### Citation

On Krause's Multi-Agent Consensus Model With State-Dependent Connectivity
Automatic Control, IEEE Transactions on
Volume 54, Issue 11, Nov. 2009 Page(s):2586 - 2597

### As Published

http://dx.doi.org/10.1109/TAC.2009.2031211

### Publisher

Institute of Electrical and Electronics Engineers

### Version

Author's final manuscript

### Accessed

Tue Dec 11 22:47:41 EST 2018

### Citable Link

http://hdl.handle.net/1721.1/51690

### Terms of Use

Attribution-Noncommercial-Share Alike 3.0 Unported

### Detailed Terms

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
On Krause’s multi-agent consensus model with state-dependent connectivity (Extended version)

Vincent D. Blondel, Julien M. Hendrickx and John N. Tsitsiklis

Abstract—We study a model of opinion dynamics introduced by Krause: each agent has an opinion represented by a real number, and updates its opinion by averaging all agent opinions that differ from its own by less than 1. We give a new proof of convergence into clusters of agents, with all agents in the same cluster holding the same opinion. We then introduce a particular notion of equilibrium stability and provide lower bounds on the inter-cluster distances at a stable equilibrium. To better understand the behavior of the system when the number of agents is large, we also introduce and study a variant involving a continuum of agents, obtaining partial convergence results and lower bounds on inter-cluster distances, under some mild assumptions.

Keywords: Multi-agent system, consensus, opinion dynamics, decentralized control.

I. INTRODUCTION

There has been an increasing interest in recent years in the study of multi-agent systems where agents interact according to simple local rules, resulting in a possibly coordinated global behavior. In a prominent paradigm dating back to [10] and [28], each agent maintains a value which it updates by taking a linear, and usually convex combination of other agents’ values; see e.g., [5], [16], [17], [25], [28], and [26], [27] for surveys. The interactions between agents are generally not all-to-all, but are described by an interconnection topology. In some applications, this topology is fixed, but several studies consider the more intriguing case of changing topologies. For example, in Vicsek’s swarming model [30], animals are modeled as agents that move on the two-dimensional plane. All agents have the same speed but possibly different headings, and at each time-step they update their headings by averaging the headings of those agents that are sufficiently close to them. When the topology depends on the combination of the agent states, as in Vicsek’s model, an analysis that takes this dependence into account can be difficult. For this reason, the sequence of topologies is often treated as exogenous (see e.g. [4], [17], [25]), with a few notable exceptions [7], [8], [18]. For instance, the authors of [7] consider a variation of the model studied in [17], in which communications are all-to-all, but with the relative importance given by one agent to another weighted by the distance separating the agents. They provide conditions under which the agent headings converge to a common value and the distance between any two agents converges to a constant. The same authors relax the all-to-all assumption in [8], and study communications restricted to arbitrarily changing but connected topologies.

We consider here a simple discrete-time system involving endogenously changing topologies, and analyze it while taking explicitly into account the dependence of the topology on the system state. The discrete-agent model is as follows. There are $n$ agents, and every agent $i$ ($i = 1, \ldots, n$), maintains a real value $x_i$. These values are synchronously updated according to

$$x_i(t + 1) = \frac{\sum_{j:|x_i(t) - x_j(t)| < 1} x_j(t)}{\sum_{j:|x_i(t) - x_j(t)| < 1}}.$$  

(1)

Two agents $i, j$ for which $|x_i(t) - x_j(t)| < 1$ are said to be neighbors or connected (at time $t$). Note that with this definition, an agent is always its own neighbor. Thus, in this model, each agent updates its value by computing the average of the values of its neighbors. In the sequel, we usually refer to the agent values as “opinions,” and sometimes as “positions.”

The model (1) was introduced by Krause [19] to capture the dynamics of opinion formation. Values represent opinions on some subject, and an agent considers another agent as “reasonable” if their opinions differ by less than 1. Each agent thus updates its opinion by computing the average of the opinions it finds “reasonable”. This system is also sometimes referred to as the Hegselmann-Krause model, following [14]. It has been abundantly studied in the literature [19], [20], [22], [23], and displays some peculiar properties that have remained unexplained. For example, it has been experimentally observed that opinions initially uniformly distributed on an interval tend to converge to clusters of opinions separated by a distance slightly larger than 2, as shown in Figure 1. In contrast, presently available results can only prove convergence to clusters separated by at least 1. An explanation of the inter-cluster distances observed for this system, or a proof of a nontrivial lower bound is not available.

Inter-cluster distances larger than the interaction radius (which in our case was set to 1) have also been observed by Defuant et al. [9] for a related stochastic model, often referred to as the Defuant-Weisbuch model. In that model, two randomly selected agents update their opinions at any

$^1$In Krause’s initial formulation, all opinions belong to $[0, 1]$, and an agent considers another one as reasonable if their opinions differ by less than a pre-defined parameter $\epsilon$. 

This research was supported by the National Science Foundation under grant ECCS-0701623, by the Concerted Research Action (ARC) “Large Graphs and Networks” of the French Community of Belgium and by the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office. The scientific responsibility rests with its authors. Julien Hendrickx holds a F.R.S.-FNRS postdoctoral fellowship (Belgian Fund for Scientific Research). V. D. Blondel and J. M. Hendrickx are with Department of Mathematical Engineering, Université catholique de Louvain, Avenue Georges Lemaître 4, B-1348 Louvain-la-Neuve, Belgium: vincent.blondel@uclouvain.be, julien.hendrickx@uclouvain.be. J. N. Tsitsiklis is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139, USA; jnt@mit.edu.
an equilibrium is perturbed by introducing an additional agent, their positions. This ensures that an initially connected set of agents, we introduce and study a variation of the model, of agents. To better understand the case of a large numbers for the model (1). We then introduce a particular notion of moves, its new position is a convex combination of its previous converging to a stable equilibrium increases with the number of agents, and examining the partial differential equation describing the evolution of this density [2], [3]. Other models, involving either discrete or continuous time, and finitely or infinitely many agents, have also been proposed [1], [12], [29]. For a survey, see for example [24].

Thus, the Krause and Deffuant-Weisbuch models rely on the average of its previous opinion and that of the other agent. Each agent moves to a new opinion which is a weighted given time step. If their opinions differ by more than a certain threshold, their opinions remain unchanged; otherwise, each agent moves to a new opinion which is a weighted average of its previous opinion and that of the other agent. This is similar, including inter-cluster distances significantly larger than the interaction radius. The behavior of the Deffuant-Weisbuch model — and in particular the final positions of the clusters — has also been studied by considering a continuous density approximating the discrete distribution of agents, and examining the partial differential equation describing the evolution of this density [2], [3].

The model that we consider also has similarities with certain rendezvous algorithms (see, e.g., [21]) in which the objective is to have all agents meet at a single point. Agents are considered neighbors if their positions are within a given radius $R$. The update rules satisfy two conditions. First, when an agent moves, its new position is a convex combination of its previous position and the positions of its neighbors. Second, if two agents are neighbors, they remain neighbors after updating their positions. This ensures that an initially connected set of agents is never split into smaller groups, so that all agents can indeed converge to the same point.

In this paper, we start with a simple convergence proof for the model (1). We then introduce a particular notion of equilibrium stability, involving a robustness requirement when an equilibrium is perturbed by introducing an additional agent, and prove that an equilibrium is stable if and only if all inter-cluster distances are above a certain nontrivial lower bound. We observe experimentally that the probability of converging to a stable equilibrium increases with the number of agents. To better understand the case of a large numbers of agents, we introduce and study a variation of the model, which involves a continuum of agents (the “continuous-agent” model). We give partial convergence results and provide a lower bound on the inter-cluster distances at equilibrium, under some regularity assumptions. We also show that for a large number of discrete agents, the behavior of the discrete-agent model indeed approximates the continuous-agent model.

Our continuous-agent model, first introduced in [6], is obtained by indexing the agents by a real number instead of an integer. It is equivalent to the so-called “discrete-time density based Hegselmann-Krause model” proposed independently in [24], which is in turn similar to a model presented in [12] in a continuous-time setup. Furthermore, our model can also be viewed as the limit, as the number of discrete opinions tends to infinity, of the “interactive Markov chain model” introduced by Lorenz [23]; in the latter model, there is a continuous distribution of agents, but the opinions take values in a discrete set.

We provide an analysis of the discrete-agent model (1) in Section II. We then consider the continuous-agent model in Section III. We study the relation between these two models in Section IV, and we end with concluding remarks and open questions, in Section V.

II. THE DISCRETE-AGENT MODEL

A. Basic properties and convergence

We begin with a presentation of certain basic properties of the discrete-agent model (1), most of which have already been proved in [14], [20], [22].

Proposition 1 (Lemma 2 in [20]): Let $(x(t))$ be a sequence of vectors in $\mathbb{R}^n$ evolving according to (1). The order of opinions is preserved: if $x_i(t_0) \leq x_j(t_0)$, then $x_i(t) \leq x_j(t)$ for all $t$.

Proof: We use induction. Suppose that $x_i(t) \leq x_j(t)$. Let $N_i(t)$ be the set of agents connected to $i$ and not to $j$, $N_j(t)$ the set of agents connected to $j$ and not to $i$, and $N_{ij}(t)$ the set of agents connected to both $i$ and $j$, at time $t$. We assume here that these sets are nonempty, but our argument can easily be adapted if some of them are empty. For any $k_1 \in N_i(t)$, $k_2 \in N_{ij}(t)$, and $k_3 \in N_j(t)$, we have $x_{k_1}(t) \leq x_{k_2}(t) \leq x_{k_3}(t)$. Therefore, $\bar{x}_{N_i} \leq \bar{x}_{N_{ij}} \leq \bar{x}_{N_j}$, where $\bar{x}_{N_i}, \bar{x}_{N_{ij}}, \bar{x}_{N_j}$, respectively, is the average of $x_k(t)$ for $k$ in the corresponding set. It follows from (1) that

$$x_i(t + 1) = \frac{|N_{ij}| \bar{x}_{N_{ij}} + |N_i| \bar{x}_{N_i}}{|N_{ij}| + |N_i|} \leq \bar{x}_{N_{ij}},$$

and

$$x_j(t + 1) = \frac{|N_{ij}| \bar{x}_{N_{ij}} + |N_j| \bar{x}_{N_j}}{|N_{ij}| + |N_j|} \geq \bar{x}_{N_{ij}},$$

where we use $|A|$ to denote the cardinality of a set $A$.

In light of this result, we will assume in the sequel, without loss of generality, that the initial opinions are sorted: if $i < j$ then $x_i(t) \leq x_j(t)$. The next Proposition follows immediately from the definition of the model.

Proposition 2: Let $(x(t))$ be a sequence of vectors in $\mathbb{R}^n$ evolving according to (1), and such that $x(0)$ is sorted, i.e., if $i < j$, then $x_i(0) \leq x_j(0)$. The smallest opinion $x_1$ is nondecreasing with time, and the largest opinion $x_n$ is nonincreasing with time. Moreover, if at some time the distance between two consecutive agent opinions $x_i(t)$ and
$x_{i+1}(t)$ is larger than or equal to 1 it remains so for all subsequent times $t' \geq t$, so that the system can then be decomposed into two independent subsystems containing the agents $1, \ldots, i$, and $i+1, \ldots, n$ respectively.

Note that unlike other related models as the Deffuant-Weisbusch model [9] or the continuous-time model in [15], the average of the opinions is not necessarily preserved, and the “variance” (sum of squared differences from the average) may occasionally increase. See [15] for examples with three and eight agents respectively. The convergence of (1) has already been established in the literature (see [11], [22]), and is also easily deduced from the convergence results for the case of exogenously determined connectivity sequences (see e.g., [5], [16], [22], [25]), an approach that extends to the case of higher-dimensional opinions. We present here a simple alternative proof, which exploits the particular dynamics we are dealing with.

