case when everyone gets a signal (again, we call this 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. Information such as the lengths of paths are irrelevant. Formally, Golub and Jackson (2010) show that a sequence of graphs $(g_n)_{n\in\mathbb{N}}$ is wise only if

$$\max_{1\leq i\leq n} \frac{d_i(g_n)+1}{\sum_k d_k(g_n)+n} \to 0.$$

Notice that this condition puts restrictions only on local structure, that is, only the degree distribution of the network affects the limiting belief.

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 that over time lose track of all but one signal. That is, the society's converged opinion may reflect exactly one signal and therefore be arbitrarily close

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

to having the maximal possible variance. More generally, this parable suggests that networks with striking asymmetries may destroy considerable information in sparse learning environments.

We also explore networks that are best described as lattice graphs with short-cuts, leaning on Watts and Strogatz (1998). This models environments best described by homogenous small-world networks, that may be realistic in many contexts. First, we first show that lattice-like graphs exhibit wisdom. Second, we prove that adding a small number of shortcuts to a lattice graph induces only small changes in the variance of limit beliefs and therefore preserves wisdom.

4.1. **Belief Dictators.** 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.

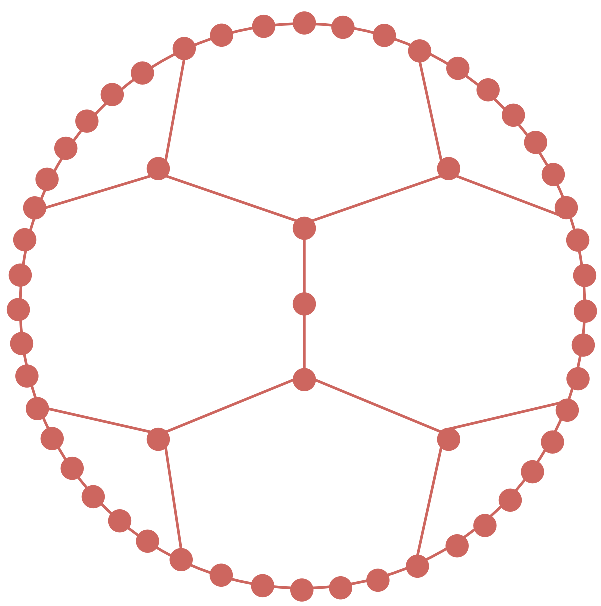


FIGURE 2. Belief Dictators Example

For each integer r we define a social network $T(r)$ 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:

$$1 + 3 + 3 \times 2 + 3 \times 2^2 + ... + 3 \times 2^{r-1} = 1 + 3\left(2^r - 1\right). \tag{4.1}$$

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.

Theorem 2. *Consider the class $T(r)$ of social networks and assume that k seeds are randomly chosen on the network. The expected value of the largest weight $E_S(w_k(S))$ (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. Therefore, it becomes increasingly unlikely that any of the seeds are located in the center for large r. 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.

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

4.2. **Wisdom in small world lattices.** The previous example is a case where almost all information is destroyed leaving just one signal to dominate. This happens because in almost any allocation of initial seeds, the induced Voronoi sets are such that one set is much, much larger than all of the others.

In this section we study a class of networks where for a typical seeding, the Voronoi sets of seeds are essentially all of the same order of magnitude. In this case, wisdom is actually preserved in the sense that the final opinion reflects information from all k seeds.

For this exercise, we look at small world networks on lattice graphs, building on Watts and Strogatz (1998). First, we first show that lattice-like graphs exhibit wisdom. Second, we prove that adding a small number of shortcuts to a lattice graph induces only small changes in the variance of limit beliefs and therefore preserves wisdom.

4.2.1. *Lattice-like Graphs.* As the name suggests, lattice-like graphs resemble lattice graphs such as the one-dimensional line or circle or the two-dimensional plane or torus. We start by defining the concept of an r-ball $B_i(r)$ which is the set of nodes at distance at most r from agent i.

