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 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. We then simulate the modeled learning process on a set of real world networks; for these networks there is relatively little information loss. 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 can be substantial.

Keywords: Social Networks, Social Learning, DeGroot Learning

Date: July 17, 2018.
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, Secretary of Defense, (2002)*

1. Introduction

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 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, 2000; 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.[2,3] However, the DeGroot model makes the somewhat unrealistic assumption that everyone is informed about the issue at hand to start with – 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, agents in our *Generalized DeGroot* (GDG) updating rule only average the opinion of their *informed* neighbors while ignoring uninformed neighbors. Hence, an agent who received a seed signal and is surrounded by uninformed neighbors will stick to this 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 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 with normal signals where uninformed agents have weak and diffuse signals that are ignored in the aggregation while the stronger signals of informed agents are averaged.

[2]The DeGroot model has a number of clear advantages. 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 model than the Bayesian alternative (Chandrasekhar et al., 2015; Mengel and Grimm, 2015; Mueller-Frank and Neri, 2013). Finally the model has attractive long-run properties under some relatively weak assumptions.

[3]Molavi et al. (2017) develop axiomatic foundations of DeGroot social learning. They begin with the assumption of imperfect recall – that the current belief of an individual's neighbor is a sufficient statistic for all available information, ignoring how and why these opinions were formed. They show how under this assumption and other restrictions on information processing, DeGroot or DeGroot-like log-linear learning emerges.

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. 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.[4]

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 tessalation therefore describes the seeded agents' social influence in our model unlike 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 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 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 crowds" effect analyzed by Golub and Jackson (2010)), may not do so in the Generalized DeGroot case.[5] In other words, the long-run outcome may reflect only a fraction of the initially available signals. To demonstrate a worst-case version

[4]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.

[5]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

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 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 (Banerjee et al., 2015) 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. 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.

Throughout 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 we look at how the geometry of the network

---

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.

influences the extent of 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 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 and symmetric graph g such that $(i, j) \in g$ implies $(j, i) \in g$ for any two agents i and j. 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).

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:

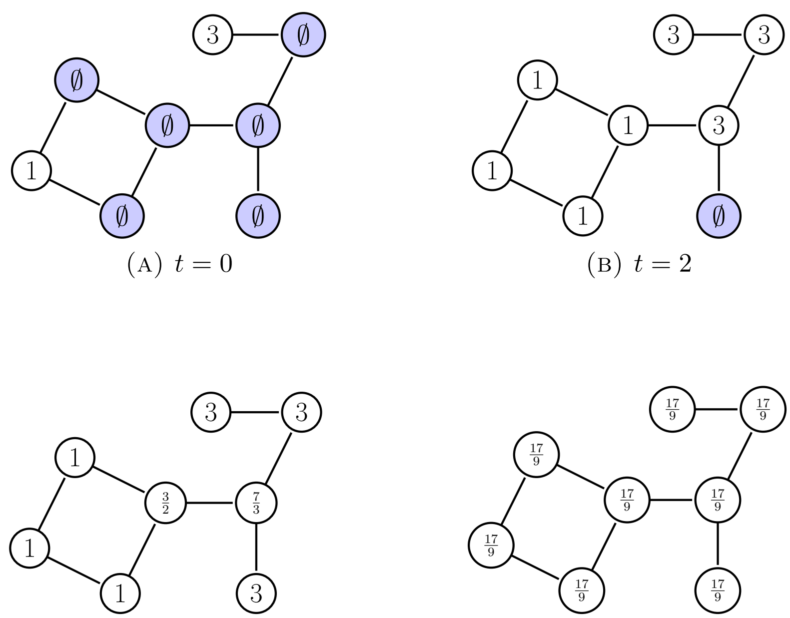
$$x_i^0 = \theta + \epsilon_i \quad \text{where} \quad \epsilon_i \sim F(0, \sigma^2). \tag{2.1}$$

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 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.[6] Note, that our updating rule reduces to the standard DeGroot model once every agent is informed. Also 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.



(A) $t = 0$ (B) $t = 2$ (C) $t = 3$ (D) $t = \infty$

FIGURE 1. Evolving beliefs in a sample social network

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, 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.

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

3. How network geometry affects 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.

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 $\sum w_i(S) = 1$.*

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 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.