**Theorem 1:** If $x(t)$ evolves according to (1), then for every $i$, $x_i(t)$ converges to a limit $x_i^*$ in finite time. Moreover, for any $i, j$, we have either $x_i^* = x_j^*$ or $|x_i^* - x_j^*| \geq 1$.

**Proof:** Since $x(0)$ is assumed to be sorted, the opinion $x_1$ is nondecreasing and bounded above by $x_n(0)$. As a result, it converges to a value $x_1^*$. Let $p$ be the highest index for which $x_p$ converges to $x_1^*$.

We claim that if $p < n$, there is a time $t$ such that $x_{p+1}(t) - x_p(t) \geq 1$. Suppose, to obtain a contradiction, that the claim does not hold, i.e., that $x_{p+1}(t) - x_p(t)$ is always smaller than 1. Fix some $\epsilon > 0$ and a time after which the distance of $x_i$ from $x_1^*$, for $i = 1, \ldots, p$, is less than $\epsilon$. Since $x_{p+1}$ does not converge to $x_1^*$, there is a further time at which $x_{p+1}$ is larger than $x_1^* + \delta$ for some $\delta > 0$. For such a time $t$, $x_p(t+1)$ is at least

$$\frac{1}{p+1}\sum_{i=1}^{p+1} x_i(t) \geq \frac{1}{p+1} (p(x_1^* - \epsilon) + (x_1^* + \delta)),$$

which is larger than $x_1^* + \epsilon$ if $\epsilon$ is chosen sufficiently small. This however contradicts the requirement that $x_p$ remain within $\epsilon$ from $x_1^*$. This contradiction shows that there exists a time $t$ at which $x_{p+1}(t) - x_p(t) \geq 1$. Subsequent to that time, using also Proposition 2, $x_p$ cannot increase and $x_{p+1}$ cannot decrease, so that the inequality $x_{p+1} - x_p \geq 1$ continues to hold forever. In particular, agents $1, \ldots, p$ will no more interact with the remaining agents. Thus, if $p < n$, there will be some finite time after which the agents $p+1, \ldots, n$ behave as an independent system, to which we can apply the same argument. Continuing recursively, this establishes the convergence of all opinions to limiting values that are separated by at least 1.

It remains to prove that convergence takes place in finite time. Consider the set of agents converging to a particular limiting value. It follows from the argument above that there is a time after which none of them is connected to any agent outside that set. Moreover, since they converge to a common value, they eventually get sufficiently close so that they are all connected to each other. When this happens, they all compute the same average, reach the same opinion at the next time step, and keep this opinion for all subsequent times. Thus, they converge in finite time. Finite time convergence for the entire systems follows because the number of agents is finite.

We will refer to the limiting values to which opinions converge as *clusters*. With some abuse of terminology, we will also refer to a set of agents whose opinions converge to a common value as a cluster.

It can be shown that the convergence time is bounded above by some constant $c(n)$ that depends only on $n$. On the other hand, an upper bound that is independent of $n$ is not possible, even if all agent opinions lie in the interval $[0, L]$ for a fixed $L$. To see this, consider $n$ agents, with $n$ odd, one agent initially placed at 1, and $(n-1)/2$ agents initially placed at 0.1 and 1.9. All agents will converge to a single cluster at 1, but the convergence time increases to infinity as $n$ grows.

We note that the convergence result in Theorem 1 does not hold if we consider the same model but with a countable number of agents. Indeed, consider a countably infinite number of agents, all with positive initial opinions. Let $m(y)$ be the number of agents having an initial opinion $y$. Suppose that $\alpha \in (1/2, 1)$, and consider an initial condition for which $m(0) = 0$, $m(\alpha) = 1$, $m(\alpha(k+1)) = m(\alpha k) + 3m(\alpha(k-1))$ for every integer $k > 1$, and $m(y) = 0$ for every other value of $y$. Then, the update rule (1) implies that $x_i(t+1) = x_i(t) + \alpha/2$, for every agent $i$ and time $t$, and convergence fails to hold. A countable number of agents also admits equilibria where the limiting values are separated by less than 1. An example of such an equilibrium is obtained by considering one agent at every integer multiple of $1/2$.

We also note that equilibria in which clusters are separated by less than 1 become possible when opinions are elements of a manifold, instead of the real line. For example, suppose that opinions belong to $[0, 2\pi]$ (identified with elements of the unit circle), and that two agents are neighbors if and only if $|x_i - x_j (mod 2\pi)| < 1$. If every agent updates its angle by moving to the average of its neighbors’ angles, it can be seen that an initial configuration with $n$ agents located at angles $2\pi k/n$, $k = 0, \ldots, n-1$, is an equilibrium. Moreover, more complex equilibria also exist. Convergence has been experimentally observed for models of this type, but no proof is available.

**B. Experimental observations**

Theorem 1 states that opinions converge to clusters separated by at least 1. Since the smallest and largest opinions are nondecreasing and nonincreasing, respectively, it follows that opinions initially confined to an interval of length $L$ can converge to at most $\lceil L \rceil + 1$ clusters. It has however been observed in the literature that the distances between clusters are usually significantly larger than 1 (see [20], [23], and Figure 1), resulting in a number of clusters that is significantly smaller than the upper bound of $\lceil L \rceil + 1$. To further study this phenomenon, we analyze below different experimental results, similar to those in [23].

Figure 2 shows the dependence on $L$ of the cluster number and positions, for the case of a large number of agents and initial opinions that are uniformly spaced on an interval of length $L$. Such incremental analyses also appear in the
literature for various similar systems [2], [13], [23], [24]. We see that the cluster positions tend to change with $L$ in a piecewise continuous (and sometimes linear) manner. The discontinuities correspond to the emergence of new clusters, or to the splitting of a cluster into two smaller ones. The number of clusters tends to increase linearly with $L$, with a coefficient slightly smaller than $1/2$, corresponding to an inter-cluster distance slightly larger than 2. Note however that this evolution is more complex than it may appear. Irregularities in the distance between clusters and in their weights can be observed for growing $L$, as already noted in [23]. Besides, for larger scale simulations ($L = 1000, n = 10^6$), a small proportion of clusters take much larger or much smaller weights than the others, and some inter-cluster distances are as large as 4 or as small as 1.5. These irregularities could be inherent to the model, but may also be the result of the particular discretization chosen or of the accumulation of numerical errors in a discontinuous system.

Because no nontrivial lower bound is available to explain the observed inter-cluster distances in Krause’s model, we start with three observations that can lead to some partial understanding. In fact, the last observation will lead us to a formal stability analysis, to be developed in the next subsection.

(a) We observe from Figure 2 that the minimal value of $L$ that leads to multiple clusters is approximately 5.1, while Theorem 1 only requires that this value be at least 1. This motivates us to address the question of whether a more accurate bound can be derived analytically. Suppose that there is an odd number of agents whose initial opinions are uniformly spaced on $[0, L]$. An explicit calculation shows that all opinions belong to an interval $[\frac{1}{2} - O(\frac{1}{n}), L - \frac{1}{2} + O(\frac{1}{n})]$ after one iteration, and to an interval $[\frac{1}{2} - O(\frac{1}{n}), L - \frac{1}{2} + O(\frac{1}{n})]$ after two iterations. Furthermore, by Proposition 2, all opinions must subsequently remain inside these intervals. On the other hand, note that with an odd number of agents, there is one agent that always stays at $L/2$. Thus, if all opinions eventually enter the interval $(L/2 - 1, L/2 + 1)$, then there can only be a single cluster. This implies that there will be a single cluster if $L - \frac{11}{12} + O(\frac{1}{n}) < L/2 + 1$, that is, if $L < \frac{23}{6} - O(\frac{1}{n}) \simeq 3.833$. This bound is smaller than the experimentally observed value of about 5.1. It can be further improved by carrying out explicit calculations of the smallest position after a further number of iterations. Also, as long as the number of agents is sufficiently large, a similar analysis is possible if the number of agents is even, or in the presence of random initial opinions.

(b) When $L$ is sufficiently large, Figure 2 shows that the position of the leftmost clusters becomes independent of $L$. This can be explained by analyzing the propagation of information: at each iteration, an agent is only influenced by those opinions within distance 1 of its own, and its opinion is modified by less than 1. So, information is propagated by at most a distance 2 at every iteration. For the case of uniformly spaced initial opinions on $[0, L]$, with $L$ large, the agents with initial opinions close to 0 behave, at least in the first iterations, as if opinions were initially distributed uniformly on $[0, +\infty)$. Moreover, once a group of opinions is separated from other opinions by more than 1, this group becomes decoupled. Therefore, if the agents with initial opinions close to 0 become separated from the remaining agents in finite time, their evolution under a uniform initial distribution on $[0, L]$ for a sufficiently large $L$ is the same as in the case of a uniform initial distribution on $[0, +\infty)$.

We performed simulations with initial opinions uniformly spaced on $[0, \infty)$, as in [23]. We found that every agent eventually becomes connected with a finite number of agents and disconnected from the remaining agents. The groups formed then behave independently and converge to clusters. As shown in Figure 3, the distances between two consecutive clusters are close to 2.2. These distances partially explain the evolution of the number of clusters (as a function of $L$) shown in Figure 2. However, a proof of these observed properties is not available, and it is unclear whether the successive inter-cluster distances possess some regularity or convergence properties.

(c) A last observation that leads to a better understanding of the size of the inter-cluster distances is the following. Suppose
that \( L \) is just below the value at which two clusters are formed, and note the special nature of the resulting evolution, shown in Figure 4. The system first converges to a “meta-stable state” in which there are two groups, separated by a distance slightly larger than 1, and which therefore do not interact directly with each other. The two groups are however slowly attracted by some isolated agents located in between; furthermore, these isolated agents are being pulled by both of these groups and remain at the weighted average of the opinions in the two groups. Eventually, the distance between the two groups becomes smaller than 1, the two groups start attracting each other directly, and merge into a single cluster. (This corresponds to one of the slow convergence phenomena observed in [23].) The initial convergence towards a two-cluster equilibrium is thus impossible by the presence of a few agents in system (1), which is a measure of the distance between the original and perturbed equilibria. We say that \( \bar{x} \) is stable if \( \sup_{\delta} \Delta_{\bar{x}_{0},\delta} \), the supremum of distances between initial and perturbed equilibria caused by a perturbing agent of given weight \( \delta \), converges to zero as \( \delta \) vanishes. Equivalently, an equilibrium is unstable if a substantial change in the equilibrium can be induced by a perturbing agent of arbitrarily small weight.

**Theorem 2:** An equilibrium is stable if and only if for any two clusters \( A \) and \( B \) with weights \( W_A \) and \( W_B \), respectively, the following holds: either \( W_A = W_B \) and the inter-cluster distance is greater than or equal to 2; or \( W_A \neq W_B \) and the inter-cluster distance is strictly greater than \( 1 + \frac{\min(W_A,W_B)}{\max(W_A,W_B)} \).

(Note that the two cases are consistent, except that the second involves a strict inequality.)

