# The Dynamics of Influence Systems

Bernard Chazelle
*Department of Computer Science*
*Princeton University and Collège de France*
*Princeton, NJ 08540, USA*
*chazelle@cs.princeton.edu*

***Abstract*—Influence systems form a large class of multiagent systems designed to model how influence, broadly defined, spreads across a dynamic network. We build a general analytical framework which we then use to prove that, while Turing-complete, influence dynamics of the diffusive type is almost surely asymptotically periodic. Besides resolving the dynamics of a popular family of multiagent systems, the other contribution of this work is to introduce a new type of renormalization-based bifurcation analysis for multiagent systems.**

## I. Introduction

This paper has three objectives: (i) to bring under one roof a wide variety of popular multiagent systems; (ii) to build an "algorithmic calculus" to help us analyze them; (iii) to resolve the complexity of their "diffusive" restriction. *Influence systems* are discrete-time dynamical systems specified by a map $\mathbf{x} \mapsto f(\mathbf{x})$ from $(\mathbb{R}^d)^n$ to $(\mathbb{R}^d)^n$ and a function $\mathcal{G}$ mapping each $\mathbf{x}$ to an $n$-node graph: the point $\mathbf{x} = (x_1, \ldots, x_n)$ encodes the position $x_i \in \mathbb{R}^d$ of each agent $i = 1, \ldots, n$; the map $\mathbf{x} \mapsto \mathcal{G}(\mathbf{x})$ specifies the *communication graph*, with one node per agent. Each coordinate function $f_i$ of $f = (f_1, \ldots, f_n)$ takes as input the neighbors of agent $i$ in $\mathcal{G}(\mathbf{x})$, together with their locations, and outputs the new position $f_i(\mathbf{x})$ of agent $i$ in $\mathbb{R}^d$. By distinguishing between $\mathcal{G}$ and $f$, the model separates the syntactic (where the information travels across the dynamic network) from the semantic (how it is used by each agent's personal algorithm $f_i$). This distinction reflects the focus on systems in which emergence owes more to the flow of communication among the agents than to the sheer computational power of $f$. A deterministic influence system is called *diffusive* if the map $f$ keeps each agent within the convex hull of its neighbors.

An overarching ambition of social dynamics is to understand and predict the collective behavior of agents influencing one another across an endogenously changing network [10]. Influence systems provide a versatile platform for such investigations [14]. The model includes swarming, synchronization, consensus systems, neural nets, Bayesian social learning, protein interaction networks, the Ising model, etc.[1] Diffusive systems remain bounded and make consensus (all $x_i$ being equal) a fixed point. *HK systems* have emerged in the last decade as a prototypical platform in social dynamics [18]. Diffusive influence systems unify their varied strands (eg, bounded-confidence, bounded-influence, truth-seeking, Friedkin-Johnsen type, deliberative exchange) into a single framework and supply closed-loop analogs to standard consensus models [3], [25], [27].

In a diffusive influence system, $f(\mathbf{x}) = (P(\mathbf{x}) \otimes \mathbf{I}_d\,)\,\mathbf{x}$, where $P(\mathbf{x})$ is a stochastic matrix whose positive entries correspond to the edges of $\mathcal{G}(\mathbf{x})$ and are rationals assumed larger than some arbitrarily small $\rho > 0$; the Kronecker product with the $d$-by-$d$ identity $\mathbf{I}_d$ makes the *transition* matrix $P(\mathbf{x})$ act on $(\mathbb{R}^d)^n$ and not $\mathbb{R}^n$. We grant the agents a measure of self-confidence by adding a self-loop to each node of $\mathcal{G}(\mathbf{x})$. Agent $i$ computes the $i$-th row of $P(\mathbf{x})$ by means of its own algebraic decision tree; that is, on the basis of the signs of a finite number of $dn$-variate polynomials evaluated at the coordinates of $\mathbf{x}$. This high level of generality allows $\mathcal{G}(\mathbf{x})$ to be specified by any first-order sentence over the reals:[2] in a recent bird flocking model [2], for instance, the communication graph joins every agent to its 7 nearest neighbors. We state our main result:[3]

THEOREM 1.1: *Given any initial state, the orbit of an influence system is attracted exponentially fast to a limit cycle whp under an arbitrarily small random perturbation. The period and preperiod are bounded by a polynomial in the reciprocal of the failure probability. Without perturbation, the model is Turing-complete. In the bidirectional case, the system is attracted to a fixed point. The convergence time is $\rho^{-O(n)}|\log \varepsilon|$ whp, where $n$ is the number of agents and $\varepsilon$ is the distance to the fixed point.*

[1] The states of an influence system can be opinions, Bayesian beliefs, neuronal spiking sequences, animal herd locations, chemotactic responses, cell populations, schooling fish velocities, sensor networks data, synchronization phases, heart pacemaker cell signals, cricket chirpings, firefly flashings, yeast cell suspensions, microwave oscillator frequencies, flocking headings, etc [6], [8], [10], [30], [32].

[2] This is the language of geometry and algebra over the reals, with statements specified by any number of quantifiers and polynomial (in)equalities. It was shown to be decidable by Tarski and amenable to quantifier elimination and algebraic cell decomposition by Collins [15].

[3] All influence systems in the remainder of this paper are assumed to be diffusive, so we drop the qualifier; "whp"= with high probability.

0272-5428/12 $26.00 © 2012 IEEE
DOI 10.1109/FOCS.2012.70

*Remarks.* The Turing machine simulation can be done with linear decision trees and $d = 1$. The (infinite) number of limit cycles is actually finite up to foliation. A system is called bidirectional if all the communication graphs are undirected. To perturb the system means: to apply a random shift, ie, to pick a small random $\delta$ and replace each test polynomial $q(\mathbf{x})$ by $q(\mathbf{x}) + \delta$; and to apply a *perturbation rule* stipulating that (a) the status of an edge $(i, j)$ is constant when agents $i, j$ are infinitesimally close to each other; and (b) no edge that disappears indefinitely can return; in both cases, the threshold can be an arbitrary function of $n$, so the perturbation rule is unnecessary in practice. Even in theory it can sometimes be relaxed: for example, (b) is not needed in the bidirectional case. We need to emphasize, however, that some form of perturbation rule is required: without (a, b) or some variant, Theorem 1.1 is provably false; in general, randomization is necessary but not sufficient. Note that the perturbation rule is not a heuristic assumption but a local rule that agents can easily implement. It is not a roundabout way to enforce connectivity either, since agents are given free rein to drop edges at any time. In the context of social dynamics, our results might be disconcerting. Influence systems model how people change opinions over time as a result of human interaction and knowledge acquisition. Strangely, unless people keep varying the modalities of their interactions, as mediated by trust levels, self-confidence, etc, they will be caught forever recycling the same opinions in the same order.

Following their introduction by Sontag [35], piecewise-linear systems have become the subject of an abundant literature, which we do not attempt to review here. Influence systems with undirected communication graphs always converge to a fixed point [13], [17], [20], [25], [27] but convergence times are known only in a few cases [6], [13]. Without bidirectionality, known convergence results are conditional [9], [11], [12], [21], [26]–[28], [31], [36].[4] The standard assumption is that some form of joint connectivity property should hold in perpetuity; as we show below, however, to check such a property is usually undecidable. A significant recent advance was Bruin and Deane's unconditional resolution of planar piecewise contractions, which are special kinds of influence systems with a single mobile agent [5].

Piecewise-linear systems are known to be Turing-complete [1], [4], [22], [34]. A typical simulation relies on the existence of Lyapunov exponents of both signs, negative ones to move the head in one direction and positive ones to move it the other way. Influence systems have no positive exponents and yet are Turing-complete. In dynamics, chaos is typically associated with positive topological entropy, which entails expansion, hence positive Lyapunov exponents. But piecewise linearity blurs this picture. With only null Lyapunov exponents, isometries are not chaotic [7] but contractions, with only negative exponents, can be [23]. Influence systems, which, with only null and negative Lyapunov exponents, sit in the middle, can be chaotic. Plainly, the spectral lens breaks down in the face of piecewise linearity and calls for a different approach: we use an algorithmic brand of bifurcation analysis.

[4] As they should be, since convergence is not assured. An exception is *truth-seeking HK systems*, which have been shown to converge unconditionally [13], [19], [24].