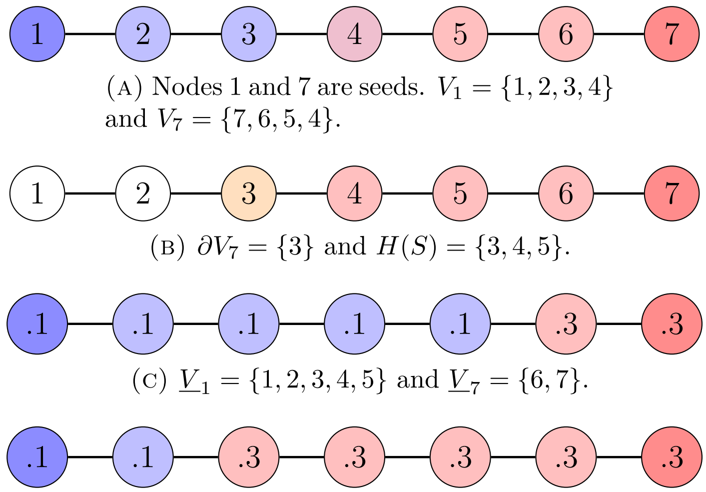
Loosely speaking, we say that the generalized DeGroot process exhibits "wisdom" if the variance of the limit belief is close to the lower bound, which is precisely achieved by the optimal estimator. On the contrary, if the variance in the limit belief is close to the upper bound we say the process exhibits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

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 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. 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 define to 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 one. 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:

$$V_i^{\min} \subseteq \underline{V}_i \subseteq V_i^{\max}$$
$$V_i^{\min} \subseteq \overline{V}_i \subseteq V_i^{\max} \tag{3.3}$$

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:*[7]

$$\sum_i \underline{v}_i^* x_i^0 \leq x^\infty \leq \sum_i \overline{v}_i^* x_i^0 \tag{3.5}$$

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 $w_i(S)$.

Corollary 1. *The social influence $w_i(S)$ of seed i satisfies $v_i^{*,\min} \leq w_j(S) \leq v_i^{*,\max}$ where $v_i^{*,\min}$ and $v_i^{*,\max}$ are the link-weighted shares of the minimal and maximal Voronoi sets, respectively.*

The proof of Corollary 1 follows immediately from inequality 3.3 and Theorem 1 by setting all signal's 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.

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.

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

It is instructive to start by comparing the case of sparse signals to the case when everyone gets a signal (which we call 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.[8]

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 may reflect just one signal and therefore be arbitrarily close to having the maximal possible variance. More generally, this suggests that networks with important asymmetries may destroy considerable part of the available information in sparse learning environments.

We then explore the class of networks that are best described as lattice graphs with shortcuts, leaning on Watts and Strogatz (1998). This models environments best described by homogenous small-world networks, which may be realistic in many contexts. First, we first show that lattice-like graphs exhibit wisdom even with sparse signals. Second, we prove that adding shortcuts to a lattice graph—wherein a random set of links in the lattice are rewired randomly to other nodes, thereby creating short paths across the network—preserves this result.

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.

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:

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

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

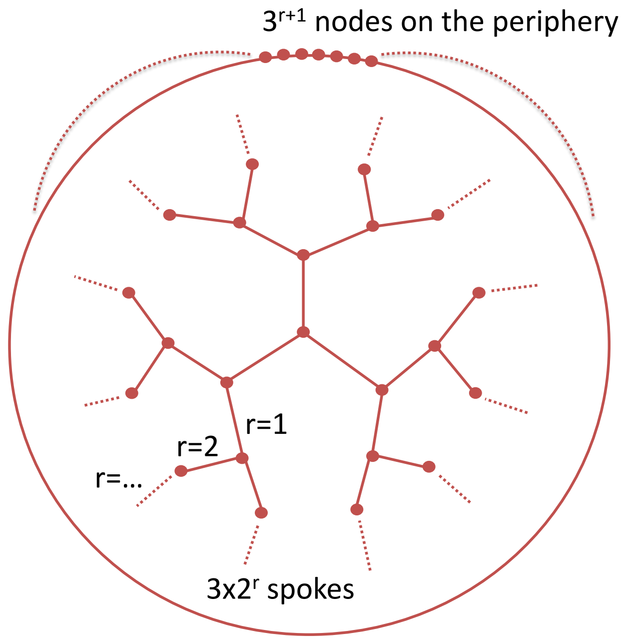


FIGURE 3. Belief Dictators Example

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 2. *Consider the class $G_T(r)$ of social networks and assume that k seeds are randomly chosen on the network. The expected 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, it becomes highly unlikely for large enough r 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.