Definition 1. *The class $G(a, A, m, d)$ of social networks consists of all social networks of bounded degree d whose r-balls satisfy the following property for all r and i:*

$$ar^m \leq B_i(r) \leq Ar^m \tag{4.2}$$

Intuitively, the parameter m describes the dimensionality of the network. For example, the class of circle networks where agents interact with their direct neighbors belongs to the class $G(1, 1, 1, 2)$ while the class of torus networks belongs to $G(1, 1, 2, 4)$. At the same time, the definition is flexible enough to allow for *local* re-wiring. For example, consider a circle network and add, for each agent, up to two more links to neighbors at most distance R away. The resulting network belongs to the class $G(1, 2, 1, 4)$ of lattice-like networks: the network is no longer a regular network as agents can have degree ranging from 2 to 4 but it still resembles a one-dimensional line. Similarly, the geographic networks studied by Ambrus et al. (2014) belong to the class $G(a, A, 2, d)$ for appropriately chosen parameters a, A and d which is a generalization of regular two-dimensional torus networks.

Theorem 3. *Consider the class $G(a, A, m, d)$ of social networks and assume that k seeds are randomly chosen on the network. Then there is a constant C that does not depend on k or n such that we can bound the variance in the limit opinion as follows:*

$$\mathrm{E}_S\left[\mathrm{var}(x^\infty)\right] \leq \frac{C\sigma^2}{k}. \tag{4.3}$$

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

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

The theorem shows that for most seeds sets the variance in the limit belief is at most a constant factor (which is independent of both n and k) 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 and therefore the generalized DeGroot process aggregates opinions far better than belief dictatorships. We can therefore view Theorem 3 as an approximate "wisdom of crowds" result similar to Golub and Jackson (2010) for this class of networks.

4.2.2. *Small World Graphs.* The seminal work of Watts and Strogatz (1998) emphasizes that real-world social networks have small average path length, and note that this cannot be generated from lattice graphs by local re-wiring only. Instead, we have to allow for limited *long-range* rewiring that creates shortcuts in the social network.

Formally, we define a $R(\eta, D)$ rewiring of the class of lattice-like graphs $G(a, A, m, d)$ by adding – independently, and identically for each node i in the network – between 0 and D links to randomly selected nodes in the social network such that the expected number of new nodes for each agent i is η.

Theorem 4. *Consider a $R(\eta, D)$ re-wiring of the class $G(a, A, m, d)$ of social networks and assume that k seeds are randomly chosen on the network. Then there is a function $C(\eta)$ that is continuous over the range $[0, D]$ and does that does not depend on k or n such that we can bound the variance in the limit opinion as follows:*

$$\mathrm{E}_{S,R(\eta,D)}\left[\mathrm{var}(x^\infty)\right] \leq \frac{C(\eta)\sigma^2}{k}. \tag{4.4}$$

This result implies that a small amount of long-range rewiring only changes the upper bound on the variance of the limit belief slightly for *most* seed sets and *most* rewirings.[6] Importantly, the change is independent of k and n.

The intuition behind this result is that even though long-range re-wiring has a dramatic effect on average path length it affects the diffusion ability of every seed in an equal manner. Therefore, it does not exacerbate the imbalance of the Voronoi set size distribution.

[6]Note, that we take the expectation both over seed sets and re-wirings. In particular, there is always a positive probability of obtaining a network akin to the $T(r)$ class of social networks that we studied in Section 2 which gives rise to belief dictatorships.

4.3. **Simulations in Indian Village Networks.** We have explored network geometries where belief dictatorships arise (i.e.., where $k-1$ units of information are destroyed) as well as cases where there is wisdom (i.e., all k units of information are preserved).

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.

4.3.1. *Data Description.* For this exercise we use the household network data collected by Banerjee et al. (2015). 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.[7] Thus, our resulting empirical networks are undirected.[8] 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 which 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 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 binary, $\theta \in \{0, 1\}$. Further, we assume signals to be Bernoulli and to be correct with probability

[7]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.