## II. Preliminaries

We show in §II-A that influence systems can have periodic orbits of length exponential in the number of agents: this result is resistant to perturbation. Quite the opposite, the next two results require careful finetuning. In §II-B, we build a conjugation with the baker's map to exhibit chaos and, in §II-C, we show how to simulate a Turing machine. All three constructions use linear decision trees. This is not surprising in view of §II-D, where we show how to linearize the decision procedure of any influence system.

### A. Long periods

Periodic orbits can be made arbitrarily long by increasing the bit-length of the encoding. More interesting is the fact that exponential periods can be achieved with only logarithmic bit-length. We simulate a counter modulo 2 by building a system with $d = 1$ and $n = 3$: the first two agents are fixed at 0 and 3 while the third oscillates between positions 1 and 2; this is trivially achieved with a two-test linear decision tree. Add another mobile agent oscillating between 1 and 2 like the previous one, but which moves only when the first oscillating agent is at position 1. (Adding a single test makes this possible.) Iterating in this fashion produces an $n$-agent influence system with $O(n)$ tests whose period is exactly $2^{n-2}$.

### B. Why perturbation is necessary

Random shifts are required for any uniform convergence bound. To see why, set $d = 1$ and $n = 3$. The first two agents move toward each other according to the rule:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \overset{f}{\longmapsto} \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{1}$$

Starting at positions $-1$ and 1, agents 1 and 2 move to $\pm 3^{-t}$ at time $t$. Imagine now a third agent starting at position $0.9 < x_3 < 1$ and set to join with agent 1 when their distance is no more than one: this happens after on the order of $|\log(1 - x_3)|$ steps. The convergence time goes to infinity as $x_3$ approaches 1, indicating the impossibility of a uniform bound.

We claimed earlier than random shifting is not enough and a perturbation rule is needed. To see why, we set $d = 1$ and $n = 4$. The first two agents stay on opposite sides of the origin, with the agent further from it moving toward it while the other one stays put:

$$(x_1, x_2) \overset{f}{\longmapsto} \frac{1}{2}\begin{cases} (\,2x_1, x_1 + x_2\,) & \text{if } x_1 + x_2 \geq 0 \\ (\,x_1 + x_2, 2x_2\,) & \text{else.} \end{cases}$$

The two agents converge toward 0 but the order in which they proceed (ie, their symbolic dynamics) is chaotic. Let $x_i(t)$ be the position of agent $i$ at time $t$. Assume that $x_1(0) < 0 < x_2(0)$ and consider the trajectory of a line $L$: $X_2 = uX_1$, for $u < 0$. If the point $(x_1(t), x_2(t))$ is on the line, then $x_1(t) + x_2(t) \geq 0$ implies that $u \leq -1$ and $L$ is mapped to $X_2 = \frac{1}{2}(u+1)X_1$; if $x_1(t) + x_2(t) < 0$, then $u > -1$ and $L$ becomes $X_2 = \frac{2u}{u+1}X_1$. The parameter $u$ obeys the dynamics: $u \mapsto \frac{1}{2}(u+1)$ if $u \leq -1$ and $u \mapsto 2u/(u+1)$ if $-1 < u \leq 0$. Writing $u = (v+1)/(v-1)$ gives $v \mapsto 2v+1$ if $v < 0$ and $v \mapsto 2v-1$ else. The system $v$ escapes for $|v(0)| > 1$ and otherwise conjugates with the baker's map [16]. To turn this into actual chaos, the third agent oscillates in $[x_1, x_4] \approx [0, 1]$, with $x_4 = 1$, depending on the order in which the first two agents move: $x_3 \mapsto \frac{1}{3}(x_3 + 2x_1)$ if $x_1 + x_2 \geq 0$ and $x_3 \mapsto \frac{1}{3}(x_3 + 2x_4)$ else. Agent 3 is either at most 0.4 or at least 0.6 depending on which of agent 1 or 2 moves. This implies that the system has positive topological entropy: to know where agent 3 is at time $t$ requires on the order of $t$ bits of accuracy in the initial state. We easily check that no random shift can prevent this and a perturbation rule is indeed necessary to prevent chaos.

### C. Turing completeness

Absent perturbation, an influence system can simulate a general piecewise-linear system and hence a Turing machine. We show how this is done. Given a nonzero $n$-by-$n$ real-valued matrix $A$, let $A^+$ (resp. $A^-$) be the matrix obtained by zeroing out the negative entries of $A$ (resp. $-A$), so that $A = A^+ - A^-$. Define the matrices

$$B = r\begin{pmatrix} A^+ & A^- \\ A^- & A^+ \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} B & (\mathbf{I}_{2n} - B)\mathbf{1} & \mathbf{0} \\ \mathbf{0} & 1 & 0 \\ \mathbf{0} & 1-r & r \end{pmatrix},$$

where $r = \min_i\{1, 1/\sum_j |A_{ij}|\}$. It is immediate that $C$ is stochastic and semiconjugates with the dynamics of $A$ (up to scaling). Indeed, given $\mathbf{x} \in \mathbb{R}^n$, if $\overline{\mathbf{x}}$ denotes the $(2n+2)$-dimensional column vector $(\mathbf{x}, -\mathbf{x}, 0, 1)$, then $C\,\overline{\mathbf{x}} = r\overline{A\mathbf{x}}$; hence the commutative diagram:

$$\begin{array}{ccc} \mathbf{x} & \longrightarrow & A\mathbf{x} \\ \big\downarrow & & \big\downarrow \\ \overline{\mathbf{x}} & \longrightarrow & r^{-1}C\,\overline{\mathbf{x}}\,. \end{array}$$

Imagine now a piecewise-linear system consisting of a number of matrices $\{A_k\}$ and a hyperplane arrangement with a matrix $A_k$ associated with each cell.[5] We add $n$ negated clones to the existing set of $n$ agents, plus a *stochasticity* agent permanently positioned at $x_{-1} = 0$ as well as a *projectivity* agent initialized at $x_0$. This allows us to form the vector $\overline{\mathbf{x}} = (\mathbf{x}, -\mathbf{x}, x_{-1}, x_0)$. The system scales down, so we rewrite any hyperplane $\mathbf{a}^T\mathbf{x} = a_0$ with homogeneous coordinates as $\mathbf{a}^T\mathbf{x} = a_0x_0$. We can use the same value of $r$ throughout by picking the smallest one among all the matrices $A_k$ used in the piecewise-linear system.

Koiran et al [22] have shown how to simulate a Turing machine with a 3-agent piecewise-linear system, so we set $n = 3$. We need an *output* agent to indicate whether the system is in an accepting state: this is done by pointing to one of two fixed agents. We can enlist one of the three original agents for that purpose, which keeps the total agent count below 10. Predicting nontrivial state properties of an influence system (such as basic connectivity properties of the communication graph) is therefore undecidable.

### D. Linearization

Beginning with the case $d = 1$, we can write $\mathbf{x}$ more simply as $(x_1, \ldots, x_n) \in \mathbb{R}^n$. We show how to linearize an influence system by tensor powering. Let $\mathtt{d}$ be the maximum total degree[6] of the polynomial tests used in the algebraic decision trees (recall that each agent comes equipped with its own). We can always assume the existence of an agent confined to position 1 with no in/out-link: we use it to homogeneize the test polynomials, so that every monomial has degree exactly $\mathtt{d}$. We define the monomial $y_{k_1,\ldots,k_{\mathtt{d}}} = \prod_{i=1}^{\mathtt{d}} x_{k_i}$ $(1 \leq k_1, \ldots, k_{\mathtt{d}} \leq n)$ and, listing them in lexicographic order, form $\mathbf{y} = (y_{k_1,\ldots,k_{\mathtt{d}}}) \in \mathbb{R}^N$, where $N = n^{\mathtt{d}}$; note that $\mathbf{y}$ lies on a (real) algebraic variety $\mathcal{V}$ smoothly parametrized injectively by $\mathbf{x}$. The map $\mathbf{x} \mapsto f(\mathbf{x})$ induces the lifted map $\mathbf{y} \mapsto g(\mathbf{y})$, where $g(\mathbf{y}) = P(\mathbf{x})^{\otimes \mathtt{d}}\,\mathbf{y}$ and

$$P(\mathbf{x})^{\otimes \mathtt{d}} \;=\; \overbrace{P(\mathbf{x}) \otimes \cdots \otimes P(\mathbf{x})}^{\mathtt{d}}.$$

Being the Kronecker product of stochastic matrices, $P(\mathbf{x})^{\otimes \mathtt{d}}$ is stochastic: its diagonal is positive and its nonzero entries all exceed $\rho^{\mathtt{d}}$. Its associated graph, whose edges map out its nonzero entries, is the tensor graph product $\mathcal{G}(\mathbf{x})^{\otimes \mathtt{d}}$. We use the term *ground agents* to refer to the $n$ agents positioned at $\mathbf{x}$. Including all the test polynomials from all the ground agents' decision trees gives us as many hyperplanes in $\mathbb{R}^N$ and the sign conditions of a cell $c$ specify a unique stochastic matrix $Q_c$. This matrix is always a tensor power $P^{\otimes \mathtt{d}}$ but it is guaranteed to be of the form $P(\mathbf{x})^{\otimes \mathtt{d}}$ only if $c$ contains a point $\mathbf{y}$ of $V$ parametrized by $\mathbf{x}$.