**Proof:** We start with an interpretation of the strict inequality in the statement of the theorem. Consider two clusters \( A \) and \( B \), at positions \( x_A \) and \( x_B \), and let \( m = (W_A x_A + W_B x_B)/(W_A + W_B) \), which is their center of mass. Then, an easy calculation shows that

\[
|x_A - x_B| > 1 + \frac{\min(W_A,W_B)}{\max(W_A,W_B)}
\]

if and only if

\[
\max\{|m - x_A|, |m - x_B|\} > 1
\]

Suppose that an equilibrium \( \bar{x}_{0} \) satisfies the conditions in the theorem. We will show that \( \bar{x}_{0} \) is stable. Let us insert a perturbing agent of weight \( \delta \). Note that since \( \bar{x}_{0} \) is an equilibrium, and therefore the clusters are at least 1 apart, the perturbing agent is connected to at most two clusters. If this agent is disconnected from all clusters, it has no influence, and \( \Delta_{\bar{x}_{0},\delta} = 0 \). If it is connected to exactly one cluster \( A \), with position \( x_A \) and weight \( W_A \), the system reaches a new equilibrium after one time step, where both the perturbing agent and the cluster have an opinion \((\bar{x}_{0} + x_A)/(\delta + W_A)\). Then,

\[
\Delta_{\bar{x}_{0},\delta} = |\bar{x}_{0} - x_A| \cdot \frac{\delta}{\delta + W_A} \leq \frac{\delta}{\delta + W_A},
\]

which converges to 0 as \( \delta \to 0 \). Suppose finally that the perturbing agent is connected to two clusters \( A, B \). This implies that the distance between these two clusters is less than 2, and since \( \bar{x}_{0} \) satisfies the conditions in the theorem, it must be greater than \( 1 + \frac{\min(W_A,W_B)}{\max(W_A,W_B)} \). Therefore, using (3), the distance of one these clusters from their center of mass \( m \) is greater than 1. The opinion of the perturbed agent after one iteration is within \( O(\delta) \) from \( m \), while the two clusters only move by an \( O(\delta) \) amount. Since the original distance between
one of the two clusters and $m$ is greater than 1, it follows that after one iteration, and when $\delta$ is sufficiently small, the distance of the perturbing agent from one of the clusters is greater than 1, which brings us back to the case considered earlier, and again implies that $\Delta x_{\alpha, \delta}$ converges to zero as $\delta$ decreases.

To prove the converse, we now suppose that the distance between two clusters $A$ and $B$, at positions $x_A$ and $x_B$, is less than 2, and also less than $1 + \frac{\min(W_A, W_B)}{\max(W_A, W_B)}$. Assuming without loss of generality that $x_A < x_B$, their center of mass $m$ is in the interval $(x_B - 1, x_A + 1)$. Let us fix an $\epsilon > 0$ such that $(m - \epsilon, m + \epsilon) \subseteq (x_B - 1, x_A + 1)$. Suppose that at some time $t$ after the introduction of the perturbing agent we have

$$\tilde{x}_0(t) \in (m(t) - \epsilon, m(t) + \epsilon) \subseteq (x_B(t) - 1, x_A(t) + 1),$$

with $x_B(t) - x_A(t) \geq 1$, where $\tilde{x}_0(t)$, $x_A(t)$, $x_B(t)$, and $m(t)$ represent the positions at time $t$ of the perturbing agent, of the clusters $A$ and $B$, and of their center of mass, respectively. One can easily verify that $x_A(t+1) = x_A(t) + \Theta(\delta)$ and $x_{\tilde{B}}(t+1) = x_B(t) + \Theta(\delta)$, so that $x_B(t+1) - x_A(t+1) < x_B(t) - x_A(t)$, and $(m(t+1) - \epsilon, m(t+1) + \epsilon) \subseteq (x_B(t+1) - 1, x_A(t+1) + 1)$.

Moreover, observe that if $\delta$ were 0, we would have $\tilde{x}_0(t+1) = m(t)$. For $\delta \neq 0$, $\tilde{x}_0(t+1)$ is close to $m(t)$, and we have $\tilde{x}_0(t+1) = m(t) + O(\delta)$. Since

$$m(t+1) = \frac{W_A x_A(t+1) + W_B x_B(t+1)}{W_A + W_B} = m(t) + O(\delta),$$

we obtain $|m(t+1) - m(t)| = O(\delta)$, and therefore $\tilde{x}_0(t+1) \in (m(t+1) - \epsilon, m(t+1) + \epsilon)$, as long as $\delta$ is sufficiently small with respect to $\epsilon$.

We have shown that if $\tilde{x}_0(0) = \tilde{x}_0$ is chosen so that the condition (4) is satisfied for $t = 0$, and if $\delta$ is sufficiently small, the condition (4) remains satisfied as long as $x_B(t) - x_A(t) \geq 1$. The perturbing agent remains close to the center of mass, attracting both clusters, until at some time $t^*$ we have $x_B(t^*) - x_A(t^*) < 1$. The two clusters then merge at the next time step. The result of this process is independent of the weight $\delta$ of the perturbing agent, which proves that $\tilde{x}$ is not stable. Finally, a similar but slightly more complicated argument shows that $\tilde{x}$ is not stable when $|x_A - x_B| = 1 + \frac{\min(W_A, W_B)}{\max(W_A, W_B)}$, and $|x_A - x_B| < 2$.

Theorem 2 characterizes the stable equilibria in terms of a lower bound on the inter-cluster distances. It allows for inter-cluster distances at a stable equilibrium that are smaller than 2, provided that the clusters have different weights. This is consistent with experimental observations for certain initial opinion distributions, as shown in Figure 5. On the other hand, for the frequently observed case of clusters with equal weights, stability requires the inter-cluster distances to be at least 2. Thus, this result comes close to a full explanation of the observed inter-cluster distances of about 2.2.

In general, there is no guarantee that the system (1) will converge to a stable equilibrium. (A trivial example is obtained by initializing the system at an unstable equilibrium, such as $x_i(0) = -\frac{1}{2}$ for half of the agents and $x_i(0) = \frac{1}{2}$ for the other half). On the other hand, we have observed that for a given distribution of initial opinions, and as the number of agents increases, we almost always obtain convergence to a stable equilibrium. This leads us to the following conjecture.

**Conjecture 1:** Suppose that the initial opinions are chosen randomly and independently according to a particular continuous and bounded probability density function (PDF) with connected support. Then, the probability of convergence to a stable equilibrium tends to 1, as the number of agents increases to infinity.

Besides the extensive numerical evidence (see e.g., Figure 6), this conjecture is supported by the intuitive idea that if the number of agents is sufficiently large, whenever two groups of agents start forming two clusters, there will still be a small number agents in between, whose presence will preclude convergence to an unstable equilibrium. The conjecture is also supported by Theorem 7 in Section III, which deals with a continuum of agents, together with the results in Section IV that provide a link between the discrete-agent and continuous-agent models.
III. THE CONTINUOUS-AGENT MODEL

The discussion in the previous section indicates that much insight can be gained by focusing on the case of a large number of agents. This motivates us to consider a model involving a continuum of agents. We use the interval \( I = [0, 1] \) to index the agents, and we consider opinions that are nonnegative and bounded above by a positive constant \( L \). We denote by \( x_t(\alpha) \) the opinion of agent \( \alpha \in I \) at time \( t \). We use \( X \) to denote the set of measurable functions \( x : I \to \mathbb{R} \), and \( X_L \subseteq X \) the set of measurable functions \( x : I \to [0, L] \). The evolution of the opinions is described by

\[
x_{t+1}(\alpha) = \frac{ \int_{\beta} \chi_x(\alpha, \beta) x_t(\beta) d\beta }{ \int_{\beta} \chi_x(\alpha, \beta) d\beta },
\]

where \( C_x \subseteq I^2 \) is defined for any \( x \in X \) by

\[
C_x := \{ (\alpha, \beta) \in I^2 : |x(\alpha) - x(\beta)| < 1 \}.
\]

If the denominator in (5) is zero, we use the convention \( x_{t+1}(\alpha) = x_t(\alpha) \). However, since the set of agents \( \alpha \) for which this convention applies has zero measure, we can ignore such agents in the sequel. We assume that \( x_0 \in X_L \). We then see that for every \( t > 0 \), we have \( x_t \in X_L \), so that the dynamics are well-defined. In the sequel, we denote by \( \chi_x \) the indicator function of \( C_x \), that is, \( \chi_x(\alpha, \beta) = 1 \) if \( (\alpha, \beta) \in C_x \), and \( \chi_x(\alpha, \beta) = 0 \) otherwise.

We note that for the same reasons as in the discrete-agent model, if for some \( \alpha \) and \( \beta \) we have the relation \( x_t(\alpha) \leq x_t(\beta) \) or \( x_t(\alpha) = x_t(\beta) \) at some \( t \), then the same relation continues to hold at all subsequent times. Furthermore, if \( x_0 \) only takes a finite number of values, the continuous-agent model coincides with the weighted discrete-agent model (2), with the same range of initial opinions, and where each discrete agent’s weight is set equal to the measure of the set of indices \( \alpha \) for which \( x_0(\alpha) \) takes the corresponding value.

In the remainder of this section, we will study the convergence properties of the continuous-agent model, and the inter-cluster distances at suitably defined stable equilibria.

A. Operator formalism

To analyze the continuous-agent model (5), it is convenient to introduce a few concepts, extending well known matrix and graph theoretic tools to the continuous case. By analogy with interaction graphs in discrete multi-agent systems, we define for \( x \in X \) the adjacency operator \( A_x \), which maps the set \( X \) of measurable functions on \( I \) into itself, by letting

\[
(A_x y)(\alpha) = \int \chi_x(\alpha, \beta) y(\beta) d\beta.
\]

Applying this operator can be viewed as multiplying \( y \) by the “continuous adjacency matrix” \( \chi_x \), and using an extension of the matrix product to the continuous case. We also define the degree function \( d_x : I \to \mathbb{R}^+ \), representing the measure of the set of agents to which a particular agent is connected, by

\[
d_x(\alpha) = \int \chi_x(\alpha, \beta) d\beta = (A_x 1)(\alpha),
\]

where \( 1 : I \to \{1\} \) is the constant function that takes the value 1 for every \( \alpha \in I \). Multiplying a function by the degree function can be viewed as applying an operator \( D_x : X \to X \) defined by

\[
(D_x y)(\alpha) = d_x(\alpha) y(\alpha) = \int \chi_x(\alpha, \beta) y(\alpha) d\beta.
\]

When \( d_x \) is positive everywhere, we can also define the operator \( D_x^{-1} \), which multiplies a function by \( 1/d_x \). Finally, we define the Laplacian operator \( L_x = D_x - A_x \). It follows directly from these definitions that \( L_x 1 = 0 \), similar to what is known for the Laplacian matrix. In the sequel, we also use the scalar product \( \langle x, y \rangle = \int x(\alpha) y(\alpha) d\alpha \). We now introduce two lemmas to ease the manipulation of these operators.

**Lemma 1:** The operators defined above are symmetric with respect to the scalar product: for any \( x, y, z \in X \), we have

\[
\langle z, A_x y \rangle = \langle A_x z, y \rangle, \quad \langle z, D_x y \rangle = \langle D_x z, y \rangle, \quad \langle z, L_x y \rangle = \langle L_x z, y \rangle.
\]

**Proof:** The result is trivial for \( D_x \). For \( A_x \), we have

\[
\langle z, A_x y \rangle = \int z(\alpha) \left( \int \chi_x(\alpha, \beta) y(\beta) d\beta \right) d\alpha = \int y(\beta) \left( \int \chi_x(\alpha, \beta) z(\alpha) d\alpha \right) d\beta.
\]

Since \( \chi_x(\alpha, \beta) = \chi_x(\beta, \alpha) \) for all \( \alpha, \beta \), this implies \( \langle z, A_x y \rangle = \langle A_x z, y \rangle \). By linearity, the result also holds for \( L_x \) and any other linear combination of those operators.