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 aggregates final opinions (almost) efficiently in the standard (dense) DeGroot model. However, since our process adds a diffusion stage to the social learning process, 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, the wisdom of crowds result continues to hold in the sense that the final opinion reflects equally weighted 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 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 a circle or the two-dimensional plane or a 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 $G(a, A, m, d, \rho)$ of social networks consists of all n finite, integral and positive, social networks with n nodes and bounded degree d with the property that for each node i in a given network z of size n there is an $r_{max}(z)$ with $B_i(r_{max}(z)) \geq \rho n$ and the following property holds for all $r \leq r_{max}$:*

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

where $a, A > 0$ and m is a positive integer.

This regularity property ensures that the networks that we consider do not have regions that grow at very different rates. For example we cannot have one region that is tree-like and another region that is a line.[9] Intuitively, the parameter m describes the dimensionality of the network while ρ represents the minimum slice of the social network for which this property has to hold.[10] For example, the class of circle networks where agents interact with their direct neighbors belongs to the class $G(1,1,1,2,\frac{1}{2})$ while the class of torus networks belongs to $G(1,1,2,4,\frac{1}{2})$. At the same time, the definition is flexible enough to allow for *local* rewiring. 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 $G(1,2,1,4,\frac{1}{2})$ 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 $G(a,A,2,d,\frac{1}{2})$ for appropriately chosen parameters a, A and d which is a generalization of regular two-dimensional torus networks.

Theorem 2. *Consider the class $G(a,A,m,d,\rho)$ 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:*

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

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

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

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 2 as an approximate "wisdom of crowds" result similar to Golub and Jackson (2010) for this class of networks.

[9]This property obviously cannot hold for all r because eventually the balls will cover the entire network, which is why we only require to hold up to some r_{max}.

[10]For example, we cannot take a torus and make it very thin so that it resembles a circle rather than a plane.

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 rewiring only. Instead, we have to allow for limited *long-range* rewiring that creates shortcuts in the social network.

Formally, we define a $R(\eta)$ rewiring of the class of lattice-like graphs $G(a, A, m, d, \rho)$ as follows: we randomly pair all agents in the network with a random partner and with probability η we add a link between these two nodes for each of the $n/2$ pairs.[11] By construction, the degree of every node increases by η in expectation and the maximum degree is now $d + 1$.

Theorem 3. *Consider the class $G(a, A, m, d, \rho)$ of social networks and an associated $R(\eta)$ rewiring. 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:*

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

This result implies that small worlds exhibit wisdom *on average* across rewiring.[12] The intuition behind this result is that even though long-range rewiring has a dramatic effect on average path length (as shown in Watts and Strogatz (1998)) 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.

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.

---

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

[12]Note, that we take the expectation both over seed sets and rewirings. 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.

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

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.[13] Thus, our resulting empirical networks are undirected.[14] 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

[13]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.

[14]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.

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 substantial heterogeneity. On average, the 25th percentile village experiences 33% information loss, while the 75th percentile village experiences only 13% information loss.