Whereas a random shift produces affine forms $a_1y_1 + \cdots + a_Ny_N + \delta$, the perturbation rule acts in a more subtle

[5] A cell is the solution set of any collection (finite or infinite) of linear (strict or nonstrict) inequalities. If it lies in an affine subspace of dimension $k$ but not $k-1$, it is called a $k$-cell.

[6] Not to be confused with $d$.

way. While the whole point of the lifting is to forget about the variety $\mathcal{V}$, the tensor structure of the matrices $Q_c$ brings benefits we will want to exploit. Given $K \subseteq \{1, \dots, n\}$, the *cluster* $C_K$ refers to the subset of $|K|^{\mathtt{d}}$ agents with labels in $K^{\mathtt{d}}$. If all the agents of a cluster fit within a tiny interval then so do their ground agents; to see why, just expand $(x_i - x_j)^{\mathtt{d}}$. By the perturbation rule, therefore, the induced subgraph of the cluster cannot change until it is pulled apart by outside agents. We revisit this point below in greater detail. Assume now that $d > 1$. We write

$$\mathbf{x} = (x_{1,1}, \dots, x_{1,d}, \dots, x_{n,1}, \dots, x_{n,d}),$$

with the homogeneizing agent 1 permanently positioned at $(x_{1,1}, \dots, x_{1,d}) = \mathbf{1}_d$. Next, we define $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$, where $N = (dn)^{\mathtt{d}}$ and $\mathbf{y}_l = \prod_{i=1}^{\mathtt{d}} x_{k_i, j_i}$ with $l$ denoting the lexicographic rank of the string $(k_1, j_1, \dots, k_{\mathtt{d}}, j_{\mathtt{d}})$ for $k_i \in \{1, \dots, n\}$ and $j_i \in \{1, \dots, d\}$. The matrix $Q_c$ associated with cell $c$ is of the form $(P \otimes \mathbf{I}_d)^{\otimes \mathtt{d}}$; furthermore, $P = P(\mathbf{x})$ whenever $\mathbf{y}$ satisfies the $N$ conditions $\mathbf{y}_l = \prod_{i=1}^{\mathtt{d}} x_{k_i, j_i}$ for some $\mathbf{x} \in \mathbb{R}^{dn}$. The cluster $C_K$ consists now of $(d|K|)^{\mathtt{d}}$ agents. For notational simplicity, we assume that $d$ and $\mathtt{d}$ are constants although no such requirement is actually required.

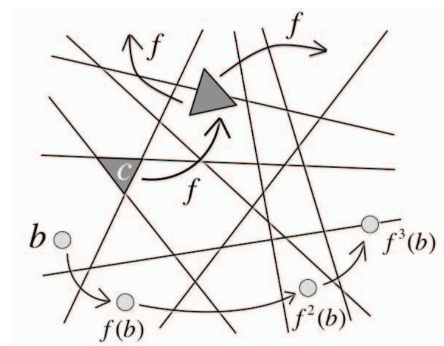
## III. An Algorithmic Calculus

We assume that $P(\mathbf{x}) = P_c$, for any $\mathbf{x} \in c$, where $c$ is any *atom* (open $n$-cell) of an arrangement of hyperplanes in $\mathbb{R}^n$, called the *switching partition* (*SP*). Given a shift $\delta$, we define the *margin*

$$\mathcal{M}_\delta = \bigcup_{SP} \{ \, \mathbf{x} \in \mathbb{R}^n \, | \, \mathbf{a}^T \mathbf{x} = 1 + \delta \, \}, \tag{2}$$

over all the hyperplanes $\mathbf{a}^T \mathbf{x} = 0$ of the switching partition. Given an atom $c$ of $\mathcal{M}_\delta$, the stochastic matrix $P_c = (\, \underline{P}_c \otimes \mathbf{I}_d \,)^{\otimes \mathtt{d}}$ is a tensor power of a *ground matrix* $\underline{P}_c$. We assume that all the relevant parameters (matrix entries, number and coefficients of hyperplanes, $\rho$, etc) can be encoded as rationals over $O(\log n)$ bits: this assumption can be freely relaxed—in fact, the bit lengths can be arbitrarily large as a function of $n$—and is only made to simplify the notation.

As in statistical mechanics, the system's complexity arises from the tension between two opposing forces: one, caused by the map's discontinuities, is "entropic" and leads to chaos; the other one, related to the Lyapunov exponents, is "energetic" and pulls the system toward an attracting manifold within which the dynamics is periodic. The goal is to show that, outside a vanishingly small critical region in parameter space, entropy always loses. What does it mean? If the iterated image of any ball $b$ never intersected the *SP* hyperplanes, as is easily shown, it would bounce around until eventually periodicity kicked in. In the figure below, however, $f^3(b)$ refuses to follow this script and splits into two smaller bodies. Both of them will bounce around until possibly splitting again and so on. If this branching, "entropic" process gets out of control, chaos will ensue. To squelch it, we can count on the paracontractivity of the map, which causes the ball $b$ to shrink—at least in directions outside the dominant eigenspace (alas of arbitrary dimension)—and thus dissipate a form of "energy." Entropy vs energy: which one will win? For entropy to lose out, the ball $b$ must avoid splitting too frequently. This can be expressed by an (infinite) system of linear inequalities. Feasibility then hinges on a type of *matrix rigidity* question: in this case, given a certain matrix, how many rows must be removed before we can express the first column as a linear combinations of the others? The matrix in question is extracted from the system's stochastic matrices and the *SP* equations and hence is highly structured: this is the key to order.




### A. Phase space refinement

By scale invariance and convexity, we may confine the phase space to the open unit box $\Omega = (0,1)^n$. It is useful to classify the initial states by how long it takes their orbits to hit the margin $\mathcal{M}_\delta$, if ever. With $f^0 = \mathbf{I}_n$ and $\min \emptyset = \infty$, we define the label $\ell(\mathbf{x})$ of $\mathbf{x} \in \Omega$ as the minimum integer $t$ such that $f^t(\mathbf{x}) \in \mathcal{M}_\delta$. The point $\mathbf{x}$ is said to *vanish* at time $\ell(\mathbf{x})$ if its label is finite. The points that do not vanish before time $t$ form the set $\mathcal{S}_t$: we have $\mathcal{S}_0 = \Omega$; and, for $t > 0$,

$$\mathcal{S}_t = \Omega \setminus \bigcup_{k=0}^{t-1} f^{-k}(\mathcal{M}_\delta) \, .$$

We impose the condition $\delta > -1$ to keep the preimages of the hyperplanes of $\mathcal{M}_\delta$ empty or of codimension one, which implies that the volume of $\mathcal{S}_t$ is always 1. Each of $\mathcal{S}_t$'s connected components is specified by a set of strict linear inequalities in $\mathbb{R}^n$, so $\mathcal{S}_t$ is a union of disjoint open $n$-cells, whose number we denote by $\#\mathcal{S}_t$. Each cell of $\mathcal{S}_{t+1}$ lies within a cell of $\mathcal{S}_t$. The limit set $\mathcal{S}_\infty = \bigcap_{t \geq 0} \mathcal{S}_t$ collects the points that never vanish. We say that the system is *nesting at $t$* if $\mathcal{S}_t = \mathcal{S}_{t+1}$. The minimum value of $t$ (or $\infty$) is called the *nesting time* $\nu$ of the system. Observe that labels cannot be skipped: if $k$ is a label, then so is $k-1$. The following facts follow easily from this observation.