**Lemma 2:** For any \( x, y \in X \), we have

\[
\langle y, (D_x \pm A_x) y \rangle = \frac{1}{2} \int \chi_x(\alpha, \beta) (y(\alpha) \pm y(\beta))^2 d\alpha d\beta.
\]

In particular, \( L_x = D_x - A_x \) is positive semi-definite.

**Proof:** From the definition of the operators, we have

\[
\langle y, (D_x \pm A_x) y \rangle = \int \chi_x(\alpha, \beta) (y(\alpha) \pm y(\beta)) d\alpha d\beta.
\]

The right-hand side of this equality can be rewritten as

\[
\frac{1}{2} \left( \int \chi_x(\alpha, \beta) y(\alpha) (y(\alpha) \pm y(\beta)) d\alpha d\beta \right) + \frac{1}{2} \left( \int \chi_x(\alpha, \beta) y(\beta) (y(\beta) \pm y(\alpha)) d\alpha d\beta \right).
\]

The symmetry of \( \chi_x \) then implies that \( \langle y, (D_x \pm A_x) y \rangle \) equals

\[
\frac{1}{2} \int \chi_x(\alpha, \beta) (y(\alpha)^2 \pm 2y(\alpha)y(\beta) + y(\beta)^2) d\alpha d\beta,
\]

from which the results follows directly.

The update equation (5) can be rewritten, more compactly, in the form

\[
\Delta x_t := x_{t+1} - x_t = -L_x^{-1} L_x x_t, \quad \text{or} \quad D_{x_t} \Delta x_t = -L_x x_t,
\]

where the second notation is formally more general as it also holds on the possibly nonempty zero-measure set on which \( d_x = 0 \). We say that \( x_t \in X_L \) is a fixed point of the system

\[
\Delta x_t = 0 \quad \text{almost everywhere (a.e., for short), that is, except possibly on a zero-measure set. It follows from (6) that the set of fixed points is characterized by the equality } L_x x = 0, \text{ a.e. One can easily see that the set of fixed points contains the set } F \equiv \{ x \in X_L : x(\alpha) \neq x(\beta) \Rightarrow |x(\alpha) - x(\beta)| > 1 \} \text{ of opinion functions taking a discrete number of values that are at least one apart. Let } F \text{ be the set of functions } x \in X_L \text{ for which there exists } s \in F \text{ such that } s = x, \text{ a.e. We prove later that } F \text{ is exactly the set of solutions to } L_x x = 0, \text{ a.e., and thus the set of fixed points of (6).}
B. Convergence

In this section we present some partial convergence results. In particular, we show that the change \( \Delta x_t \) of the opinion function decays to 0, and that \( x_t \) tends to the set of fixed points. We begin by proving the decay of a quantity related to \( \Delta x_t \).

**Theorem 3:** For any initial condition of the system (6), we have

\[
\sum_{t=0}^{\infty} \int \chi_{x_t}(\alpha, \beta) (\Delta x_t(\alpha) + \Delta x_t(\beta))^2 \, d\alpha \, d\beta < \infty.
\]

**Proof:** We consider the nonnegative potential function \( V : X \to \mathbb{R}^+ \) defined by

\[
V(x) = \frac{1}{2} \int \min \left( 1, (x(\alpha) - x(\beta))^2 \right) \, d\alpha \, d\beta \geq 0,
\]
and show that

\[
V(x_{t+1}) - V(x_t) \leq -\langle \Delta x_t, (A_{x_t} + D_{x_t}) \Delta x_t \rangle,
\]

where which by Lemma 2 implies the desired result.

We observe that for every \( y \in X \),

\[
V(y) \leq \frac{1}{2} \int_{C_x} (y(\alpha) - y(\beta))^2 \, d\alpha \, d\beta + \frac{1}{2} \int_{I^2 \setminus C_x} 1 \, d\alpha \, d\beta = \langle y, L_x y \rangle + \frac{1}{2} |I^2 \setminus C_x|,
\]

where Lemma 2 was used to obtain the last equality. For \( y = x \), it follows from the definition of \( C_x \) that the above inequality is tight. In particular, the following two relations hold for any \( s \) and \( t \):

\[
\begin{align*}
V(x_t) &= \langle x_t, L_{x_t} x_t \rangle + \frac{1}{2} |I^2 \setminus C_{x_t}| \\
V(x_s) &
\end{align*}
\]

Taking \( s = t + 1 \), we obtain

\[
V(x_{t+1}) - V(x_t) \leq \langle x_{t+1}, L_{x_{t+1}} x_{t+1} \rangle - \langle x_t, L_{x_t} x_t \rangle = 2 \langle \Delta x_t, x_t \rangle + \langle \Delta x_t, x_t \rangle,
\]

where we have used the symmetry of \( L_{x_t} \). It follows from (6) that \( L_{x_t} x_t = -D_{x_t} x_t \), so that

\[
V(x_{t+1}) - V(x_t) \leq -2 \langle \Delta x_t, D_{x_t} x_t \rangle + \langle \Delta x_t, L_{x_t} x_t \rangle = -\langle \Delta x_t, (A_{x_t} + D_{x_t}) \Delta x_t \rangle,
\]

since \( L_{x_t} = D_{x_t} - A_{x_t} \).

As will be seen below, this result implies the convergence of \( \Delta x_t \) to 0 in a suitable topology. Since the proof of Theorem 3 does not rely on the particular form of the dependence on \( x \) of the interconnection topology \( C_x \), the theorem holds for any time-varying symmetric interconnection topology. This is not the case though for subsequent results.

We now show that for the specific interconnection topology used in our model, namely, \( (\alpha, \beta) \in C_x \iff |x(\alpha) - x(\beta)| < 1 \), \( L_{x_t} \) is small only if \( x_t \) is close to \( F \), the set of functions taking discrete values separated by at least 1. As a corollary, we then obtain the result that \( F \) is exactly the set of fixed points, as also shown in [24]. The intuition behind the proof of these results parallels our proof of Theorem 1, and is as follows. Consider an agent \( \alpha \) with one of the smallest opinions \( x(\alpha) \). If the change in \( x(\alpha) \) is small, its attraction by agents with larger opinions must be small, because almost no agents have an opinion smaller than \( x(\alpha) \). Therefore, there must be very few agents with an opinion significantly larger than \( x(\alpha) \) that interact with \( \alpha \), while there might be many of them who have an opinion close to \( x(\alpha) \). In other words, possibly many agents have approximately the same opinion \( x(\alpha) \), and very few agents have an opinion in the interval \( [x(\alpha) + \epsilon, x(\alpha) + 1] \), so that \( x \) is close to a function in \( F \) in that zone. Take now an agent \( \alpha' \) with an opinion larger than \( x(\alpha) + 1 + \epsilon \), and such that very few agents have an opinion in \( (x(\alpha) + 1 + \epsilon, x(\alpha')) \). This agent interacts with very few agents having an opinion smaller than its own. Thus, if the change in such an agent’s opinion is small, this implies that its attraction by agents having larger opinions is also small, and we can repeat the previous reasoning.

In order to provide a precise statement of the result, we associate an opinion function \( x \) with a measure that describes the distribution of opinions, and use a measure-theoretic formalism. For a measurable function \( x : I \to [0, L] \) (i.e., \( x \in X_L \)), and a measurable set \( S \subseteq [0, L] \), we let \( \mu_S (x) \) be the Lebesgue measure of the set \( \{ x(\alpha) \in S \} \). By convention, we let \( \mu_S (x) = 0 \) if \( S \subseteq \mathbb{R} \setminus [0, L] \). To avoid confusion with \( \mu \), we use \( |S| \) to denote the standard Lebesgue measure of a set \( S \). We also introduce a suitable topology on the set of opinion functions. We write \( x \leq_\mu \epsilon \) if \( \{ x(\alpha) > \epsilon \} \leq \epsilon \). Similarly, \( x <_\mu \epsilon \) if \( \{ x(\alpha) < \epsilon \} < \epsilon \) and \( x =_\mu \epsilon \) if \( \{ x(\alpha) = \epsilon \} = 0 \). We define the “ball” \( B_\mu (x, \epsilon) \) as the set \( \{ y \in X_L : |x - y| <_\mu \epsilon \} \). This allows us to define a corresponding notion of limit. We say that \( x_t \to_\mu y \) if for all \( \epsilon > 0 \), there is a \( t' \) such that for all \( t > t' \) there is \( x_t \in B_\mu (y, \epsilon) \). We write \( x_t \to_\mu S \) for a set \( S \) if for all \( \epsilon > 0 \), there is a \( t' \) such that for all \( t > t' \), there is a \( y \in S \) for which \( x_t \in B_\mu (y, \epsilon) \).

The result below, proved in Appendix A, states that the distance between \( x \in X_L \) and \( F \) (the subset of \( X_L \) consisting of functions taking discrete values separated by at least 1) decreases to 0 (in a certain uniform sense) when \( L_x \to_\mu 0 \).

**Theorem 4:** For any \( \epsilon > 0 \), there exists a \( \delta > 0 \) such that if \( |L_x x| <_\mu \delta \), then there exists some \( s \in F \) with \( |x - s| <_\mu \epsilon \). In particular, if \( L_x x =_\mu 0 \), then \( x \in F \).

The next theorem compiles our convergence results.

**Theorem 5:** Let \( (x_t) \) be a sequence of functions in \( X_L \) evolving according to the model (5), and let \( F \) be the set of functions taking discrete values separated by at least 1. Then \( (x_{t+1} - x_t) \to_\mu 0 \), and \( x_t \to_\mu F \). (In particular, periodic trajectories, other than fixed points, are not possible.) Furthermore, \( x \) is a fixed point of (5) if and only if \( x \in F \).

**Proof:** We begin by proving the convergence of \( \Delta x_t \). Suppose that \( \Delta x_t \to_\mu 0 \) does not hold. Then, there is an \( \epsilon > 0 \) such that for arbitrarily large \( t \), there is a set of measure at least \( \epsilon \) such that \( |\Delta x_t| > \epsilon \) for every \( \alpha \) in that set. Consider such a time \( t' \). Without loss of generality, assume that there is a set \( S \subseteq I \) of measure at least \( \epsilon/2 \) on which \( \Delta x_t(\alpha) > \epsilon \). (Otherwise, we can use a similar argument for the set on which \( \Delta x_t(\alpha) < -\epsilon \).) Fix some \( L' > L \) for \( \alpha \in S \). For any \( i \in \{1, \ldots, 2[L']\} \), let \( A_i \subseteq I \) be the set on which \( x_t \in [(i - 1)/2, i/2] \). For any \( i \) and for any \( \alpha, \beta \in A_i \), holds \( |x_t(\alpha) - x_t(\beta)| < 1 \) and thus \( (\alpha, \beta) \in C_{x_t} \). Therefore, \( A_i \subseteq C_{x_t} \) for all \( i \). Moreover, the sets \( A_i \) cover \( [0,1] \), so that
obtain convergence to a single cluster.) Moreover, it can be verified that the notion of stability used here is equivalent to both \( L_1 \) and \( L_2 \) stability. In the sequel, and to simplify the presentation, we will neglect any zero measure sets on which \( \Delta x_t(\alpha) \neq 0 \), and will give the proof for a fixed point in \( F \). The extension to fixed points in \( F \) is straightforward. The proof of the following result is similar to that of its discrete counterpart, the necessary part of Theorem 2, and is presented in the Appendix B.

**Theorem 6:** Let \( s \in F \) be a fixed point of (5), and let \( a, b \) two values taken by \( s \). If \( s \) is stable, then

\[
|b - a| \geq \frac{\min (\mu_s(a), \mu_s(b))}{\max (\mu_s(a), \mu_s(b))}.
\]

With a little extra work, focused on the case where the distance \( |a - b| \) between the two clusters is exactly equal to 2, we can show that the strict inequality version of condition (9) is necessary for stability. We conjecture that this strict inequality version is also sufficient.