[8]See Banerjee et al. (2013) and Banerjee et al. (2015) for a detailed description of the data collection methodology and for a general discussion of the data.

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

$\alpha > 1/2$.[9] Here, we set $\alpha = 0.6$. 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 $\frac{\sigma^2}{k} = \frac{\alpha(1-\alpha)}{k}$.

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

$$k^{effective} := \frac{\alpha(1-\alpha)}{\sigma^2_{x^\infty}}.$$

Given that $\sigma^2_{x^\infty} \leq \frac{\alpha(1-\alpha)}{k}$, $k^{effective}$ (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 $\frac{k^{effective}}{k}$.

[9]This is a slight abuse of notation. Formally, since we have mean-zero signals, we should denote the signals as recentered: $1 - \alpha$ and $0 - \alpha$.

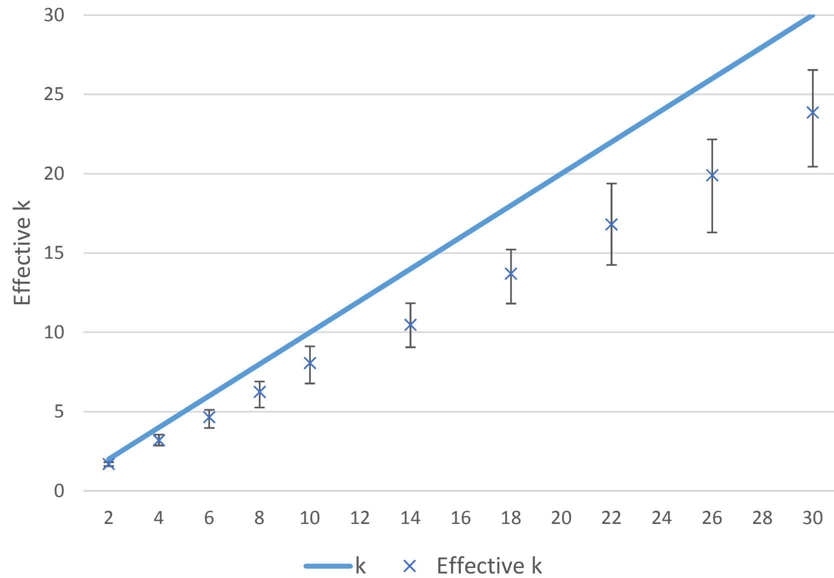


FIGURE 3. Plots of the mean $k^{effective}$ against k, where the average is taken over all simulations and all networks. The bars represent the interquartile range of $k^{effective}$ across networks, for each k.

4.3.3. *Results.* In Figure 3, 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. To investigate the value of an additional signal for GDG agents, we run a regression of $k^{effective}$ on k (i.e., calculate the slope of the line through the points in Figure 3). The results are displayed in Column 1 of Table 2. On average, adding one additional signal improves $k^{effective}$ by 0.775 signals. To further explore information loss, we plot in Figure 4 the fraction of information preserved, $\frac{k^{effective}}{k}$ as a function of k. Averaging across each k, we find that 78.4% of signals are preserved. We find no clear relationship between information loss and k.

Further, we find substantial heterogeneity in the degree of information loss across the 75 networks. In both Figure 3 and Figure 4, 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 substantial heterogeneity. On average, the 25th percentile village experiences 33% information loss, while the 75th percentile village experiences only 13% information loss.