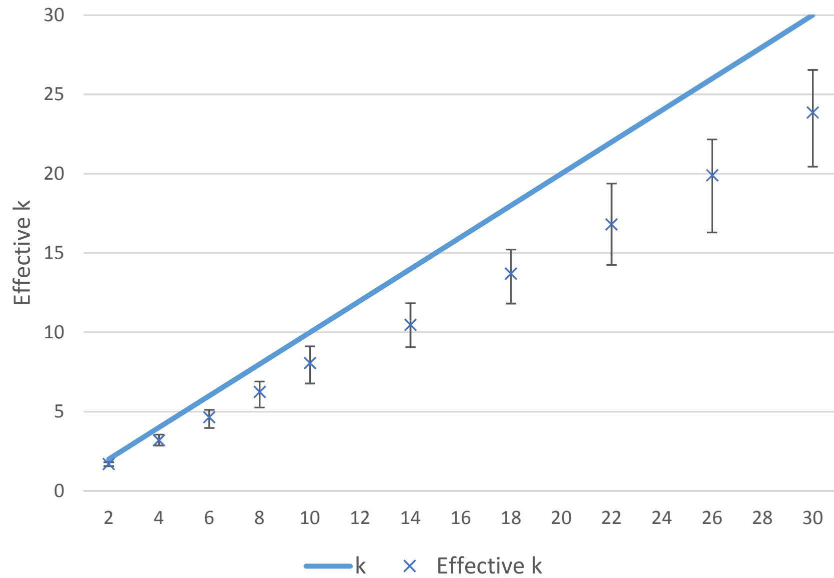


FIGURE 4. Simulations on 75 Indian village networks. The average is taken over all simulations and all networks. The bars represent the interquartile range across networks, for each k.

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). 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. 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,

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

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 will have far more variance than under a more dispersed allocation.

Let's return to the 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 other 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.

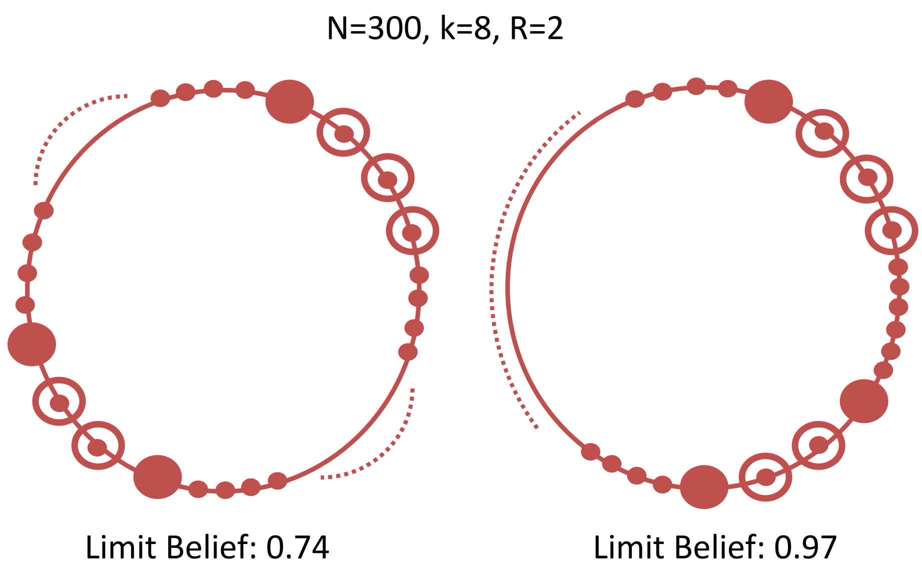


FIGURE 5. Example: Clustered Seeding

Figure 5 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 largely reflects only the outer two signals, that is the two signals with realization 1. Here the limit belief, 0.97, is close to 1.

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

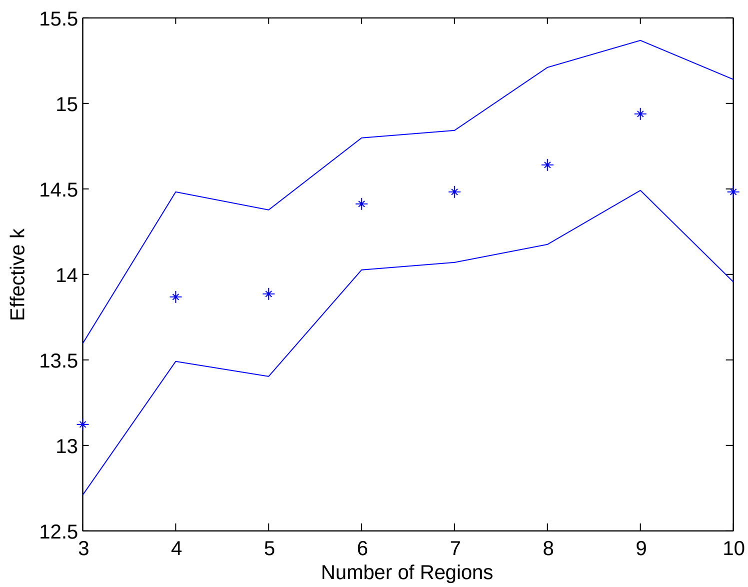


FIGURE 6. Plots of the mean $k^{effective}$ against R, where the average is taken over all simulations and all networks. The solid lines represent the 5th and 95th percentiles of Effective k, bootstrapped across the simulation draws.