We will now proceed to show that under an additional smoothness assumption on the initial opinion function, we can never have convergence to a fixed point that violates condition (9). We start by introducing the notion of a regular opinion function. We say that a function \( x \in X_L \) is regular if there exist \( M \geq m > 0 \) such that any interval \( J \subseteq [\inf_0 x, \sup_0 x] \) satisfies \( m |J| \leq \mu_x(J) \leq M |J| \). Intuitively, a function is regular if the set of opinions is connected, and if the density of agents on any interval of opinions is bounded from above and from below by positive constants. (In particular, no single value is taken by a positive measure set of agents.) For example, any piecewise differentiable \( x \in X_L \) with positive upper and lower bounds on its derivative is regular.

We will show that if \( x_0 \) is regular and if \( (x_t) \) converges, then \( x_t \) converges to an equilibrium satisfying the condition (9) on the minimal distance between opinions, provided that \( \sup_0 x_t - \inf_0 x_t \) remains always larger than 2. For convenience, we introduce a nonlinear update operator \( U \) on \( X_L \), defined by \( U(x) = x - D_x^{-1}L_x x = D_x^{-1}A_x x \), so that the recurrence (5) can be written as \( x_{t+1} = U(x_t) \). The proof of the following proposition is presented in Appendix C.

**Proposition 3:** Let \( x \in X_L \) be a regular function such that \( \sup_0 x - \inf_0 x > 2 \). Then \( U(x) \) is regular.

We note that the assumption \( \sup_0 x - \inf_0 x > 2 \) in Proposition 3 is necessary for the result to hold. Indeed, if the opinion values are confined to a set \([a, b]\), with \( b - a = 2 - \delta < 2 \), then all agents with opinions in the set \([a + 1 - \delta, a + 1]\) are connected with every other agent, and their next opinions will be the same, resulting in a non-regular opinion function.

As a consequence of Proposition 3, together with Theorem 5, if \( x_0 \) is regular, then there are two main possibilities: (i) There exists some time \( t \) at which \( \sup_0 x_t - \inf_0 x_t < 2 \). In this case, the measure \( \mu_{x_t} \) will have point masses shortly thereafter, and will eventually converge to the set of fixed points with at most two clusters. (ii) Alternatively, in the “regular” case, we have \( \sup_0 x_t - \inf_0 x_t > 2 \) for all times. Then, every \( x_t \) is regular, and convergence cannot take place in finite time. Furthermore, as we now proceed to show, convergence to a fixed point that violates the stability condition

\[
\sum_{i=1}^{2|L'|} |A_i \cap S| \geq |S| \geq \epsilon/2.
\]

Thus, there exists some \( i^* \) such that \( |A_{i^*} \cap S| \geq \epsilon/4|L'| \). We then have

\[
\int_{C_{x_t}} (\Delta x_t(\alpha) + \Delta x_t(\beta))^2 \ d\alpha \ d\beta
\geq \int_{(A_{i^*} \cap S)^2} (\Delta x_t(\alpha) + \Delta x_t(\beta))^2 \ d\alpha \ d\beta
\geq 4\epsilon^2 (A_{i^*} \cap S)^2 \geq \epsilon^4/4|L'|^2.
\]

Thus, if \( \Delta x_t \to \mu \) does not hold, then \( \int_{(\alpha, \beta) \in C_{x_t}} (\Delta x_t(\alpha) + \Delta x_t(\beta))^2 \ d\alpha \ d\beta \) does not decay to 0, which contradicts Theorem 3. We conclude that \( \Delta x_t \to \mu \). Using also (6) and the fact \( d_{C_{x_t}}(\alpha) \leq 1 \), we obtain \( L_{x_t} x_t \to \mu \). Theorem 4 then implies that \( x_t \to \mu \).

If \( x \in F \), it is immediate that \( x \) is a fixed point. Conversely, if \( x_0 = x \) is a fixed point, then \( x_t = x_0 \), a.e., for all \( t \). Then, the fact \( x_t \to \mu \) implies that \( x \). We note that the fact \( x_t \to \mu \) means that the measure \( \mu_x \) associated with any limit point \( x \) of \( x_t \) is a discrete measure whose support consists of values separated by at least 1. Furthermore, it can be shown that at least one such limit point exists, because of the semi-compactness of the set of measures under the weak topology.

Theorem 5 states that \( x_t \) tends to the set \( F \), but does not guarantee convergence to an element of this set. We make the following conjecture, which is currently unresolved.

**Conjecture 2:** Let \( (x_t) \) be a sequence of functions in \( X \), evolving according to the model (5). Then, there is a function \( x^* \in F \) such that \( x_t \to \mu \) x^*.

### C. Inter-cluster distances and stability of equilibria

We have found that \( x \) is a fixed point of (5) if and only if it belongs to \( F \), that is, with the exception of a zero-measure set, the range of \( x \) is a discrete set of values that are separated by at least one. As before, we will refer to these discrete values as clusters. In this section, we consider the stability of equilibria, and show that a condition on the inter-cluster distances similar to the one in Theorem 2 is necessary for stability. Furthermore, we show that under a certain smoothness assumption, the system cannot converge to a fixed point that does not satisfy this condition.

In contrast to the discrete case, we can study the continuous-agent model using the classical definition of stability. We say that \( s \in F \) is stable if for any \( \epsilon > 0 \), there is a \( \delta > 0 \) such that for any \( x_0 \in B_\delta(s, \epsilon) \), we have \( x_t \in B_\epsilon(s, \epsilon) \) for all \( t \). It can be shown that this notion encompasses the stability with respect to the addition of a perturbing agent used in Section II-C. More precisely, if we view the discrete-agent system as a special case of the continuum model, stability under the current definition implies stability with respect to the notion used in Section II-C. The introduction of a perturbing agent with opinion \( \tilde{x}_0 \) can indeed be simulated by taking \( x_0(\alpha) = s(\alpha) \) everywhere except on an appropriate set of measure less than \( \delta \), and \( x_0(\alpha) = \tilde{x}_0 \) on this set. (However, the converse implication turns out to not hold in some pathological cases. Indeed, consider two agents separated by exactly 2. They are stable with respect to the definition of Section II-C, but not under the current definition. This is because if we introduce a small measure set of additional agents that are uniformly spread between the two original agents, we will
(9) is impossible. Let us note however that tight conditions for a sequence of regular functions to maintain the property \( \sup_{\alpha} x_{t\alpha} - \inf_{\alpha} x_{t\alpha} > 2 \) at all times appear to be difficult to obtain.

**Theorem 7:** Let \((x_t)\) be a sequence of functions in \(X_L\) that evolve according to (5). We assume that \(x_0\) is regular and that \(\sup_{\alpha} x_{t\alpha} - \inf_{\alpha} x_{t\alpha} > 2\) for all \(t\). If \((x_t)\) converges, then it converges to a function \(s \in F\) such that
\[
|b - a| \geq 1 + \min \left( \mu_x(a), \mu_y(b) \right) \max \left( \mu_x(a), \mu_y(b) \right),
\]
for any two distinct values \(a, b\), with \(\mu_x(a), \mu_y(b) > 0\). In particular, if \(\mu_x(a) = \mu_y(b)\), then \(|b - a| \geq 2\).

**Proof:** Suppose that \((x_t)\) converges to some \(s\). It follows from Theorem 5 that \(s \in F\), and from Proposition 3 that all \(x_t\) are regular. Suppose now that \(s\) violates the condition in the theorem, for some \(a, b\), with \(a < b\). Then, \(b - a < 2\), and we must have \(\mu_x(a), \mu_y(b) > 0\) because all discrete values taken by \(s\) (with positive measure) must differ by at least 1. We claim that there exists a positive length interval \(J \subseteq (a, b)\) such that \(\mu_x(J) \geq \mu_x(J)\) whenever \(x_t \in B_p(s, \epsilon)\), for some sufficiently small \(\epsilon > 0\). Since \(x_t\) converges to \(s\), this will imply that there exists a finite time \(t^*\) after which \(\mu_x(J)\) is nondecreasing, and \(\lim_{t \to -\infty} \mu_x(J) \geq \mu_x(J)\) is violated. On the other hand, since \(\mu_x(a, b) = 0\), \(\mu_x(J)\) must converge to zero. This is a contradiction and establishes the desired result.

We now establish the above claim. Let \(c = \mu_x(a, b)\) be the weighted average of \(a\) and \(b\). The fact that the condition in the theorem is violated implies (cf. (3)) that \(|c - a| < 1\) and \(|c - b| < 1\). Let \(\delta > 0\) be such that \(c - \delta + 1 < b\) and \(c + \delta - 1 < a\), and consider the interval \(J = [c - \delta, c + \delta]\). For any \(s \in B_p(s, \epsilon)\), we have
\[
\mu_x([a - \epsilon, a + \epsilon]) \subseteq \mu_x(a) - \epsilon, \mu_x(a) + \epsilon],
\]
\[
\mu_x([b - \epsilon, b + \epsilon]) \subseteq \mu_x(b) - \epsilon, \mu_x(b) + \epsilon],
\]
\[
\mu_x((a - 1, b + 1) \setminus ([a - \epsilon, a + \epsilon] \cup [b - \epsilon, b + \epsilon])) \subseteq \epsilon,
\]
where we have used the fact that the values taken by \(s\) are separated by at least 1. Suppose now that \(\epsilon\) is sufficiently small so that \(c - \delta + 1 > b + \epsilon\) and \(c + \delta - 1 < a - \epsilon\). This implies that for every \(\gamma\) such that \(x(\gamma) \in J\), we have \((a - \epsilon, b + \epsilon) \subseteq (x(\gamma) - 1, x(\gamma) + 1)\). If \(\epsilon\) are equal to zero, we would have \(u_x(d) = c\). When \(\epsilon\) is small, the location of the masses at \(a\) and \(b\) moves by an \(O(\epsilon)\) amount, and an additional \(O(\epsilon)\) mass is introduced. The overall effect is easily shown to be \(O(\epsilon)\) (the detailed calculation can be found in (15)). Thus, \(\|U_x(\gamma) - \gamma\| \leq \|x(\epsilon)\| \leq \epsilon + \delta\), i.e., \((U_x(\gamma))(\gamma) \in J\) for all \(\gamma\) such that \((x(\gamma)) \in J\). This implies that \(\mu(U_x(J)) \geq \mu_x(J)\), and completes the proof.

**IV. RELATION BETWEEN THE DISCRETE AND THE CONTINUOUS-AGENT MODELS**

We now analyze the extent to which the continuous-agent model (5) can be viewed as a limiting case of the discrete-agent model (1), when the number of agents tends to infinity. As already explained in Section III, the continuous-agent model can simulate exactly the discrete-agent model. In this section, we are interested in the converse; namely, the extent to which a discrete-agent model can describe, with arbitrarily good precision, the continuous-agent model. We will rely on the following result on the continuity of the update operator.

**Proposition 4:** Let \(x \in X_L\) be a regular function. Then, the update operator \(U\) is continuous at \(x\) with respect to the norm \(\|\cdot\|\). More precisely, for any \(\varepsilon > 0\) there exists some \(\delta > 0\) such that if \(\|x - y\| < \delta\) then \(\|U(y) - U(x)\| < \varepsilon\).