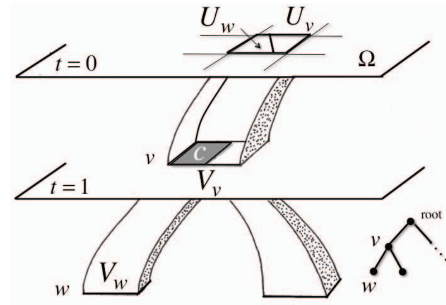
LEMMA 3.1: *The nesting time $\nu$ is the minimum $t$ such that, for each cell $c$ of $\mathcal{S}_t$, $f^t(c)$ lies within an atom. If $c$ is a cell of $\mathcal{S}_\nu$, then $f(c)$ intersects at most one cell of $\mathcal{S}_\nu$ and $\mathcal{S}_\nu = \mathcal{S}_\infty$. Any nonvanishing orbit is eventually periodic and the sum of its period and preperiod is bounded by $\#\mathcal{S}_\nu$.*

We define the directed graph $F$ with one node per cell $c$ of $\mathcal{S}_\nu$ and an edge from $(c, c')$, where $c'$ is the unique cell of $\mathcal{S}_\nu$, if it exists, that intersects $f(c)$. The edge $(c, c')$ is labeled by the linear map $f_{|c}$ defined by the matrix $P_a$, where $a$ is the unique atom $a \supseteq c$. The graph defines a *sofic shift* (ie, a regular language) of the functional kind, meaning that each node has exactly one outgoing edge, possibly a self-loop, so any infinite path leads to a cycle. Periodicity follows immediately. The *trajectory* of a point $\mathbf{x}$ is the string $s(\mathbf{x}) = c_0 c_1 \cdots$ of atoms that its orbit visits: $f^t(\mathbf{x}) \in c_t$ for all $0 \leq t < \ell(\mathbf{x})$. It is infinite if and only if $\mathbf{x}$ does not vanish, so all infinite trajectories are eventually periodic. A serious obstacle is that influence systems are rarely nesting. Some points can take infinitely long to vanish. In the 2-agent system (1), for example, the margin $\mathcal{M}_\delta$ consisting of the line $x_3 - x_1 = 1 + \delta$ yields an infinite cell decomposition $\mathcal{S}_\infty$; this holds for any $\delta$, so randomization is of no help. There are two solutions: one is to thicken the margin by a tiny amount; the other is to break up the phase space into invariant manifolds and argue that most of them are "good" in a technical sense. We follow the latter approach.

### B. The coding tree

The previous discussion hints at the tree structure of the space of orbits. We explore this idea further. The *coding tree* $\mathcal{T}$ encodes into one geometric object the set of all orbits and the full symbolic dynamics. It is the system's "Rosetta stone," from which everything of interest can be read off. Intuitively, the tree divides up the phase space into maximal regions over which the iterated map is linear. It is embedded in $\Omega \times \mathbb{N}$, with the last dimension representing time. Each child $v$ of the root is associated with an atom $U_v$. The *phase tube* $(U_v, V_v)$ of each child $v$ is the "time cylinder" whose cross-sections at times 0 and 1 are $U_v$ and $V_v = f(U_v)$, respectively. In general, a phase tube is a discontinuity-avoiding sequence of iterated images of a given cell in phase space.

The coding tree $\mathcal{T}$ is built recursively by subdividing $V_v$ into the cells $c$ formed by its intersection with the atoms, and attaching a new child $w$ for each $c$: we set $V_w = f(c)$ and $U_w = U_v \cap f^{-t_v}(c)$, where $t_v$ is the depth of $v$. Whereas $U_v$ is always an open $n$-cell, $V_v$ and $c$ can be of lower dimension. By $\delta > -1$, the cell $V_v$ cannot lie inside the margin, so at least one cell $c$ exists and the coding tree has no leaves. We denote by $P_w$ the matrix of the map's




restriction to $c$. The phase tube $(U_v, V_v)$ consists of all the cylinders whose cross-sections at $t = 0, \ldots, t_v$ are, respectively, $U_v, f(U_v), \ldots, f^{t_v}(U_v) = V_v$.

Let $ww'w'' \cdots$ denote the upward, $t_w$-node path from $w$ to the root (but excluding the root). Using the notation $P_{\leq w} = P_w P_{w'} P_{w''} \cdots$, we have the identities $V_w = P_{\leq w} U_w$ and $\mathcal{S}_k = \bigcup_w \{ U_w \,|\, t_w = k \}$, with $\mathcal{S}_k \supseteq \mathcal{S}_{k+1}$. Labeling each node $w$ by the atom that contains the cell $c$ allows us to interpret any path as the prefix of a trajectory and define the language $L(\mathcal{T})$ of all such words. Each infinite path $v_0, v_1, v_2, \ldots$ down the tree has its own limit cell $\bigcap_{t \geq 0} U_{v_t}$ which, unlike those of $\mathcal{S}_t$, might not always be open: collectively, they form the cells of $\mathcal{S}_\infty$.

- The *nesting time* $\nu = \nu(\mathcal{T})$ is the minimum depth at which all nodes have a single child (Lemma 3.1); the number can be infinite. A node $v$ is *deep* if $t_v > \nu$ and *shallow* otherwise.
- The *word-entropy* $h(\mathcal{T})$ expresses the growth rate of the language $L(\mathcal{T})$: it is defined as the logarithm of the number of shallow nodes; $\#\mathcal{S}_\nu \leq 2^{h(\mathcal{T})}$.

We need additional parameters, such as the attraction rate and the augmented word-entropy, but we postpone their introduction. Later, we will randomize $\delta$ within a small interval $\Delta$, so it is useful to define the *global coding tree* $\mathcal{T}^\Delta$ as the coding tree derived from the system $(\mathbf{x}, \delta) \mapsto (f(\mathbf{x}), \delta)$, with the phase space $\Omega \times \Delta$. The sets $\mathcal{M}_\delta$, $U_v$ and $V_v$ are now polyhedra in $\mathbb{R}^{n+1}$.

### C. The arborator

We assemble the coding tree by glueing together smaller coding trees defined recursively. We entrust this task to the *arborator*, a recursive algorithm expressed in a language for "lego-like" assembly. The arborator needs two (infinite) sets of parameters to do its job, the *coupling times* and the *renormalization scales*. To produce these numbers, we use the *flow tracker*, which is a form of breadth-first search for dynamic graphs. The arborator relies on a few primitives that we now describe. The direct sum and direct product are tensor-like operations that we use to assemble the coding tree from smaller pieces. We can also compile a *dictionary* to keep track of the tree's parameters (nesting time, word-entropy, etc) as we build it up one piece at a time.

*Direct sum:* The coding tree $\mathcal{T} = \mathcal{T}_1 \oplus \mathcal{T}_2$ models two independent systems of size $n_1$ and $n_2$. The phase space of the direct sum is of dimension $n = n_1 + n_2$. A path $w_0, w_1, \ldots$ of $\mathcal{T}$ is a pairing of paths in the constituent trees: the node $w_t$ is of the form $(u_t, v_t)$, where $u_t$ (resp. $v_t$) is a node of $\mathcal{T}_1$ (resp. $\mathcal{T}_2$) at depth $t$. The direct sum is commutative and associative; furthermore, $U_w = U_u \times U_v$, $V_w = V_u \times V_v$, and $P_w = P_u \oplus P_v$.

*Direct product:* We begin with a few words of intuition. Consider two systems $S_1$ and $S_2$, governed by different dynamics yet evolving in the same phase space $\Omega$. Given an arbitrary region $\Lambda \subset \Omega$, define the hybrid system $S$ with the dynamics of $S_2$ over $\Lambda$ and $S_1$ elsewhere. Suppose we had complete knowledge of the coding tree $\mathcal{T}_i$ of each $S_i$ ($i = 1, 2$). Could we then combine them in some ways in cut-and-paste style to assemble the coding tree $\mathcal{T}$ of $S$? The direct product $\mathcal{T}_1 \otimes \mathcal{T}_2$ provides the answer. The operation is associative but (being chronological) not commutative. It begins by marking certain nodes of $\mathcal{T}_1$ as *absorbed* and pruning the subtrees below. This operation is called *absorption* by analogy with the absorbing states of a Markov chain: any orbit reaching an absorbed leaf comes to a halt, broken only after we reattach a copy of $\mathcal{T}_2$ at that leaf. The copy must be properly cropped.