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 iid, in Figure 6, we see that this represents a 13pp further decline in effective number of signals on a base of

22pp loss just due to the GDG process. This shows that in empirical networks, when information is not disturbed uniformly at random, the loss can be sizable.

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 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".[15]

[15]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 4.1 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 learnt. 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 DeGroot model. However, we also characterize a large class of networks where this does not happen and our simulations using 75 real world social networks from Indian villages suggest that the outcome may be quite close to full information aggregation even with sparse signals. Finally we observe that the extent of information aggregation depends on the clustering of signals on the network.

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, 14

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2015): "Gossip: Identifying central individuals in a social network," . 1, 4.3.1, 14

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*. 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

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, 4.1, 4.2.1

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," . 2

Molavi, P., A. Tahbaz-Salehi, and A. Jadbabaie (2017): "Foundations of Non-Bayesian Social Learning," *Working Paper*. 3

Mueller-Frank, M. and C. Neri (2013): "Social Learning in Networks: Theory and Experiments," . 2

Rosenblat, T. S. and M. M. Mobius (2004): "Getting Closer or Drifting Apart?*," *The Quarterly Journal of Economics*, 119, 971–1009. A.5

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

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 $j \in S$.

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 $|N(j)|+1$. 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. [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.

We next prove Theorem 1. Without loss of generality, we assume that all initial opinions of seeds are positive.[16]

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).

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

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.

A.3. **Proof of Proposition 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 $r \to \infty$.

A.4. **Proof of Theorems 2.** Our proof proceeds in three steps. We prove the result for a slightly distinct seeding process first: we assume that every node in the graph independently becomes a seed with probability $\frac{k}{n}$. Hence, the *expected* number of seeds is equal to k. We will show at the end that the result extends when there are exactly k seeds randomly distributed on the graph.

Step 1: Fix the seed set S. For any seed $i \in S$ we define the *maximal* Voronoi set $\tilde{V}_i$ as the union of the Voronoi set V_i and its immediate boundary. Note, that $\overline{V}_i \subset \tilde{V}_i$. We also define $\tilde{v}_i = \frac{|\tilde{V}_i|}{n}$.

Using Theorem 1 we can upper-bound the variance in the limit belief x^∞ as follows:

$$\mathrm{var}(x^\infty) \leq d^2\sigma^2 \sum_i \tilde{v}_i^2 \tag{A.10}$$

This follows because $\overline{v}_i^* \leq d\overline{v}_i \leq d\tilde{v}_i$. We can upper-bound the link-weighted share of agents in the Voronoi set with $d\overline{v}_i$ because the maximum degree is d by assumption. Therefore, we can focus on finding a bound for $\sum_i \tilde{v}_i^2$.

Now consider the following thought experiment: draw two random nodes z and z' and consider the event that both of them are in the same maximal Voronoi set. We define the random indicator variable $I_{zz'}$ which equals 1 iff z and z' are in the same maximal Voronoi set.

Lemma 2. *The following holds:*

$$\text{var}(x^\infty) \le 2d^2\sigma^2 \mathrm{E}_{z,z'}\left(I_{zz'}\right) \tag{A.11}$$

Proof. We first note that the probability that both points are in a specific Voronoi set $\tilde{V}_i$ equals $\tilde{v}_i^2$. However, the probability that both points are in *some* Voronoi set is not simply $\sum_i \tilde{v}_i^2$ because the maximal Voronoi sets overlap at their boundaries. For any two distinct seeds i and j denote the pair-wise overlap of the two associated maximal Voronoi sets with $\tilde{V}_{ij} = \tilde{V}_i \bigcap \tilde{V}_i$ and $\tilde{v}_{ij} = \frac{|\tilde{V}_{ij}|}{n}$. We then have:

$$\mathrm{E}_{z,z'}\left(I_{zz'}\right) = \sum_i \tilde{v}_i^2 - \sum_{i\neq j} \tilde{v}_{ij}^2 \tag{A.12}$$