**Proof:** Consider a regular function \(x \in X_L\), and an arbitrary \(\varepsilon > 0\). Let \(\delta = \frac{\varepsilon}{2M}\), where \(m = M\) (with \(m \leq M\) are the bounds in the definition of regular opinion functions applied to \(x\). We will show that if a function \(y \in X_L\) satisfies \(\|x - y\| < \delta\), then \(\|U(y) - U(x)\| < \varepsilon\).

Fix some \(\alpha \in I\), and let \(S_y, S_x \subseteq I\) be the set of agents connected to \(\alpha\) according to the interconnection topologies \(C_x\) and \(C_y\) defined by \(x\) and \(y\), respectively. We let \(S_x \cap y - S_x \cap y = S_x \setminus S_y \cap y = S_y \cap y = S_y \cap y\). Since \(\|x - y\| < \delta\), the values \(|x(\alpha) - x(\beta)|\) and \(|y(\alpha) - y(\beta)|\) differ by at most \(2\delta\), for any \(\beta \in I\). As a consequence, \(x(\alpha) - x(\beta) \leq |y(\alpha) - y(\beta)| + 2\delta\). Similarly, if \(\beta \notin I\), then \(|x(\alpha) - x(\beta)| \geq |y(\alpha) - y(\beta)| - 2\delta\). Combining these two inequalities with the definitions of \(S_y\), \(S_x\), \(S_y \cap y\), and \(S_y \setminus y\), we obtain
\[
|x(\alpha) - 1 - 2\delta, x(\alpha) + 1 + 2\delta| \subseteq S_x \cap y - S_y \cap y, S_y \cap y \cap S_x \subseteq S_x \setminus S_y \cap y \subseteq S_y \cap y \subseteq S_y \cap y.
\]
Since \(x\) is regular, we have \(S_y \cap y \geq m(2 - 4\delta) \geq m\) and \(S_x \cap y, S_y \setminus y \geq M4\delta\). Let now \(\bar{x}_x, \bar{y}_x, \bar{y}_y\), be the average value of \(x\) on \(S_x \cap y\) and \(S_y \setminus y\), respectively. Similarly, let \(\bar{y}_x, \bar{y}_y, \bar{y}_y\), be the average value of \(y\) on \(S_x \cap y\) and \(S_y \cap y\). Since \(\|x - y\| < \delta\), \(\bar{x}_x, \bar{y}_x, \bar{y}_y\) differ by at most \(\delta\). It follows from the definition of the model (5) that
\[
(U(x)(\alpha)) = \bar{x}_x + \frac{S_x \cap y}{|S_x \cap y|}, (\bar{x}_x y - \bar{y}_x),
\]
\[
(U(y)(\alpha)) = \bar{y}_y + \frac{S_y \cap y}{|S_y \cap y|}, (\bar{y}_x y - \bar{y}_y).
\]
It can be seen that \(\bar{x}_x y - \bar{y}_x \leq 3\) and \(\bar{y}_y \setminus y \leq 3\), from which we obtain that \(\|U(x)(\alpha) - U(y)(\alpha)\|\) is upper \(\bar{x}_x y - \bar{y}_y + \frac{S_x \cap y}{|S_x \cap y|} + \frac{S_y \cap y}{|S_y \setminus y|} \leq \delta + 6M4\delta \leq \delta\). Since the above is true for any \(\alpha \in I\), we conclude that \(\|U(y) - U(x)\| \leq \varepsilon\).

Let \(U^t : X_L \to X_L\) be the composition of the update operator, defined by \(U^t(x) = U(U^{t-1}(x))\), so that \(U^t(x_0) = x_t\). Proposition 4 readily extended to a continuity result for \(U^t\).

**Corollary 1:** Let \(x_0 \in X_L\) be a regular function such that \(\sup_{\alpha} U^t(x_0) - \inf_{\alpha} U^t(x_0) > 2\) for every \(t \geq 0\). Then for any finite \(t\), \(U^t\) is continuous at \(x\) with respect to the norm \(\|\cdot\|\).

**Proof:** Since \(x\) is regular and since \(\sup_{\alpha} U^t(x_0) - \inf_{\alpha} U^t(x_0) > 2\) for every \(t \geq 0\), for any finite \(t\), \(U^t\) is continuous at \(x\), and therefore the composition \(U^t\) is continuous at \(x\).

**Corollary 1** allows us to prove that, in the regular case, and for any given finite time horizon, the continuous-agent model is the limit of the discrete-agent model, as the number of agents grows. To this effect, for any given partition of \(I = \)
We define $\hat{\alpha}_i$ for $i = 1, \ldots, n$, such that the sequence of vectors $\hat{x}_t$ generated by the discrete-agent model (1), starting from $\hat{x}_0$, satisfies $\|x_t - \mathcal{G}\hat{x}_t\|_\infty \leq \epsilon$, for $t = 1, \ldots, t^\star$.

Proof: Fix $\epsilon > 0$. Since all $U^t$ are continuous at $x_0$, there is some $\delta > 0$ such that if $\|y - x_0\|_\infty \leq \delta$, then $\|U^t(y) - x_t\|_\infty \leq \epsilon$, for $t \leq t^\star$. Since $x_0$ is regular, we can divide $I$ into subsets $J_1, J_2, \ldots, J_n$, so that $|J_i| = 1/n$ for all $i$, and $|x_0(\alpha) - x_0(\beta)| \leq \delta$ for all $\alpha, \beta$ in the same set $J_i$. (This is done by letting $c_i$ be such that $\mu_{x_0}(0, c_i) = i/n$, and defining $J_i = \{\alpha : c_{i-1} \leq x_0(\alpha) \leq c_i\}$, where $n$ is sufficiently large.) We define $\tilde{x}_0 \in [0, L]^n$ by letting its $i$th component be equal to $c_i$. We then have $\|x_t - \mathcal{G}\tilde{x}_0\|_\infty \leq \delta$. This implies that $\|x_t - U^t(\mathcal{G}\hat{x}_0)\|_\infty \leq \epsilon$, for $t \leq t^\star$. Since the continuous-agent model, initialized with a discrete distribution, simulates the discrete-agent model, we have $U^t(\mathcal{G}\hat{x}_0) = \mathcal{G}\hat{x}_t$, and the desired result follows.

Theorem 8 supports the intuition that for large values of $n$, the continuous-agent model behaves approximately as the discrete-agent model, over any finite horizon. In view of Theorem 6, this suggests that the discrete-agent system should always converge to a stable equilibrium (in the sense defined in Section II) when $n$ is sufficiently large, as stated in Theorem 1, and observed in many examples (see, e.g., Figure 6). Indeed, Theorem 6 states that under the regularity assumption, the continuum system cannot converge to an equilibrium that does not satisfy condition (9) on the inter-cluster distances. However, this argument does not translate to a proof of the conjecture because the approximation property in Theorem 8 only holds over a finite time horizon, and does not necessarily provide information on the limiting behavior.

V. CONCLUSIONS AND OPEN QUESTIONS

We have analyzed the model of opinion dynamics (1) introduced by Krause, from several angles. Our motivation was to provide an analysis of a simple multi-agent system with an endogenously changing interconnection topology while taking explicitly advantage of the topology dynamics, something that is rarely done in the related literature.

We focused our attention on an intriguing phenomenon, the fact that equilibrium inter-cluster distances are usually significantly larger than 1, and typically close to 2. We proposed an explanation of this phenomenon based on a notion of stability with respect to the addition of a perturbing agent. We showed that such stability translates to a certain lower bound on the inter-cluster distances, with the bound equal to 2 when the clusters have identical weights. We also discussed the conjecture that when the number of agents is sufficiently large, the system converges to a stable equilibrium for “most” initial conditions.

To avoid granularity problems linked with the presence or absence of an agent in a particular region, we introduced a new opinion dynamics model that allows for a continuum of agents. For this model we proved that under some regularity assumptions, there is always a finite density of agents between any two clusters during the convergence process. As a result, we could prove that such systems never converge to an unstable equilibrium. We also proved that the continuous-agent model is indeed the limit of a discrete model, over any given finite time horizon, as the number of agents grows to infinity. These results provide some additional support for the conjectured, but not yet established, generic convergence to stable equilibria.

We originally introduced the continuous-agent model as a tool for the study of the discrete-agent model, but it is also of independent interest and raises some challenging open questions. An important one is the question of whether the continuous-agent model is always guaranteed to converge. (We only succeeded in establishing convergence to the set of fixed points, not to a single fixed point.)

Finally, the study of the continuous-agent model suggests some broader questions. In the same way that the convergence of the discrete-agent model can be viewed as a special case of convergence of inhomogeneous products of stochastic matrices, it may be fruitful to view the convergence of the continuous-agent model as a special case of convergence of inhomogeneous compositions of stochastic operators, and to develop results for the latter problem.

The model (1) can of course be extended to higher dimensional spaces, as is often done in the opinion dynamics literature (see [24] for a survey). Numerical experiments again show the emergence of clusters that are separated by distances significantly larger than 1. The notion of stability with respect to the addition of an agent can also be extended to higher dimensions. However, stability conditions become more complicated, and in particular cannot be expressed as a conjunction of independent conditions, one for each pair of clusters. For example, it turns out that adding a cluster to an unstable equilibrium may render it stable [15]. In addition, a formal analysis appears difficult because in $\mathbb{R}^n$, with $n > 1$, the support of the opinion distribution can be connected without being convex, and convexity is not necessarily preserved by our systems. For this reason, even under “regularity” assumptions, the presence of perturbing agents between clusters is not guaranteed.

REFERENCES


Vincent D. Blondel received the M.Sc. degree in mathematics from Imperial College, London, U.K., in 1990 and the Ph.D. degree in applied mathematics from the Université catholique de Louvain, Louvain-la-Neuve, Belgium, in 1992.

He was a Visiting Researcher at the Royal Institute of Technology, Stockholm, Sweden, and at the Institut National de Recherche en Informatique et en Automatique (INRIA), Rocquencourt, France. During 2005-2006, he was an Invited Professor and a Fulbright Scholar at Massachusetts Institute of Technology, Cambridge. He is currently a Professor and Department Head at the Université catholique de Louvain, Louvain-la-Neuve, Belgium. Dr. Blondel was the recipient of the Prize Wetterns of the Belgian Royal Academy of Science, the Society for Industrial and Applied Mathematics (SIAM) Prize on Control and Systems Theory, and the Ruberti Prize in Systems and Control of the IEEE in 2006.

Julien M. Hendrickx received an engineering degree in applied mathematics and a PhD in mathematical engineering from the Université catholique de Louvain, Belgium, in 2004 and 2008, respectively.

He has been a visiting researcher at the University of Illinois at Urbana Champaign in 2003-2004, at the National ICT Australia in 2005 and 2006, and at the Massachusetts Institute of Technology in 2006 and 2008. He is currently a postdoctoral fellow at the Laboratory for Information and Decision systems of the Massachusetts Institute of Technology, and holds postdoctoral fellowships of the F.R.S.-FNRS (Fund for Scientific Research) and of Belgian American Education Foundation.

Doctor Hendrickx was the recipient of the 2008 EECI award for the best PhD thesis in Europe in the field of Embedded and Networked Control.

John N. Tsitsiklis (F’99) received the B.S. degree in mathematics and the B.S., M.S., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1980, 1981, and 1984, respectively. He is currently a Clarence J. Lebel Professor with the Department of Electrical Engineering, MIT. He has served as a Codirector of the MIT Operations Research Center from 2002 to 2005, and in the National Council on Research and Technology in Greece (2005-2007). His research interests are in systems, optimization, communications, control, and operations research. He has coauthored four books and more than a hundred journal papers in these areas.