*Renormalization:* Directs sums model independent subsystems through parallel composition. Direct products model sequential composition. What are the benefits? In pursuit of some form of contractivity, the flow tracker (discussed below) classifies the communication graphs by their connectivity properties and breaks up orbits into sequential segments accordingly. It partitions the set of stochastic matrices into classes and decompose the coding tree $\mathcal{T}$ into maximal subtrees consisting of nodes $v$ with matrices $P_v$ from the same class. The power of this "renormalization" procedure is that it can be repeated recursively. We classify the ground communication graphs by their *block-directionality* type: $\mathcal{G}(\mathbf{x})$ is of type $m \to n-m$ if the agents can be partitioned into $A, B$ ($|A| = m$) so that no $B$-agent ever links to an $A$-agent; if in addition, no $A$-agent links to any $B$-agent, $\mathcal{G}(\mathbf{x})$ is of type $m \,\|\, n-m$.

### D. The flow tracker

A little imagery will help. Suppose that $m < n$. Pour water on the $B$-agents while keeping the $A$-agents dry. Whenever an edge of the communication graph links a dry agent to a wet one, the former gets wet; note how the water flows in the *reverse* direction of the edges. As soon as all agents become wet (if ever), dry them but leave the $B$-agents wet; repeat. The case $m = n$ is identical, with one agent designated wet once and for all. The sequence of times at which water spreads or drying occurs plays a central role in building the arborator. Assume that $n > 1$ and $0 < m \le n$ from now on. Let $\mathcal{T}_{m \to n-m}$ denote the coding tree of a block-directional system of type $m \to n-m$: we assume inheritance, so it can also be written, albeit less informatively, as $\mathcal{T}_n$. Likewise, $\mathcal{T}_m \oplus \mathcal{T}_{n-m}$ can be expressed as $\mathcal{T}_{m \,\|\, n-m}$ but the converse is not true. When the initial state $\mathbf{x}$ is undersood, we use the shorthand $G_t = \mathcal{G}(f^t(\mathbf{x}))$ to designate the communication graph at time $t$ and we denote by $W_t$ the set of wet agents at that time. The flow tracker monitors communication among the ground agents: information exchanges among lifted agents are implied.

**[1]** $t_0 \leftarrow 0$.
**[2]** Repeat forever:
- **[2.1]** If $m < n$ then $W_{t_0} \leftarrow \{m+1, \ldots, n\}$ else $W_{t_0} \leftarrow \{1\}$.
- **[2.2]** For $t = t_0, t_0+1, \ldots, \infty$
  $W_{t+1} \leftarrow W_t \cup \{\, i \mid \exists\, (i,j) \in G_t \ \&\ j \in W_t \,\}$.
- **[2.3]** If $|W_\infty| = n$ then $t_0 \leftarrow \min\{\, t > t_0 : |W_t| = n \,\}$ else stop.

The set $W_t$ of wet agents is never empty. The assignments of $t_0$ in step [2.3] divide the timeline into *epochs*, time intervals during which either all agents become wet or, failing that, the flow tracker comes to a halt. Each epoch is itself divided into subintervals by the *coupling times* $t_1 < \cdots < t_\ell$, such that $W_{t_k} \subset W_{t_k+1}$. The last coupling time $t_\ell$ marks either the end of the flow tracking (if not all $A$-agents become wet) or one less than the next value of $t_0$ in the loop.

If we define the renormalization scale $w_k = |W_{t_k+1}| - n + m$ for $k = 1, \ldots, \ell - 1$, any path of the coding tree can be expressed as $\mathcal{T}_{m \to n-m} \implies$

$$\underline{\mathcal{T}_{m \,\|\, n-m}}_{\,|\, t_1} \otimes \mathcal{T}_n^{|1} \otimes \\ \Big\{ \bigotimes_{k=1}^{\ell-1} \Big( \underline{\mathcal{T}_{w_k \to n-w_k}}_{\,|\, t_{k+1}-t_k-1} \otimes \mathcal{T}_n^{|1} \Big) \Big\} \otimes \mathcal{T}_{m \to n-m}\,. \tag{3}$$

The expression above describes a maximal (infinite) path of the coding tree. Recursion operates in two distinct ways: first, via the rewriting rule $\mathcal{T}_{m \to n-m} \Rightarrow \cdots \{\,\} \otimes \mathcal{T}_{m \to n-m}$; second, through calls to the inductively smaller subsystems $\mathcal{T}_{w_k \to n-w_k}$. All these derivations extend easily to the global coding trees.

## IV. Bidirectional Systems

We prove Theorem 1.1 for undirected communication graphs. We run the flow tracker with respect to the ground agents and their communication graphs. This induces wetness among the actual agents (in lifted space) in the obvious way: if $W_t$ is the set of ground agents that are wet at time $t$, the cluster $C_{W_t}$ consists of the $(dn)^{\mathsf{d}}$ wet agents. We use the perturbation space to $\Delta = (0, n^{-b})$, where $b$

is a suitably large constant (the higher $b$ the smaller the perturbation). We only need part (a) of the perturbation rule: the status of an edge between two ground agents apart by at most $n^{-b}$ is fixed and independent of the other agents.[7] Let *diam* $(s)$ be the diameter of the system after the $s$-th epoch. If $\|W_t\|$ denotes the length of the smallest interval enclosing $W_t$, it can be easily shown by induction that $\|W_{t_k+1}\| \leq 1 - \rho^{O(k)}$ (see (14) in [13]), where $\rho$ is the smallest nonzero entry among the ground matrices. We conclude that water propagation to all the agents entails the shrinking of the system's diameter by at least a factor of $1 - \rho^{O(n)}$. Since an epoch witnesses the wetting of all the agents, repeated applications of this principle yields

$$\textit{diam}\,(s) \leq e^{-s\rho^{O(n)}}. \tag{4}$$

After $\rho^{-cn}$ epochs have elapsed (if ever), for a large enough constant $c$, the diameter of the system falls beneath $n^{-b}$ and, by convexity, never rises again. By the perturbation rule, the communication subgraph is now frozen and can no longer change. Fix the initial (ground) state $\mathbf{x} \in \Omega$ once and for all. The sets $U_v$ and $V_v$ become open intervals of $\Delta$, so a node $v$ has at most $n^{O(1)}$ children. With the outer product enumerating the first $\rho^{-O(n)}$ epochs leading to the combinatorial "freezing" of the system, we rewrite (3) as: $\mathcal{T}_n^{\Delta} \Longrightarrow$

$$\Big\{ \bigotimes_{s=1}^{\rho^{-O(n)}} \bigotimes_{k=1}^{\ell_s - 1} \Big( \underline{\mathcal{T}^{\Delta}_{w_k \,\|\, n-w_k}}_{\,|\, t_{k+1}-t_k-1} \otimes \mathcal{T}_n^{|1} \Big) \Big\} \otimes \mathcal{T}_n^{*}. \tag{5}$$

Note that $w_k = w_k(s)$, $t_k = t_k(s)$. A single communication graph is associated with $\mathcal{T}_n^*$, hence a fixed matrix $P = P(\mathbf{x})$. The rewriting rule in (5) produces terms of the form $\mathcal{T}^{\Delta}_{w_1\|w_2\|\ldots\|w_k}$, where $\sum w_i = n$. To keep the notation simple, we denote by $\mathcal{T}^{\Delta}_{\|w}$ any such coding tree, with $w = \max\{w_i\}$: the $n$ ground agents are partitioned into groups of size at most $w$ with no edges between them; the status of an edge may depend on all the ground agents, so the system is not a direct sum. Thus,

$$\mathcal{T}^{\Delta}_{\|w} \Longrightarrow \Big\{ \bigotimes_{s=1}^{\rho^{-O(w)}} \big( \mathcal{T}^{\Delta}_{\|w-1} \otimes \mathcal{T}_n^{|1} \big) \Big\} \otimes \mathcal{T}^{*}_{\|w}, \tag{6}$$

where the matrix $P$ for $\mathcal{T}^*_{\|w}$ is of the form $\oplus_i P_i$, with each $P_i$ at most $w$-by-$w$. (The rank of $P$ is at least $n/w$ and possibly much bigger.) By basic Markov chain theory and (4), there exists another matrix $\Pi = \Pi(P)$ such that $\|P^k - \Pi\|_{\max} = e^{-k\rho^{-O(n)}}$, for any $k \geq 0$. Let $\mu(\mathcal{T}^{\Delta}_{\|w})$ be the (maximum) time at which the direct product with $\mathcal{T}^*_{\|w}$ (or earlier absorption) can take place. Given any small $\varepsilon > 0$, there is a time $\theta_\varepsilon(\mathcal{T}^{\Delta}_{\|w})$, the *attraction rate*, after which $f^t(\mathbf{x})$ is forever confined to a ball of radius $\varepsilon$, where

$$\theta_\varepsilon(\mathcal{T}^{\Delta}_{\|w}) \leq \mu(\mathcal{T}^{\Delta}_{\|w}) + \rho^{-O(n)}|\log \varepsilon|. \tag{7}$$