From this we obtain:

$$\begin{aligned} \sum_i \tilde{v}_i^2 &= \mathrm{E}_{z,z'}\left(I_{zz'}\right) + \sum_{i\neq j} \tilde{v}_{ij}^2 \\ &\le 2\mathrm{E}_{z,z'}\left(I_{zz'}\right) \end{aligned} \tag{A.13}$$

This proves the lemma. □

Lemma 2 provides us with the following upper bound for the expected variance in the limit belief:

$$\mathrm{E}_S\left[\text{var}(x^\infty)\right] \le 2d^2\sigma^2 \mathrm{E}_S\left[\mathrm{E}_{z,z'}\left(I_{zz'}\right)\right] = 2d^2\sigma^2 \mathrm{E}_{z,z'}\left[\mathrm{E}_S\left(I_{zz'}\right)\right] \tag{A.14}$$

Note, that we are changing the order of summation to obtain the right-hand side equality. This is a key step in the proof because it allows us to focus on first bounding $\mathrm{E}_S\left(I_{zz'}\right)$ which is the probability that two specific points z and z' are in the same Voronoi set (when taking the expectation over all seed sets).[17] As we will see in Step 2 this probability can be bounded from above using a simple geometric argument.

Step 2: In this step we fix z. We denote the distance between z and z' with $d(z, z')$ and attempt to upper-bound the following expectation:

$$\frac{1}{n} \sum_{z' | d(z,z') \le r_{max}} \mathrm{E}_S\left(I_{zz'}\right) \tag{A.15}$$

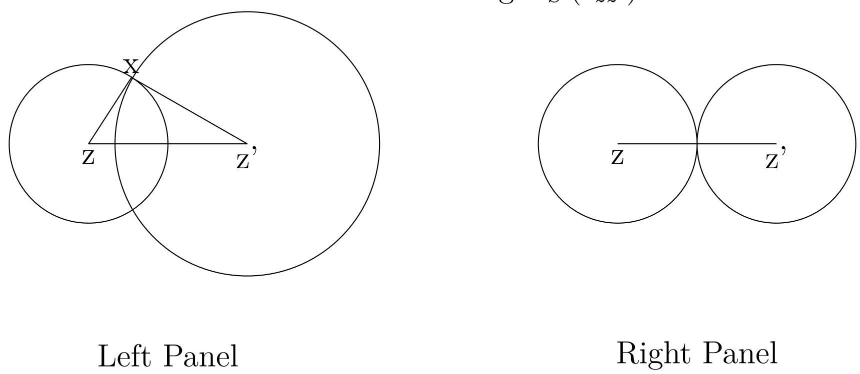
We prove the following lemma:

Lemma 3. *There is a constant C such that:*

$$\frac{1}{n} \sum_{z' | d(z,z') \le r_{max}} \mathrm{E}_S\left(I_{zz'}\right) \le \frac{C}{k} \tag{A.16}$$

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

Figure 7. Bounding $E_S(I_{zz'})$



Proof. We first prove a mini-lemma: consider two points x and x' and assume that $d(z, x') < d(z, x) - 2$. Then z cannot be part of the maximal Voronoi set $\tilde{V}_x$. Recall, that the maximal Voronoi set includes all the closest (and equidistant points) plus a boundary. Therefore, a point z that belongs to Voronoi set $\tilde{V}_x$ can be closer to a distinct seed x' but by at most a difference in length of 2.

Now fix a point z and consider the second point z' at distance $r = d(z, z') = 1, 2..$ from z. We want to bound the probability of the event $I_{zz'}$ where both points lie in the same Voronoi set (taking the expectation over all seed assignments). If z and z' are in the same maximal Voronoi set then there must be a seed x such that $z, z' \in \tilde{V}_x$. Consider the ball $B_z(d(z, x)-3)$ as indicated in the left panel of Figure 7: there can be no other seed x' inside this ball because otherwise z would not belong to the maximal Voronoi set $\tilde{V}_x$ (by our mini-lemma above). Similarly, the ball $B_{z'}(d(z', x)-3)$ cannot contain any seeds.