Prof. Tsitsiklis was a recipient of an Outstanding Paper Award from the IEEE Control Systems Society (1986), the M.I.T. Edgerton Faculty Achievement Award (1989), the Bodossakis Foundation Prize (1995), and the INFORMS/SCST Prize (1997). He is a member of the National Academy of Engineering. Finally, in 2008, he was conferred the title of Doctor honoris causa, from the Université catholique de Louvain.
b) the sequence $K_n$ for positive integer $\alpha$ be the Lebesgue measure of the set $\{\hat{x} \mid x \in [0, \Delta_i] \},$ so that for $i = 1, \ldots, N,$ we have $K_1 \Delta_i < \max(\frac{2K_1 \Delta_i}{\epsilon}, M).$ The conditions on $\bar{K}_1$ are satisfied for $i = 1, \ldots, N + 1,$ and so are those on $K_i \Delta_i$ for $i = 1, \ldots, N.$ The result then follows from

$$K_{N+1} \Delta_{N+1} = K_{N+1} \left( \frac{2K_1 \Delta_i}{\epsilon} + \frac{1}{K_i} \right) \leq K_{N+1} \left( \frac{2K_1 \Delta_i}{\epsilon} + \frac{1}{2K_{N+1}} \right).$$

To simplify the presentation of the proof, we introduce some new notation. For a given measure $\mu,$ we define the function $\hat{L}_\mu$ by

$$(y-z) \mu(z).$$

Thus, for any $x \in I,$ we have $(L_x \alpha)(x) = \hat{L}_\mu(x \alpha),$ where $\mu_x$ is the measure associated to $x,$ defined by letting $\mu_x(S)$ be the Lebesgue measure of the set $\{\alpha : x(\alpha) \in S\}$ for any measurable set $S.$ Since no ambiguity is possible here as we only use one such measure, we will refer to $\mu_x$ as $\mu$ in the sequel. We also define the nonnegative functions

$$L_\mu^+(y) = \int_{y-z} (y-z) \mu(z) \geq 0,$$

and

$$L_\mu^-(y) = \int_{y-z} (y-z) \mu(z) \geq 0,$$

so that $\hat{L}_\mu = L_\mu^+ - L_\mu^-.$ Using the definition of the relation $\leq \mu,$ we observe that if $L_x \alpha \leq \mu,$ then the set

$$S = \{y \in [0, \Delta_i] \mid \hat{L}_\mu(y) \geq \delta \}$$

satisfies $\mu(S) \leq \delta.$ As a consequence, if $L_x \alpha \leq \mu,$ then for any $z \in [0, \Delta_i]$ at least one of the following must be true:

(i) there exists $y \in [z, L_i]$ such that $L_\mu^+(y) < \hat{L}_\mu(y) + \delta$ and $\mu((z, L_i]) < \delta$; or,

(ii) we have $\mu((z, L_i]) < \delta.$ We will make use of this observation repeatedly in the proof of Theorem 4, which is

Proof: Without loss of generality, we assume that $L$ is integer. (This is because the model with $L$ not integer can be viewed as a special case of a model in which opinions are distributed on $[0, [L]].$) We fix some $\epsilon > 0,$ and without loss of generality we assume that $\epsilon < 1/2.$ Using Lemma 3, we form two sequences $K_1, \ldots, K_{L+1}$ and $\Delta_1, \ldots, \Delta_{L+1}$ that satisfy: (i) $\Delta_{i+1} = 3K_i \Delta_i + \frac{1}{K_i},$ for $i = 1, \ldots, L,$ (ii) $K_i > (L + 1)/\epsilon$ and $K_i \Delta_i < \epsilon,$ for $i = 1, \ldots, L+1.$ In particular, $\Delta_i < \epsilon^2/(L + 1).$ We then choose some $\delta$ smaller than $\Delta_i/3,$ for all $i.$ We will prove the following claim. If $L_x \alpha \leq \mu,$ then there exists some $N \leq L + 1,$ and two nondecreasing finite sequences, $(x_i)$ and $(y_i),$ that satisfy

$$-1 = y_0 \leq x_1 \leq y_1 \leq \cdots \leq x_N \leq y_N,$$

the termination condition $\mu((y_N, L)) \leq \epsilon^2/(L + 1),$ and the following additional conditions, for $i = 1, \ldots, N$:

(a) $L_\mu^+(x_i) < \Delta_i,$

(b) $x_i \geq y_{i-1} + 1,$

(c) $\mu((y_i-1, x_i)) \leq \Delta_i - \delta,$

(d) $0 \leq y_i - x_i \leq K_i \Delta_i < \epsilon.$

The above claim, once established, implies that $\mu$ is "close" to a discrete measure whose support consists of values that are separated by at least 1, and provides a proof of the theorem. To see this, note that the length of each interval $[x_i, y_i]$ is less than $\epsilon.$ Furthermore, the set $[0, \Delta_i] \setminus \bigcup_{i=1}^{N}[x_i, y_i]$ is covered by disjoint intervals, of the form $[0, x_i], (y_i-1, x_i),$ or $(y_N, L).$ Since intervals $(y_i-1, x_i)$ have at least unit length, the overall number of such intervals is at most $L + 1.$ For intervals of the form $[0, x_i], (y_i-1, x_i),$ condition (c) implies that their measure is bounded above by $\Delta_i - \delta < \Delta_i < \epsilon^2/(L + 1).$ Recall also the termination condition $\mu((y_N, L)) < \epsilon^2/(L + 1).$ It follows that the measure of the set $[0, \Delta_i] \setminus \bigcup_{i=1}^{N}[x_i, y_i]$ is at most $\epsilon^2,$ hence smaller than $\epsilon.$ Let $s \in F$ be a function which for every $\alpha$ takes a value $x_1$ which is closest to $x(\alpha).$ Since $x$ can differ from all $x_i$ by more than $\epsilon$ only on a set of measure smaller than $\epsilon,$ it follows that $|x - s| < \epsilon.$ Finally, if $L_x \alpha = \mu,$ then $L_x \alpha \leq \mu$ for all positive $\delta.$ As a consequence, the distance between $x$ and $F$ is smaller than any positive $\epsilon$ and is thus 0. Because $F$ is the closure of $F,$ it follows then that $x \in F.$ Thus, it will suffice to provide a proof of the claim.

We now use a recursive construction to prove the claim. We initialize the construction as follows. Since $L_x \alpha \leq \mu,$ there exists some $x_i \geq 0$ such that $\mu((0, x_i)) \leq \delta$ and $L_\mu^+(x_i) < L_\mu^-(x_i) + \delta.$ Since $y_0 = -1,$ $x_1$ satisfies condition (b). Since $\delta < \Delta_i/3,$ we have $\mu((y_1, x_1)) = \mu((0, x_1)) \leq \delta \leq 2\Delta_i/3 - \delta < \Delta_i - \delta,$ and condition (c) is satisfied. Moreover,

$$L_\mu^-(x_1) \mu((x_1, y_1)) \leq \int_{x_1-1, x_1} \delta \mu(z) \leq \mu((0, x_1)) \leq \delta.$$

Thus, $L_\mu^+(x_1) - L_\mu^-(x_1) + \delta \leq 2\delta < \Delta_i,$ and condition (a) is also satisfied.

We now assume that we have chosen nonnegative $x_1, \ldots, x_i$ and $y_1, \ldots, y_{i-1},$ so that $x_1, \ldots, x_i$ satisfy the four conditions (a)-(d), and $x_i$ satisfies conditions (a)-(c). We will first show that we can choose $y_i$ to satisfy condition (d). Then, if $\mu((y_i, L)) < \epsilon^2/(L + 1),$ we will set $N = i,$ and terminate the construction. Otherwise, we will show that we can choose
$x_{i+1}$ to satisfy conditions (a)-(c), and continue similarly. Note that if $y_L$ has been thus constructed and the process has not yet terminated, then the property $y_{i+1} \geq x_i \geq y_i + 1$ (from conditions (b) and (d)) implies that $y_{L+1} \geq L$, so that $\mu ((y_{L+1}, L]) = 0$, which satisfies the termination condition. This shows that indeed $N \leq L + 1$, as desired. Because all the required conditions will be enforced, this construction will indeed verify our claim.

The argument considers separately two different cases. For the first case, we assume that $\mu ((x_i, x_i + 1)) \leq \delta + \frac{1}{K_i}$, which means that very few agents have opinions between $x_i$ and $x_{i+1}$. The construction described below is illustrated in Figure 7(a). We let $y_i = x_i$, so that condition (d) is trivially satisfied by $y_i$. If $\mu ((y_i, L]) < \epsilon^2/(L+1)$, we let $i = N$ and terminate. Suppose therefore that this is not the case. Then, $y_i < L$, and

$$
\mu ((y_i, L + 1]) = \mu ((y_i, L]) \geq \frac{\epsilon^2}{L + 1} > \Delta_{i+1}.
$$

We then have

$$
\mu ((y_i + 1, L + 1]) = \mu ((y_i + 1, L]) - \mu ((y_i, y_i + 1)) > \Delta_{i+1} - \delta - \frac{1}{K_i} > \delta,
$$

where the last inequality follows from the recurrence $\Delta_{i+1} = 3K_i\Delta_i + \frac{1}{K_i}$ (cf. Lemma 3), and the fact that $2\delta < \Delta_i < K_i\Delta_i$, for all $i$. In particular, we must have $y_i + 1 \leq L$. Using the assumption $|L_x y| \leq \mu \delta$, and an earlier observation, we can find some $x_i + 1 \geq y_i$ such that $\hat{L}_\mu^+(x_i + 1) < \hat{L}_\mu^+(x_{i+1}) + \delta$ and $\mu ((y_i + 1, x_{i+1})) \leq \delta$. Condition (b) then trivially holds for $x_{i+1}$.

Remembering that $y_i = x_i$, we also have

$$
\mu ((y_i, x_{i+1})) = \mu ((y_i, x_i + 1]) + \mu ((y_i + 1, x_{i+1})) \leq \delta + \frac{1}{K_i} + \delta \leq \Delta_{i+1} - \delta,
$$

where the last inequality follows, as before, from the recurrence $\Delta_{i+1} = 3K_i\Delta_i + \frac{1}{K_i}$, and from $\delta \leq K_i\Delta_i$. As a result, $x_{i+1}$ satisfies condition (c). To prove that $x_{i+1}$ satisfies condition (a), observe that

$$
\hat{L}_\mu^+(x_{i+1}) = \int_{(x_{i+1}-1,x_{i+1})} (x_{i+1} - z) d\mu(z) \leq \mu ((x_{i+1} - 1, x_{i+1})) \leq \mu ((x_{i+1} + 1, x_{i+1}))
$$

where the last inequality follows from condition (b) for $x_{i+1}$. Because $x_{i+1}$ satisfies condition (c), we have $\hat{L}_\mu^-(x_{i+1}) \leq \Delta_{i+1} - \delta$. Then, condition (a) for $x_{i+1}$ follows from the fact that $x_{i+1}$ has been chosen so that $\hat{L}_\mu^+(x_{i+1}) < \hat{L}_\mu^-(x_{i+1}) + \delta$.

We now consider the second case, where $\mu ((x_i, x_i + 1)) > \delta + \frac{1}{K_i}$. Our construction is illustrated in Figure 7(b). We claim that $\mu ((x_i + K_i\Delta_i, x_i + 1)) \leq \frac{1}{K_i}$. This is because otherwise we would have

$$
\hat{L}_\mu^+(x_i) = \int_{(x_i,x_i+1]} (z - x_i) d\mu(z) \geq \mu ((x_i + K_i\Delta_i, x_i + 1)) \geq \Delta_i,
$$

contradicting condition (a) for $x_i$. This implies that $\mu ((x_i, x_i + K_i\Delta_i]) > \delta$. It follows that we can choose $y_i \geq x_i$ so that $y_i \leq x_i + K_i\Delta_i$, $\mu ((y_i, x_i + K_i\Delta_i)) \leq \delta$, and $\hat{L}_\mu^+(y_i) \leq \hat{L}_\mu^-(y_i) + \delta$. Then, condition (d) is satisfied by $y_i$. If $\mu ((y_i, L]) < \epsilon^2/(L + 1)$, we let $i = N$ and terminate.