Our model has empirical predictions for when we should expect a network to exhibit relatively more information loss. That is, when the sizes of the Voronoi sets are

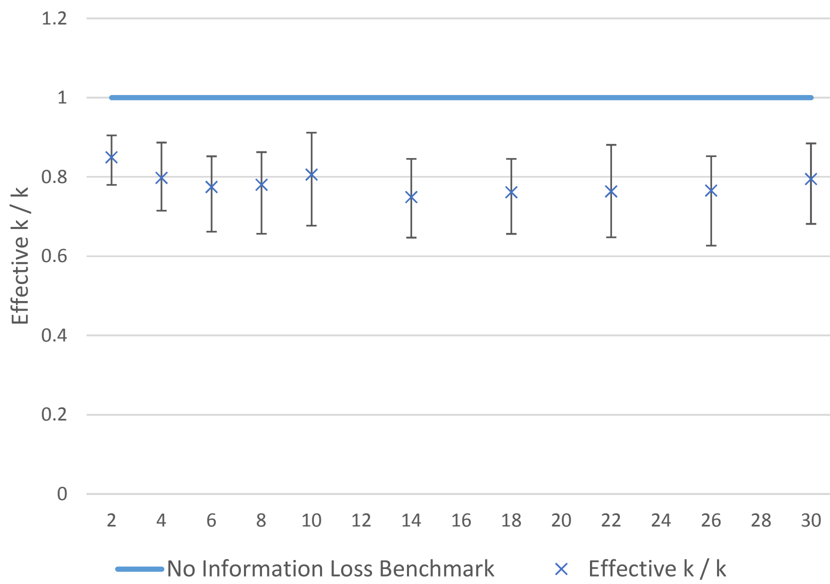


FIGURE 4. Plots of the mean $\frac{k^{effective}}{k}$ against k, where the average is taken over all simulations and all networks. The bars represent the interquartile range of $\frac{k^{effective}}{k}$ across networks, for each k.

TABLE 2. Regression results.

	(1) Effective k	(2) Preserved Information
k	0.775***	-0.000238
	(0.0156)	(0.000819)
Adjusted Herfindahl (k, village)		-0.0213***
		(0.00676)
k Fixed Effects	No	Yes
Observations	750	750
R-squared	0.874	0.022

Standard errors are clustered at the village level. Sample includes outcomes at the village level for each k, averaged across all 50 simulation draws. The Herfindahl measure is defined as the Herfindahl index of the sizes of the Voronoi cells for a given signal draw, fixing k and village, averaging over all 50 draws. Adjusted Herfindahl scales by k to make the measure comparable across values of k and divides the resulting measure by its standard deviation (0.215). Preserved information is defined as Effective k/k. *** p<0.01, ** p<0.05, * p<0.1

relatively unequal for a given k and a given seeding. To investigate this further, for each simulation draw s, for each village v, and for each k, we calculate the Herfindahl

Index, $H_{v,k,s}$ for that observation's Voronoi sets, that is $H_{v,k,s} = \sum_i (\frac{|V_{i,v,k,s}|}{n_v})^2$, where n_v is the size of village v. We then average this value across the simulation draws $\bar{H}_{v,k} = \frac{\sum_s H_{v,k,s}}{50}$ to obtain a measure that only depends on v and k. Note that in the dictator example $\bar{H}_{v,k}$ is large (and close to the maximal value of 1) for any finite graph.

In Column 2 of Table 2, we investigate whether $\bar{H}_{v,k}$ predicts information loss. In order to compare $\bar{H}_{v,k}$ across k, we consider the normalization $\frac{\bar{H}_{v,k}}{\sum_s (\frac{1}{k})^2} = k\bar{H}_{v,k}$. Note that this normalization scales the average Herfindahl for a given village, k pair by the benchmark Herfindahl value obtined under equally-sized Voronoi cells. We find that larger Herfindahl measures are indeed predictive of more information loss. Increasing $k\bar{H}_{v,k}$ corresponds to an additional information loss of 2.13 percentage points.

In sum, it appears that on average real world social networks do reasonably well at preserving information, though there is a compelling amount of heterogeneity there. We also find that networks that tend to have unequally-sized Voronoi sets on average do worse at preserving information.

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 or more likely to generate information loss.