This implies that at least $a\left[(d(z, x) - 3)^m + (d(z', x) - 3)^m\right]$ nodes cannot contain seeds. According to the triangle inequality we also have:

$$d(z, x) + d(z', x) \geq d(z, z') = r \tag{A.17}$$

Due to the convexity of the polynomial function r^m we can therefore deduce:

$$a\left[(d(z, x) - 3)^m + F(d(z', x) - 3)^m\right] \geq 2a(r/2 - 3)^m \tag{A.18}$$

The right panel of Figure 7 illustrates the simple geometric intuition for this inequality: the two tangential, equal-sized discs always cover a smaller area that the two discs on the left panel.

We can now complete the proof:

$$\begin{aligned}\frac{1}{n}\sum_{z'|d(z,z')\leq r_{max}} \mathrm{E}_S\left(I_{zz'}\right) &= \frac{1}{n}\sum_{r=1}^{r_{max}}\sum_{z'|d(z,z')=r} \mathrm{E}_S\left(I_{zz'}\right)\\ &\leq \frac{1}{n}\sum_{r=1}^{r_{max}}\sum_{z'|d(z,z')=r}\left(1-\frac{k}{n}\right)^{2a(r/2-3)^m} \qquad \text{(A.19)}\end{aligned}$$

The inequality follows because whenever z and z' are in the same Voronoi set then at least $2a(r/2-3)^m$ nodes cannot contain seeds.

Since $\left(1-\frac{k}{n}\right)^n \leq \exp(-k)$ we obtain:

$$\frac{1}{n}\sum_{z'|d(z,z')\leq r_{max}} \mathrm{E}_S\left(I_{zz'}\right) \leq \frac{1}{n}\sum_{r=1}^{r_{max}}\sum_{z'|d(z,z')=r}\exp\left(-k\frac{2a(r/2-3)^m}{n}\right) \qquad \text{(A.20)}$$

There are at most $AF(r)$ such points at distance r or less. Moreover, the function $\exp\left(-k\frac{2a(r/2-3)^m}{n}\right)$ is decreasing in r. Therefore, the left-hand side will be maximized if there are exactly $A(r^m-(r-1)^m)$ nodes at distance r:

$$\frac{1}{n}\sum_{z'|d(z,z')\leq r_{max}} \mathrm{E}_S\left(I_{zz'}\right) \leq \frac{1}{n}\sum_{r=1}^{r_{max}}\exp\left(-k\frac{2a(r/2-3)^m}{n}\right)A\left[r^m-(r-1)^m\right]$$

Using the mean-value theorem and the fact that r^m is convex we know that $r^m-(r-1)^m \leq mr^{m-1}$:

$$\frac{1}{n}\sum_{z'|d(z,z')\leq r_{max}} \mathrm{E}_S\left(I_{zz'}\right) \leq \frac{1}{n}\sum_{r=1}^{r_{max}}A\exp\left(-k\frac{2a(r/2-3)^m}{n}\right)mr^{m-1} \qquad \text{(A.21)}$$

Next, note that $r^{m-1}\leq C\frac{1}{2}(r/2-3)^{m-1}$ for some constant $C>0$. We therefore get (for some constant C'):

$$\begin{aligned}\frac{1}{n}\sum_{z'|d(z,z')\leq r_{max}} \mathrm{E}_S\left(I_{zz'}\right) &\leq \frac{1}{n}\sum_{r=1}^{r_{max}}AC\exp\left(-k\frac{2a(r/2-3)^m}{n}\right)m\frac{1}{2}(r/2-3)^{m-1}\\ &\leq AC'\int_0^\infty \exp(-2akx)dx\end{aligned}$$

For the last step we approximate the infinite sum with the corresponding integral and use the chain rule. This finally gives us:

$$\frac{1}{n}\sum_{z'|d(z,z')\leq r_{max}} \mathrm{E}_S\left(I_{zz'}\right) \leq \frac{AC'}{2ak} \qquad \text{(A.22)}$$

This proves our lemma. □

Step 3: By combining the previous 2 steps we obtain:

$$\mathrm{E}_S\left[\mathrm{var}(x^\infty)\right] \leq 2d^2\sigma^2\frac{1}{\rho}\frac{AC'}{2ak}$$

The factor $\frac{1}{\rho}$ enters because we have only focused on the case where z and z' are at most r_max apart (and hence $B_z(r_max)$ covers just more than a share ρ of the network). Take any slice size ρn which does not intersect with $B_z(r_max)$ (of which there are at most $1/\rho$) – then one can see the probability that z and z' are in the same Voronoi set is bounded above by the same bound as for Lemma 3.