Removing from $\Delta$ a mere $n^{O(1)}$ intervals of length $n^{O(1)}\varepsilon$ is sufficient to form a new set $\Delta' \subseteq \Delta$ such that $\mathcal{T}^*_{\|w}$ witnesses no inter-group communication after $\rho^{-O(n)}|\log\varepsilon|$ steps: this follows from the observation that, in the $n^{O(1)}$ equations, $\mathbf{a}^T\mathbf{x} = 1+\delta$, of the margin (2), the left-hand side can vary by at most $n^{O(1)}\varepsilon$. Extending this idea to all of the renormalized trees in (6) leads to $\underline{\Delta} \subseteq \Delta$ such that: $\nu(\mathcal{T}^{\underline{\Delta}}_{\|n}) \leq \theta_\varepsilon(\mathcal{T}^{\underline{\Delta}}_{\|n})$ and, by (7),

$$\mu(\mathcal{T}^{\underline{\Delta}}_{\|n}) \leq \rho^{-O(n)}\mu(\mathcal{T}^{\underline{\Delta}}_{\|n-1}) + \rho^{-O(n)}|\log\varepsilon| \leq \rho^{-O(n^2)}|\log\varepsilon|.$$

We prove that this holds almost surely by showing that $\Delta \backslash \underline{\Delta}$ is of arbitrarily small measure. For this, it is convenient to define the *augmented* word-entropy $\underline{h}(\mathcal{T}^{\underline{\Delta}}_{\|n})$ to be the logarithm of the (maximum) number of nodes of depth at most $\theta_\varepsilon(\mathcal{T}^{\underline{\Delta}}_{\|n})$. Since no absorption occurs at higher depths, quasi-subadditivity obtains:

$$\underline{h}(\mathcal{T}_1^{\underline{\Delta}} \otimes \mathcal{T}_2^{\underline{\Delta}}) \leq \underline{h}(\mathcal{T}_1^{\underline{\Delta}}) + \underline{h}(\mathcal{T}_2^{\underline{\Delta}}) + \log \texttt{max-degree}\,(\mathcal{T}_1^{\underline{\Delta}});$$

hence, for $\varepsilon$ small enough,

$$\underline{h}(\mathcal{T}^{\underline{\Delta}}_{\|n}) \leq \rho^{-O(n)}\big( \underline{h}(\mathcal{T}^{\underline{\Delta}}_{\|n-1}) + O\big(n|\log\rho| + \log|\log\varepsilon| + \log n\big)\big) \leq \rho^{-O(n^2)} \log|\log\varepsilon|.$$

The Lebesgue measure of $\Delta \setminus \underline{\Delta}$ is bounded by

$$\varepsilon\, n^{O(1)} 2^{\underline{h}(\mathcal{T}_n^{\underline{\Delta}})} \leq \varepsilon|\log\varepsilon|^{\rho^{-O(n^2)}} < \sqrt{\varepsilon},$$

Setting $\varepsilon$ small enough but in $\exp(-\rho^{-O(n^2)})$ proves the birectional case of Theorem 1.1, with a convergence time of $\rho^{-O(n^2)}$. We can improve the exponent to $O(n)$ by using known bounds on the total 1-energy. With $\mathbf{x}$ fixed, each edge of the ground communication graph has a length at time $t$ that depends only on $\delta$. We call a node $v$ of $\mathcal{T}_n^{\underline{\Delta}}$ *heavy* if its graph contains one or more edges of length at least $n^{-2b}$ (and *light* otherwise). For fixed $\delta$, the number of times the communication graph has at least one edge of length $\lambda$ or more is called the *communication count* $C_\lambda$: it has been shown, using the *total s-energy* [13], that $C_\lambda \leq \lambda^{-1}\rho^{-O(n)}$. It follows that, along any path of the global coding tree, the number of heavy nodes is $\rho^{-O(n)}$. The convergence bound follows then from the fact that all the light nodes between two heavy ones correspond to the same communication graph (hence the same ground matrix). We omit the rest of the proof, which repeats much of the previous argument. □

[7] We could use exponentially small thresholds or even lower, if so desired; crucially, such a rule is required to avoid chaotic behavior.

## V. Nonbidirectional Systems

We sketch the general case of Theorem 1.1, beginning with the case $d = \mathtt{d} = 1$, which removes the distinction between ground and lifted agents. We first consider a simpler system and show later how to reduce any influence system to it. Let $t_o$ be the timing threshold of the perturbation rule (b) and let $H$ be a directed $n$-node graph.[8] Given $\mathbf{x} \in \Omega$, as soon as $\mathcal{G}(f^t(\mathbf{x}))$ contains an edge not in $H$ or some edge of $H$ fails to appear within a time interval of length $t_o$, we stop the system. The coding tree $\mathcal{T}_n$ is still well defined. The difference is that some nodes are now absorbed (and their subtree pruned) because the corresponding orbits are entering a "wrong" atom. We show that whp the orbit of any point is attracted to a limit cycle or its path in the coding tree reaches an absorbed leaf. Intuitively, $H$ is our guess for the *persistent* graph, defined to include exactly the edges that appear infinitely often in $\mathcal{G}(f^t(\mathbf{x}))_{|t\geq 0}$. The new system is no longer Markovian but this is a minor technicality.

Consider the directed graph derived from $H$ by identifying each strongly connected component with a single node. Let $B_1, \ldots, B_r$ be the components whose corresponding nodes are sinks and let $n_i$ denote the number of agents in the group $B_i$; write $n = m + n_1 + \cdots + n_r$. The system is block-directional system with $m$ (resp. $n - m$) $A$-agents (resp. $B$-agents) and, for fixed $\delta$, the coding tree is of the form $\mathcal{T}_{m \to n-m}$, with

$$P_{\leq v} = \begin{pmatrix} A_{\leq v} & C_v \\ 0 & B_{\leq v} \end{pmatrix} . \tag{8}$$

We break down the bifurcation analysis in four stages: (i) we bound the rate at which phase tubes thin out; (ii) we argue that, deep enough in the coding tree, perturbations keep the expected (mean) degree below one; (iii) we show how perturbed phase tubes avoid being split by *SP* discontinuities at high depths; finally, (iv) we show to reduce any influence system to the "persistent" case. We assume throughout this section that $\rho > n^{-O(1)}$: this is not required for the proof, but it simplifies the calculations and allows us to recycle the notation $\rho$ for a different purpose.

### A. The thinning rate

As the depth of a node $v$ of the global coding tree grows, $A_{\leq v}$ and $B_{\leq v}$ tend to matrices of ranks 0 and $r$, respectively, at a "thinning" rate that we can bound.

Lemma 5.1: *Given a node $v$ of $\mathcal{T}_{m \to n-m}$, there exist vectors $\mathbf{z}_i \in \mathbb{R}^{n_i}$ $(i = 1, \ldots, r)$, such that, for any $t_v \geq t_c := n^{cnt_o}$ and a large enough constant $c$,*

(i) $\|A_{\leq v}\mathbf{1}_m\|_\infty \leq e^{-\gamma t_v}$ *and*

(ii) $\left\| B_{\leq v} - \mathrm{diag}\,(\mathbf{1}_{n_1}\mathbf{z}_1^T, \ldots, \mathbf{1}_{n_r}\mathbf{z}_r^T) \right\|_{\max} \leq e^{-\gamma' t_v}$,

*where $\gamma = 1/t_c$ and $\gamma' = n^{-cn}$.*

*Proof:* We begin with (i). Consider the initial state $\mathbf{x} = (\mathbf{1}_m, \mathbf{0}_{n-m})$, with all the $A$-agents at 1 and the $B$-agents at 0, and let $\mathbf{y} = P_{\leq v}\mathbf{x}$; obviously, $\|A_{\leq v}\mathbf{1}_m\|_\infty = \|\mathbf{y}\|_\infty$. To bound the $\ell_\infty$-norm of $\mathbf{y}$, we apply to $\mathbf{x}$ the sequence of maps specified along the path of $\mathcal{T}_{m \to n-m}$ from the root to $v$.[9] Referring to the arborator (3), let's analyze the factor

$$\underline{\mathcal{T}_{w_k \to n-w_k}}_{\,|\, t_{k+1}-t_k-1} \otimes \mathcal{T}_n^{|1} .$$