5. Discussion and Conclusions

The Generalized DeGroot rule is a natural extension of the standard DeGroot rule to the sparse signal setting in the following sense: the standard DeGroot model has been justified as the multiperiod extension of an agent applying, naively, a one-period Bayesian updating rule (DeMarzo et al., 2003). Here we will argue that our Generalized DeGroot process is the obvious parallel version of naive Bayesian learning when there are uninformed agents.

For this argument assume that the signals are drawn normally: $F(\theta, \sigma^2) = \mathcal{N}(\theta, \sigma^2)$. In the framework of DeMarzo et al. (2003) the DeGroot learner updates "correctly" in a Bayesian sense in period $t = 1$ by averaging over her own and her neighbors' opinions but then naive applies the same rule in every period without taking into account that their neighbors' beliefs are no longer independent signals at time $t > 1$. Hence we can view a DeGroot learner as a naive Bayesian.

How would a naive Bayesian learner update in our model in period $t = 1$? In order to be able to perform Bayesian updating in our model we assume that an uninformed agent i has a normally distributed but highly imprecise signal $\tilde{x}_i$:

$$\tilde{x}_i^0 = \theta + \tilde{\epsilon}_i \quad \text{where} \quad \tilde{\epsilon}_i \sim \mathcal{N}(0, \tilde{\sigma}^2) \tag{5.1}$$

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 the informed Bayesian agent will optimally update in period $t = 1$ by averaging over the opinions of all informed neighbors in the set J_i^t (possibly including herself).[10] Moreover, an uninformed agent who has only uninformed friends will continue to hold a highly imprecise belief (hence she remains "uninformed"). Therefore, the Generalized DeGroot learner can again be viewed as a naive Bayesian.

Nevertheless this rule is just one of a continuum of possible learning rules with a DeGroot flavor which differ in the way we treat the fact that we do not get signals from uninformed agents. An alternative formulation could be the following. Suppose agents do not know k but know the size of the social network n. Furthermore, they know that there is something to be learned about beginning at $t = 0$. In this variation we are allowing them to keep track of two things instead of one – the estimate of the state (θ) and an estimate of the share of uninformed agents in the population (k/n).

The opinions are updated based on Generalized DeGroot learning while the estimate of the share of uninformed agents is updated through a standard DeGroot learning process where every agent's initial opinion on the share of informed agents at time $t = 0$ is equal to the local population share in her immediate neighborhood.

Formally, consider the learning process as two-dimensional

$$x_i^t = \left(x_{i,\theta}^t, x_{i,k}^t\right)'$$

where these track estimates of θ and k/n.

This corresponds to two agents having a conversation where an agent tells the other her estimate of the mean, and also her current estimate of what share in the population got signals initially.

In this case, $x_{i,k}^1$ just encodes the share of one's neighbors that drew signals $x_{i,\theta}^0 \neq \emptyset$. And this can be thought of as essentially standard (dense) DeGroot since every agent does get a "signal", so the limit opinion will be

$$x_{i,k}^\infty = \sum_j \frac{d_j + 1}{\sum_l d_l + 1} x_{i,k}^1.$$

[10]This averaging rule becomes exact as $\sigma_\theta^2 \to \infty$ and $\tilde{\sigma}^2 \to \infty$.

If the size n of the social network is large enough (and degree distribution is bounded from above) then agents will learn the share of informed agents with high precision in the limit as in Golub and Jackson (2010), taking into account minor modifications to the argument since correlation of $x^1_{i,k}$ and $x^1_{j,k}$ now depends on the network topology. Therefore, equipped with x^∞_θ and x^∞_k, the agent has an estimate of k, $\widehat{k} = x^\infty_k \cdot n$ as well as an estimate of θ given by x^∞_θ as before. In this way she can have a sense of how many effective signals are being incorporated into her limit opinion.

In one sense this is a more attractive model because by learning k the agent learns the precision of his etimate of θ and not just the mean. This is not an issue in the standard DeGroot case with a large population, because the aggregation of a very large number of signals automatically removes the possibility of imprecision, but with small k precision is obviously important.