This completes the proof of Theorem 2.

A.5. **Proof of Theorems 3.** Our proof proceeds in four steps.

Step 1: We replicate the proof of step 1 of Theorem 2. We again focus on maximal Voronoi sets and we take into account that the maximum degree in the rewired graph is $d+1$ instead of d. We therefore obtain:

$$\mathrm{E}_{S,R(\eta)}\left[\mathrm{var}(x^\infty)\right] \leq 2(d+1)^2\sigma^2\mathrm{E}_{z,z'}\left[\mathrm{E}_{S,R(\eta)}\left(I_{zz'}\right)\right] \tag{A.23}$$

Step 2: We next show that the probability that two random points z and z' are in the same Voronoi set (averaged across all seed assignments with k seeds and all rewirings) is approximately $\frac{1}{k}$ (which proves the theorem).

To see the intuition for this result we consider a fixed k and large n. We construct the r-balls around the k seeds as well as the two points z and z' and let r increase. As soon as the r-ball around z overlaps with any of the r-balls around the seeds, we found the Voronoi set assignment for z and the shortest path to the corresponding seed (same argument for z'). Because n is large, this shortest path involves at least one rewired link with high probability. However, as long as the r-balls around the seeds are of comparable size, there is a similar number of rewired links in any r-ball. Since rewired links connect random points, the conditional probability that any particular rewired link belong to the r-ball around any particular seed has to be approximately $\frac{1}{k}$.

In the following we make this intuition precise through two two sub-steps. First, we show that the r-balls around different grow at approximately the same rate as long as the r-balls around the seeds abd the points z and z' have volume less than $\sqrt{n}$. Second, we show that as soon as the balls reach size $\sqrt{n}$ they will intersect with high probability and each of the two points z and z' will intersect with the r-balls around the seeds with approximately equal probability.

Step 2.1: We start with a lemma that show that all of these $k+2$ r-balls grow at similar rates.

Lemma 4. *Consider the class $G(a, A, m, d, \rho)$ of social networks and an associated $R(\eta)$ rewiring. Then there are positive constants C_1 to C_4 and θ such that for any graph $g \in G(a, A, m, d, \rho)$ and any node $z \in g$ we have*

$$C_1 \exp(C_3 r) \leq E_{R(\eta)} |B_r(z)| \leq C_2 \exp(C_3 r)$$

with probability greater than $\theta > 0$ for $r \leq \frac{\ln(n)}{2C_3}$.

Hence, the balls expand at an exponential rate and the relative size of balls around different two different nodes will node exceed C_3/C_1 (ratio of larger to smaller ball) with probability that's bounded away from 0 as long as the balls do not exceed size $o(\sqrt{n})$. The proof follows readily from Rosenblat and Mobius (2004). The key intuition is that (a) as the balls grow the growth rates become less noisy and (b) the balls have volume less than $\sqrt{n}$ and hence the extent of overlap can be made as small as desired because only a share $\frac{(k+2)\sqrt{n}}{n}$ of nodes belongs to some r-ball.

Step 2.1: We now show that even though the r-balls have volume less than $\sqrt{n}$ for $r \leq \frac{\ln(n)}{2C_3}$ the probability that they intersect *somewhere* becomes large *exactly* when the balls reach size $\sqrt{n}$.

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:

$$\left(1 - \eta \frac{D_x \exp(C_3 r)}{n}\right)^{D_z \exp(C_3 r)} = \left(1 - \frac{\eta}{\frac{n}{D_x \exp(C_3 r)}}\right)^{\frac{n}{D_x \exp(C_3 r)} \frac{D_x D_z \exp(2C_3 r)}{n}} \tag{A.24}$$

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

$$\exp\left(-\eta \frac{D_x D_z \exp(2C_3 r)}{n}\right) \tag{A.25}$$

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.

Therefore, node z will connect with the closest seed as soon as the ball around z reaches size $O(\sqrt{n})$. 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.