The wait period $t_{k+1} - t_k$ before wetness propagates again at time $t_{k+1}$ is at most $t_o$: indeed, by definition, any $A$-agent can reach some $B$-agent in $H$ via a directed path, so all of them will eventually get wet. It follows that the set $W_k$ cannot fail to grow in $t_0$ steps unless it already contains all $n$ nodes or the trajectory reaches an absorbing leaf. Assume that the agents of $W_{t_k+1}$, the wet agents at time $t_k + 1$ lie in $(0, 1 - \sigma]$. Because their distance to 1 can decrease by at most a polynomial factor at each step, they all lie in $(0, 1 - \sigma n^{-O(t_o)}]$ between times $t_k$ and $t_{k+1}$. The agents newly wet at time $t_{k+1} + 1$, ie, those in $W_{t_{k+1}+1} \setminus W_{t_{k+1}}$, move to a weighted average of up to $n$ numbers in $(0, 1)$, at least one of which is in $(0, 1 - \sigma n^{-O(t_o)}]$. This implies that the agents of $W_{t_{k+1}+1}$ lie in $(0, 1 - \sigma n^{-O(t_o)}]$. Since $\sigma \leq 1$, when all the $A$-agents are wet, which happens within $nt_o$ steps, their positions are confined within $(0, 1 - n^{-O(nt_o)}]$. It follows that

$$\|\mathbf{y}\|_\infty \leq e^{-\lfloor t_v/(nt_o)\rfloor n^{-O(nt_o)}},$$

which proves (i). We establish (ii) along similar lines. The coding tree $\mathcal{T}_{m \to n-m}$ can be written as an absorbed instance of $\mathcal{T}_{m \to (n_1\|\cdots\|n_r)}$ The subgraph $H_{|B_i}$ of $H$ induced by the agents of any given $B_i$ is strongly connected, so viewed as a separate subsystem, the $B$-agents are newly wetted at least once every $nt_o$ steps. By repeating the following argument for each $B_i$, we can assume, for the purposes of this proof, that $B = B_1$, $n_1 = n - m$ and $r = 1$.

Initially, place $B$-agent $j$ at 1 and all the others at 0; then apply to it the sequence of maps leading to $B_{\leq v}$ (again, this may not be the actual trajectory of that initial state). The previous argument shows that the entries of the $j$-th column of $B_{\leq v}$, which denote the locations of the agents at time $t_v$, are confined to an interval of length $e^{-\lfloor t_v/(nt_o)\rfloor n^{-O(nt_o)}}$. By the perturbation rule (a), as stated in §IV, this implies that the communication subgraph among the $B$-agents must freeze at some time $t_c = n^{cnt_o}$ for a constant $c$ large enough, hence become $H_{|B}$. Let $\{u_i\}$ be the $n^{O(nt_c)}$ nodes of the coding tree at depth $t_c$. Any deeper node $v$ is such that $B_{\leq v} = Q^{t_v - t_{u_i}} B_{\leq u_i}$ for some $i$, where $Q$ is the stochastic matrix associated with $H_{|B}$. Since that graph is strongly connected, the previous argument shows that the entries in column $j$ of $Q^k$ lies in an interval of length $e^{-kn^{-O(n)}}$; we lose the delay $t_o$. Since $Q^{k+1}$ is derived from $Q^k$ by

[8] We can pick $t_o$ as large as we please, say, doubly exponential in $n$, to make it irrelevant in practice.

[9] The path need not track the orbit of $\mathbf{x}$.

taking convex combinations of the rows of $Q^k$, as $k$ grows, these intervals are nested downwards and hence converge to a number $z_j$. It follows that $Q^k$ tends to $\mathbf{1}_{n_1}\mathbf{z}^T$, with $\|Q^k - \mathbf{1}_{n_1}\mathbf{z}^T\|_{\max} \leq e^{-kn^{-O(n)}}$. Doubling the value of $t_c$ yields part (ii) of the lemma. □

The lemma points to $C_v$ as the key to the dynamics and the necessary focus of our attention. We state the thinning rate bound in terms of the global coding tree for the perturbation interval $\mathbb{I} = (-1, 1)$.

LEMMA 5.2: *Any node $v$ of $\mathcal{T}^{\mathbb{I}}_{m \to n-m}$ of depth $t_v \geq t_c$ has an ancestor $u$ of depth $t_c$ such that*

$$\left\| P_{\leq v} - \begin{pmatrix} 0 & C_v \\ 0 & D_u \end{pmatrix} \right\|_{max} \leq e^{-\gamma t_v},$$

*where $D_u$ is a stochastic matrix of the form $D_u = \operatorname{diag}(\mathbf{1}_{n_1}\mathbf{z}_1(u)^T, \ldots, \mathbf{1}_{n_r}\mathbf{z}_r(u)^T)$.*

## *B. Sparse branching*

Bruin and Deane [5] used a simple, elegant argument to show that generic planar (single-agent) contractions do not branch out nearly as often as one could fear. We prove, likewise, that branching tapers off deep enough in the coding tree. Our argument is not nearly as simple, however, because of the bewildering complexity of the interactions among the agents. By elucidating the entropic contribution of the process, this argument constitutes the heart of the proof. Let $\mathrm{Lin}\,[x_1, \ldots, x_n]$ denote any real linear form over $x_1, \ldots, x_n$, with $\mathrm{Aff}\,[x_1, \ldots, x_n]$ designating the affine version; in neither case may the coefficients depend on $\delta$ or on the agent positions.[10] With $y_1, \ldots, y_r$ understood, a *gap* of type $\omega$ denotes an interval of the form $a + \omega\,\mathbb{I}$, where $a = \mathrm{Aff}\,[y_1, \ldots, y_r]$. We define the set

$$\mathbb{C}[y_1, \ldots, y_r] = \{\, (\xi\,, \overbrace{y_1, \ldots, y_1}^{n_1}, \ldots, \overbrace{y_r, \ldots, y_r}^{n_r}\,\},$$

where $\xi \in (0,1)^m$. The variables $y_1, \ldots, y_r$ denote the limit positions of the $B$-agents: they are linear combinations of their initial positions $x_{m+1}, \ldots, x_n$. Let $v$ be a node of the global coding tree $\mathcal{T}^{\mathbb{I}}_{m \to n-m}$. The matrix $P_{\leq v}$ is a product $P_{t_v} \cdots P_0$, with $P_0 = \mathbf{I}_n$ and $P_0, \ldots, P_{t_v}$ form what we call a *valid matrix sequence*. Fix a parameter $\rho > 0$ (not to be confused with the matrix bound $\rho$ used earlier) and a point $\mathbf{x}$ in $\mathbb{R}^n$. The phase tube formed by the cube $\mathbf{B} = \mathbf{x} + \rho\,\mathbb{I}^n$ and the matrix sequence $P_0, \ldots, P_{t_v}$ consists of the cells $P_0\,\mathbf{B}, \ldots, (P_{t_v} \cdots P_0)\mathbf{B}$. Note that it might not track an actual orbit from $\mathbf{B}$. We say that the phase tube *splits* at node $v$ if $(P_k \cdots P_0\,\mathbf{B}) \setminus \mathcal{M}_{\mathbb{I}}$ is disconnected. The following result is the key to sparse branching:

LEMMA 5.3: *Fix $\rho > 0$, $D_0 \geq 2^{(1/\gamma)^{n+1}}$, and $(y_1, \ldots, y_r) \in \mathbb{R}^r$, where $\gamma = n^{-cnt_o}$. There exists a union $W$ of $n^{O(nD_0)}$ gaps of type $\rho n^{O(n^5 D_0)}$ such that, for any interval $\Delta \subseteq \mathbb{I} \setminus W$ of length $\rho$ and any $\mathbf{x} \in \mathbb{C}[y_1, \ldots, y_r]$, the phase tube formed by the box $\mathbf{x} + \rho\,\mathbb{I}^n$ along any path of $\mathcal{T}^{\Delta}_{m \to n-m}$ of length at most $D_0$ cannot split at more than $D_0^{1-\gamma^{n+1}}$ nodes.*

*Proof:* The crux of the lemma is the *uniformity* over $\mathbf{x}$: only $(y_1, \ldots, y_r)$ needs to be fixed. We begin with a technical lemma. For $k = 0, \ldots, D$, let $a_k$ be a row vector in $\mathbb{R}^m$ with $O(\log n)$-bit rational coordinates and $A_k$ be an $m$-by-$m$ nonnegative matrix whose entries are rationals over $O(\log N)$ bits, for $N > n$. Write $v_k = a_k A_k \cdots A_0$, with $A_0 = \mathbf{I}_m$, and assume that the maximum row-sum $\alpha = \max_{k>0} \|A_k \mathbf{1}\|_\infty$ satisfies $0 < \alpha < 1$. Given $I \subseteq \{0, \ldots, D\}$, denote by $V_{|I}$ the matrix whose rows are, from top to bottom, the row vectors $v_k$ with the indices $k \in I$ sorted in increasing order. The following result is an elimination device meant to factor out the role of the $A$-agents.