On the other hand to implement this alternative rule the decision-maker needs to have started keeping track of the share of uninformed agents from the beginning of time – by the time he acquires an opinion of his own, it is typically too late, since everyone else around him also has an opinion.

Such anticipatory behavior is perhaps more natural when the state of the world is about *known unknowns* – the agent is aware of at time $t = 0$ even if he has no signals about these states. For example, agents might have no information about the state of the economy but they are aware of concepts such as GDP growth and they 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, in many circumstances we have to deal with *unknown unknowns* which are states of world whose existence we are unaware of until some signal reaches us. For example, we might be unaware of the existence of a new electronics gadget until we listen to a friend who extols on the product's quality and usefulness. At this point, it is too late to start keeping track of the share of agents who have first-hand information about the product: while we will generally learn the average experience of first-hand users it will be impossible to determine whether the average is based on the experiences of a small or large number of first-hand users.

More generally, our analysis suggests that thinking harder about how to treat the absence of signals in a DeGroot framework will be key to the future of this agenda.

References

Ambrus, A., M. Mobius, and A. Szeidl (2014): "Consumption Risk-Sharing in Social Networks," *American Economic Review*, 104, 149–82. 4.2.1

Bala, V. and S. Goyal (2000): "A noncooperative model of network formation," *Econometrica*, 68, 1181–1229. 1

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013): "Diffusion of Microfinance," *Science*, 341, DOI: 10.1126/science.1236498, July 26 2013. 1, 8

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2015): "Gossip: Identifying central individuals in a social network," . 1, 4.3.1, 8

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 (2015): "Testing models of social learning on networks: Evidence from a lab experiment in the field experiment," *NBER Working Paper 21468*. 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, 3, 5

Golub, B. and M. Jackson (2010): "Naive Learning in Social Networks and the Wisdom of Crowds," *American Economic Journal: Microeconomics*, 2, 112–149. 1, 3, 4, 4.1, 4.2.1, 5

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

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

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

Watts, D. and S. Strogatz (1998): "Collective dynamics of small-world networks," *Nature*, 393, 440–442. 1, 4, 4.2, 4.2.2

Appendix A. Proofs

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 $j \in S$. □

The proof of Theorem 1 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 $|N(j)| + 1$. 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 provides us with the well-known limit belief of the DeGoot model with symmetric links.

Proof of Theorem 1. Without loss of generality, we assume that all initial opinions of seeds are positive.[11]

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

[11]We can always ensure that by adding a constant to all opinions.

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:

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

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, 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.

Since we have $\hat{d}_j \leq d_j$ we obtain the inequality:

$$\sum_{j \in S^t} d_j x_j^t + \sum_{j \in \partial S^t} d_j \underline{x}_j \leq \sum_{j \in S^{t+1}} d_j x_j^{t+1} \leq \sum_{j \in S^t} d_j x_j^t + \sum_{j \in \partial S^t} d_j \overline{x}_j \tag{A.8}$$

Since the claim holds at time t we can deduce:

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

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

Proof of Theorem 2. 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.

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 some function of k, independent of r. And therefore, in our case, the distance must 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 $r \to \infty$. □

Proof of Theorem 3. Consider a graph with n nodes. Seeds are placed on these nodes in a way such that every node is selectd to be a seed with probability $q = k/n$, so the expectd number of seeds is given by k.

Any particular seed realization S induces a Voronoi tessalation, and we are interested in upper-bounding the second moment of this tessalation