Suppose therefore that this is not the case. By the same argument as for the previous case, we can then choose $x_{i+1}$ so that $y_i + 1 \leq x_{i+1} \leq L$, $\hat{L}_\mu^+(x_{i+1}) \leq \hat{L}_\mu^-(x_{i+1}) + \delta$, and $\mu ((y_i + 1, x_{i+1})) < \delta$. Thus $x_{i+1}$ satisfies condition (b).

To prove that $x_{i+1}$ satisfies the remaining two conditions, (c) and (a), we need an upper bound on $\mu ((x_i + 1, y_i + 1))$. Observe first that

$$
\hat{L}_\mu^+(y_i) = \int_{(y_i,y_i+1]} (z - y_i) d\mu(z) \geq \int_{x_{i+1},y_i+1} (z - y_i) d\mu(z) \geq (1 + x_i - y_i) \mu ((x_i + 1, y_i + 1)) \geq \frac{1}{2} \mu ((x_i + 1, y_i + 1)),
$$

where the last inequality follows from the fact $1 + x_i - y_i \geq 1 - \epsilon$ (condition (d)), and the fact that $\epsilon$ was assumed smaller than $\frac{1}{2}$. Thus, to derive upper bound on $\mu ((x_i + 1, y_i + 1))$ it will suffice to derive an upper bound on $\hat{L}_\mu^+(y_i)$. We start with the inequality, $\hat{L}_\mu^+(y_i) \leq \hat{L}_\mu^-(y_i) + \delta$, and also make use of the fact $y_i - 1 \geq x_i - 1 \geq y_{i-1}$, which is a consequence of conditions (b) and (d) for $x_i$. We obtain

$$
\hat{L}_\mu^-(y_i) \leq \int_{y_{i-1},x_i} (y_i - z) d\mu(z) \leq \mu ((y_i - y_{i-1}, y_i - z)) (y_i - x_i) \leq \Delta_{i-1} - \delta + K_i\Delta_i,
$$

where the last inequality follows from conditions (c) and (d) for $x_i$, and the fact that $\mu ((x_i, y_i)) \leq \mu ([0, L]) = 1$. Combining this with the lower bound (10) leads to

$$
\mu ((x_i + 1, y_i + 1)) \leq 2(K_i + 1)\Delta_i,
$$

which is the desired upper bound on $\mu ((x_i + 1, y_i + 1))$. We will now use the above upper bound to prove conditions (a) and (c) for $x_{i+1}$. Observe that $\mu ((y_i, x_{i+1}))$ can be
This completes the induction and the proof of Theorem 4.

Recall that $y_i$ has been chosen so that $\mu \left( (y_i, x_i + K_i \Delta_i) \right) \leq \delta$, and $x_{i+1}$ so that $\mu \left( (y_i + 1, x_i + 1) \right) \leq \delta$. Moreover, $\mu \left( (x_i + K_i \Delta_i, x_i + 1) \right)$ has been shown to be no greater than $1/K_i$. It then follows from (11) that

$$\mu \left( (y_i, x_i + 1) \right) \leq 2\delta + \frac{1}{K_i} + 2(K_i + 1) \Delta_i \leq 3K_i \Delta_i + \frac{1}{K_i} - \delta,$$

where we have used the facts that $3\delta \leq \Delta_i$ and $K_i \geq 3$.

Condition (c) for $x_{i+1}$ follows from the property $\Delta_{i+1} = 3K_i \Delta_i + \frac{1}{K_i}$ in the definition of the sequence $\{\Delta_i\}$ (see Lemma 3). To prove condition (a), we observe that

$$\hat{\mu} \left( x_{i+1} \right) = \int_{(x_{i+1} - 1, x_{i+1})} \left( x_{i+1} - z \right) d\mu(z) \leq \mu \left( (x_{i+1} - 1, x_{i+1}) \right),$$

and then use conditions (b) and (c) for $x_{i+1}$ to obtain

$$\hat{\mu} \left( x_{i+1} \right) \leq \mu \left( (y_i, x_i + 1) \right) \leq \Delta_i - \delta.$$

This completes the induction and the proof of Theorem 4.

\section*{B. Proof of Theorem 6}

Proof: Suppose $s$ does not satisfy the condition of this theorem, and that $a < b$. Since $s \in F$, we have $\mu_s \left( (a, b) \right) = 0$. Let $S_a, S_b \subset I$ be two sets on which $s$ takes the values $a$ and $b$, respectively. We choose these sets so that the Lebesgue measure of $S_a \cup S_b$ is $\delta$, and so that the ratio $|S_a|/|S_b|$ of their measures is equal to $\mu_s(a)/\mu_s(b)$. Let $x_0(\alpha) = s(\alpha)$ for $\alpha \notin S_a \cup S_b$, and $x_0(\alpha) = \frac{\mu_s(a) + \mu_x(b)}{\mu_s(a) + \mu_x(c)}$ for $\alpha \in S_a \cup S_b$. Observe that $\mu_{x_0}(\alpha) = \mu_s(\alpha)/\mu_s(b)$.

As already discussed, when $x_0$ takes discrete values, the evolution of $x_1$ is entirely characterized by the evolution of a corresponding weighted discrete-agents system of the form (2). We can then apply the reasoning in the proof of Theorem 2 to show that the two clusters initially at $a$ and $b$ converge to a single cluster. Since this can be done for any, arbitrarily small $\delta > 0$, $s$ is unstable.

\section*{C. Proof of Proposition 3}

To ease the reading of the proof, we introduce a new notation. For any $x \in X_L$, we let $u_x : [0, L] \rightarrow [0, L]$ be a function defined so that $u_x(a)$ is the updated opinion of an agent that held opinion $a$, namely

$$u_x(a) = \frac{\int_{a - 1}^{a + 1} z d\mu_x(z)}{\mu_x(a, a + 1)}.$$

As a consequence, $(U(x))'(a) = u_x(x'(a))$ and $x_{i+1}(\alpha) = u_x(x_{i}(\alpha))$, for any $\alpha \in I$.

Proof: Since $x$ is regular, there exist $m$ and $M$, with $0 < m \leq M$, such that for any $[a, b] \subseteq [\inf_x x, \sup_x x]$ we have $m(b - a) \leq \mu_x([a, b]) \leq M(b - a)$. Let $\delta = \min \left\{ \frac{1}{2}, \sup_x x - \inf_x x - 2 \right\}$. We first prove the existence of $M'$, $m' > 0$ such that if $[a, b] \subseteq [\inf_x x, \sup_x x]$, and $b - a < \delta$, then $\mu'(b - a) \leq u_x(b) - u_x(a) \leq M'(b - a)$. (The proof of the upper bound amounts to noting that the numerator and denominator in the definition of $u_x(a)$ are both Lipschitz continuous functions of $a$, and that the denominator is bounded below by $m$. The proof of the lower bound is essentially a strengthening of the proof of Proposition 1, which only established that $u_x(b) - u_x(a) \geq 0$.)

With our choice of $\delta$, we have either $a \geq \inf_x x + 1$ or $b \leq \sup_x x - 1$. We only consider the second case, so that $(a, b + 1) \subseteq [\inf_x x, \sup_x x]$; the first case can be treated similarly. Let $\bar{\mu}_x = \mu_x \left( (a - 1, b + 1) \right)$, $\bar{\mu}_x = x_0 \left( (b - 1, a + 1) \right)$, and $\bar{\mu}_{x(a)} = \mu_x \left( (a + 1, b) \right)$. Let also $x_{a(b)}$, $x_{a(b)}$, and $x_{a(b)}$ be the center of mass of the opinions of those agents whose opinions lie in the set $(a - 1, b - 1)$, $(b - 1, a + 1)$, and $(a + 1, b - 1)$, respectively. In case $\bar{\mu}_{a(b)} = 0$, we use the convention $x_{a(b)} = b - 1$.

From the definition of $u_x$, we have

$$u_x(a) = \frac{\mu_x(x_{a(b)})}{\mu_x(a) + \mu_x(b)} = \frac{\mu_x(x_{a(b)})}{\mu_x(a) + \mu_x(b)} = \frac{\mu_x(a)}{\mu_x(a) + \mu_x(b)}.$$

and

$$u_x(b) = \frac{\mu_x(x_{a(b)})}{\mu_x(a) + \mu_x(b)} = \frac{\mu_x(a)}{\mu_x(a) + \mu_x(b)}.$$

Note that $a - 1 \leq x_{a(b)} \leq \bar{x}_a \leq \bar{b} \leq b + 1$, so that $x_{a(b)} - \bar{x}_a \leq 2(b - a) \leq \delta$, and similarly, $\bar{x}_a - x_{a(b)} \leq \delta$. From the regularity assumption, we also have $\mu_{a(b)} \leq M(b - a)$ and $\mu_{a(b)} = \mu_x \left( (b - 1, a + 1) \right) \geq \mu_x \left( (a + 1, b) \right) \geq m$. Thus,

$$u_x(b) - u_x(a) \leq 3\frac{\mu_{a(b)}}{\mu_{a(b)}} + 3\frac{\mu_{a(b)}}{\mu_{a(b)}} = \frac{3m(b - a)}{m},$$

which proves the claimed upper bound with $M' = 3M/m$.

For the lower bound, an elementary calculation shows that if we have a density function on the interval $[0, 1]$, which is bounded above and below by $M$ and $m$, respectively, then its center of mass is at least $m/2M$. By applying this fact to the interval $(b - 1, a + 1)$ (which has length larger than 1), we conclude that its center of mass, $\bar{x}_a$, is at least $m/2M$ below the right end-point $a + 1$. Since also $\bar{x}_a \geq a + 1$, we have $\bar{x}_a \geq m/2M$, and

$$u_x(b) \geq \bar{x}_a + \frac{\mu_{a(b)}}{\mu_{a(b)}} - \frac{m}{2M} \geq u_x(a) + \frac{(b - a)}{m} \geq \frac{m}{2M}.$$

where the last inequality made use of the facts $\mu_{a(b)} \geq m(b - a)$ and $\mu_{b(a)} \leq 3M$. This establishes the claimed lower bound with $m' = m/2M^2$.

By splitting an interval $[a, b] \subseteq [\inf_x x, \sup_x x]$ into subintervals of length bounded by $\delta$, we see that the result $m'(b - a) \leq u_x(b) - u_x(a) \leq M'(b - a)$ also holds for general such intervals. Consider now an interval $[a', b'] \subseteq [\inf_x U(x), \sup_x U(x)]$, and let $a = \inf \left\{ z \in [0, L] : u_z(z) \in [a', b'] \right\}$ and $b = \sup \left\{ z \in [0, L] : u_z(z) \in [a', b'] \right\}$. As a consequence of the order preservation property, $u_z((a, b)) \subseteq [a', b']$, and $[a', b'] \subseteq [u_z(a), u_z(b)]$. Since $x$ is regular, we have $\mu_z(a) = \mu_z(b) = 0$, which implies that $\mu_{U(z)}(a', b') = \mu_z(a, b) \in [m(b - a), M(b - a)]$. Using the bounds on $\mu_{U(z)}(a', b')$, we finally obtain the desired result

$$mm'(b' - a') \leq \mu_{U(z)}(a', b') \leq M'(b' - a').$$