LEMMA 5.4: *Given any integer $D \geq 2^{(1/\beta)^{m+1}}$ and $I \subseteq \{0, \ldots, D\}$ of size $|I| \geq D^{1-\beta^{m+1}}$, where $\beta = |\log \alpha| / (cm^3 \log N)$ for a constant $c$ large enough, there exists a unit vector $u$ such that*

$$u^T V_{|I} = \mathbf{0} \quad \textit{and} \quad u^T \mathbf{1} \geq N^{-cm^3 D}.$$

The remainder of the proof and the full version of this paper can be found at:

*http://www.cs.princeton.edu/∼chazelle/pubs/focs12full.pdf*


## *Acknowledgments*

I wish to thank Pascal Koiran and John Tsitsiklis for helpful conversations.


## REFERENCES


[1] Alur, R., Henzinger, T.A., Sontag, E.D. Hybrid Systems III. Verification and Control, Springer Verlag, Berlin, LNCS 1066, 1996.

[2] Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., Viale, M., Zdravkovic, V. *Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study*, Proc. National Academy of Sciences 105 (2008), 1232-1237.

[3] Blondel, V.D., Hendrickx, J.M., Olshevsky, A., Tsitsiklis, J.N. *Convergence in multiagent coordination, consensus, and flocking*, Proc. 44th IEEE Conference on Decision and Control, Seville, Spain, 2005.

[4] Blondel, V.D., Tsitsiklis, J.N. *A survey of computational complexity results in systems and control*, Automatica 36 (2000), 1249–1274.


[10] For example, we can express $y = \delta + x_1 - 2x_2$ as $y = \delta + \mathrm{Lin}\,[x_1, x_2]$ and $y = \delta + x_1 - 2x_2 + 1$ as $y = \delta + \mathrm{Aff}\,[x_1, x_2]$.

[5] Bruin, H., Deane, J.H.B. *Piecewise contractions are asymptotically periodic*, Proc. American Mathematical Society 137, 4 (2009), 1389–1395.

[6] Bullo, F., Cortés, J., Martinez, S., Distributed Control of Robotic Networks, Applied Mathematics Series, Princeton University Press, 2009.

[7] Buzzi, J. *Piecewise isometries have zero topological entropy*, Ergod. Theory Dyn. Syst. 21 (2001), 13711377. MR1855837 (2002f:37029)

[8] Camazine, S., Deneubourg, J.L., Franks, N., Sneyd, J., Bonabeau, E., Theraulaz, G. Self-Organization in Biological Systems (2001), Princeton University Press.

[9] Cao, M., Spielman, D.A. Morse, A.S. *A lower bound on convergence of a distributed network consensus algorithm*, 44th IEEE Conference on Decision and Control, and the European Control Conference 2005, Seville, Spain, 2005

[10] Castellano, C., Fortunato, S., Loreto, V. *Statistical physics of social dynamics*, Rev. Mod. Phys. 81 (2009), 591–646.

[11] Catsigeras, E., Budelli, R. *Topological dynamics of generic piecewise continuous contractive maps in $n$ dimensions*, International Journal of Pure and Applied Mathematics, 68 (2011), 61-83.

[12] Cessac, B. *A discrete time neural network model with spiking neurons: rigorous results on the spontaneous dynamics*, J. Math. Biol. 56 (2008), 311-345.

[13] Chazelle, B. *The total $s$-energy of a multiagent system*, SIAM J. Control Optim. 49 (2011), 1680–1706.

[14] Chazelle, B. *Natural algorithms and influence systems*, CACM Research Highlights, 2012.

[15] Collins, G. E. *Quantifier elimination for real closed fields by cylindrical algebraic decomposition*, Proc. 2nd GI Conf. Automata Theory and Formal Languages, New York, Springer-Verlag (1975), 134–183.

[16] Devaney, R.L. An Introduction to Chaotic Dynamical Systems, 2nd edition, Westview Press, 2003.

[17] Dittmer, J.C. *Consensus formation under bounded confidence*, Nonlinear Analysis, Theory, Methods & Applications, 47 (2001), 4615–4622.

[18] Hegselmann, R., Krause, U. *Opinion dynamics and bounded confidence models, analysis, and simulation*, J. Artificial Societies and Social Simulation 5, 3 (2002).

[19] Hegselmann R, Krause U. *Truth and cognitive division of labor: first steps towards a computer aided social epistemology*, J. Artificial Societies and Social Simulation 9 (2006).

[20] Hendrickx, J.M., Blondel, V.D. *Convergence of different linear and non-linear Vicsek models*, Proc. 17th International Symposium on Mathematical Theory of Networks and Systems (MTNS2006), Kyoto (Japan), July 2006, 1229–1240.

[21] Jadbabaie, A., Lin, J., Morse, A.S. *Coordination of groups of mobile autonomous agents using nearest neighbor rules*, IEEE Trans. Automatic Control 48 (2003), 988–1001.

[22] Koiran, P., Cosnard, M., Garzon, M. *Computability with low-dimensional dynamical systems*, Theoretical Computer Science A 132 (1994), 113-128.

[23] Kruglikov, B., Rypdal, M. *A piecewise affine contracting map with positive entropy*, Discrete and Continuous Dynamical Systems 16, 2 (2006), 393–394.

[24] Kurz, S., Rambau, J. *On the Hegselmann-Krause conjecture in opinion dynamics*, Journal of Difference Equations and Applications (2009).

[25] Lorenz, J. *A stabilization theorem for dynamics of continuous opinions*, Physica A: Statistical Mechanics and its Applications 355 (2005), 217–223.

[26] Mirtabatabaei A., Bullo, F. *Opinion dynamics in heterogeneous networks: convergence conjectures and theorems*, SIAM Journal on Control and Optimization, March 2011.

[27] Moreau, L. *Stability of multiagent systems with time-dependent communication links*, IEEE Transactions on Automatic Control 50 (2005), 169–182.

[28] Nedic, A., Olshevsky, A., Ozdaglar, A., Tsitsiklis, J.N. *On distributed averaging algorithms and quantization effects*, IEEE Transactions on Automatic Control, 2009.

[29] Niven, I., Zuckerman, H.S., Montgomery, H.L. An Introduction to the Theory of Numbers, Wiley, 5th edition, 1991.

[30] Okubo, A., Levin, S.A. Diffusion and Ecological Problems, Springer, 2nd ed, 2002.

[31] Olshevsky, A., Tsitsiklis, J.N. *Convergence speed in distributed consensus and averaging*, SIAM Journal on Control and Optimization 48 (2009), 33–55.

[32] Parrish, J.K., Hamner, W.M. Animal Groups in Three Dimensions, Cambridge University Press, 1997.

[33] Papachristodoulou, A., Jadbabaie, A. *Synchonization in oscillator networks: switching topologies and non-homogeneous delays*, Proc. IEEE CDC-ECC (2005), 5692–5697.

[34] Siegelmann, H.T., Sontag, E.D. *On the computational power of neural nets*, J. Comput. System Sci., 50 (1995), 132–150.

[35] Sontag, E.D. *Nonlinear regulation: the piecewise linear approach*, IEEE Trans. Automat. Control 26 (1981), 346–358.

[36] Tsitsiklis, J.N., Bertsekas, D.P., Athans, M. *Distributed asynchronous deterministic and stochastic gradient optimization algorithms*, IEEE Transactions on Automatic Control 31 (1986), 803–812.