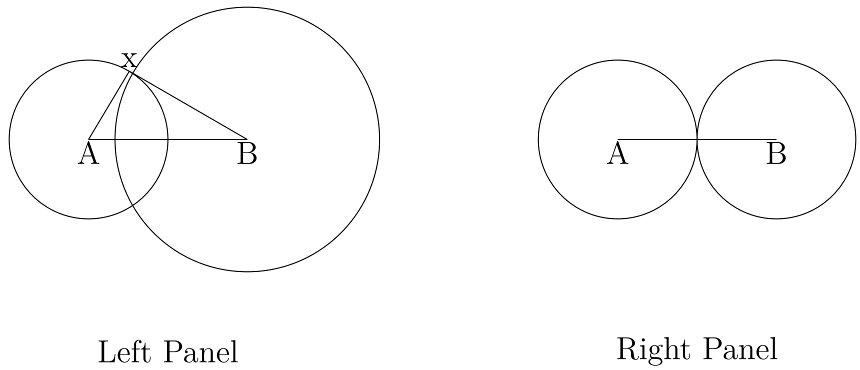
$$\mathrm{E}_S\left[\sum_{i \in S} v_i^2\right] \tag{A.10}$$

where we take the expectation over all the possible seed sets induced by the Poisson process. Note, that we know that $\mathrm{E}[|S|] = k$. We can then use Theorem 1 to derive an upper bound on the variance in the final opinion x^∞ since the link-weighted Voronoi shares are equal to v_i for regular lattice graphs.

Now consider the following thought experiment: draw two random points A and B from the set of nodes and consider the event that both of them are in the same Voronoi set. We define the random indicator variable I_{AB} which equals 1 iff A and B are in the same Voronoi set:

$$\mathrm{E}_S\left[\mathrm{E}_{A,B}\left(I_{AB}\right)\right] = \mathrm{E}_S\left[\sum_{i \in S} v_i^2\right] \tag{A.11}$$

FIGURE 5. Geometric Interpretion of $E_S\left(I_{AB}\right)$



This equation follows because the probability that point A is inside the ith Voronoi set is exactly v_i.

Note, that we are taking first the expectation over all points A and B and then the expectation over the seed sets. We now use Fubini's theorem (it is trivial for finite sums) to change the order of integration:

$$\mathrm{E}_S\left[\sum_{i\in S} v_i^2\right] = \mathrm{E}_{A,B}\left[\mathrm{E}_S\left(I_{AB}\right)\right] \tag{A.12}$$

Let's focus on the inner integral $\mathrm{E}_S\left(I_{AB}\right)$ for two fixed points A and B which is simply the probability that A and B belong to the same Voronoi set when we integrate over all possible tessalations. Let us denote the closest seed with x.

Figure 5 presents the next step graphically, in $\mathbb{R}^2$ for simplicity instead of on a discrete graph. We can illustrate the inner integral in through the left panel of Figure 5. Conditional on x, the two balls around A and B should not contain any seeds. Let F be the joint area covered by both discs.

We want to compute the probability that, conditional on x, there are no other seeds closer to either A or B. It is useful to define $d_A := d\left(A,x\right)$ and $d_B := d\left(B,x\right)$, and define F as volume of the union of the two d_A and d_B balls around A and B respectively. We want to compute the probability that there are no seeds in the F nodes:

$$\mathrm{E}_S\left[I_{AB}\right] = (1-q)^F = \left(1-\frac{k}{n}\right)^F = \left[\left(1-\frac{k}{n}\right)^n\right]^{\frac{F}{n}} \to e^{-k\frac{F}{n}}.$$

The next step is to bound F from below, which we illustrate in the right panel of Figure 5. Let $2r := d\left(A,B\right)$. Then consider the two r-balls around A and B and notice that

$$F \geq B_A(r) + B_B(r).$$

Under the main assumption of the theorem, for some m, there exist constants C_1 and C_2 such that

$$2C_1 r^m \leq B_A(r) + B_B(r) \leq 2C_2 r^m.$$

In this case

$$\mathrm{E}_S[I_{AB}] \leq e^{-kC_3 r^m} + \mathit{Remainder}.$$

Now we integrate over all points A and B – more precisely, over all possible distances $2r$ between two random points A and B, to get

$$\mathrm{E}_{A,B}[\mathrm{E}_S(I_{AB})] \leq C_4 \sum_r e^{-kC_3 r^m} \mathrm{P}(d(A,B) = 2r) = \frac{C_5}{k} + o(1/k).$$

□