Diffusive Influence Systems *

Bernard Chazelle †

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.

(*Preliminary version in Proc. 53rd IEEE FOCS 2012*)

1 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].

*This work was supported in part by NSF grants CCF-0832797, CCF-0963825, and CCF-1016250.

†Department of Computer Science, Princeton University, chazelle@cs.princeton.edu

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 ε is the distance to the fixed point.*

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 δ 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

[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. We use the shorthand *whp* for "with probability arbitrarily close to 1."

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.

2 Preliminaries

We show in §2.1 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 §2.2, we build a conjugation with the baker's map to exhibit chaos and, in §2.3, we show how to simulate a Turing machine. All three constructions use linear decision trees. This is not surprising in view of §2.4, where we show how to linearize the decision procedure of any influence system.

[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].

2.1 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}.

2.2 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:

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

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:

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

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.

2.3 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:

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

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 $\mathbf{a}^T\mathbf{x} = a_0 x_0$. 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.

2.4 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

[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.

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 $a_1 y_1 + \cdots + a_N y_N + \delta$, the perturbation rule acts in a more subtle 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 $K \subseteq \{1, \ldots, n\}$, 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

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

with the homogeneizing agent 1 permanently positioned at $(x_{1,1}, \ldots, x_{1,d}) = \mathbf{1}_d$. Next, we define $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$, 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 $(k_1, j_1, \ldots, k_\mathtt{d}, j_\mathtt{d})$ for $k_i \in \{1, \ldots, n\}$ and $j_i \in \{1, \ldots, d\}$. 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.

3 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 δ,

[6] Not to be confused with d.

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, ρ, 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.

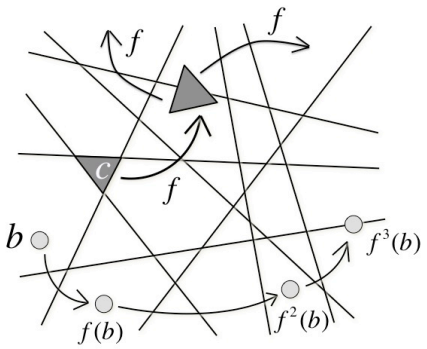

Figure 1: The atom c of the *SP* maps via $f = P_c$ to a cell intersecting two atoms.

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, unlike in Fig.1, 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, 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.

3.1 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 ∞) is called the *nesting time* ν 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 ν 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 δ, 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.

3.2 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.

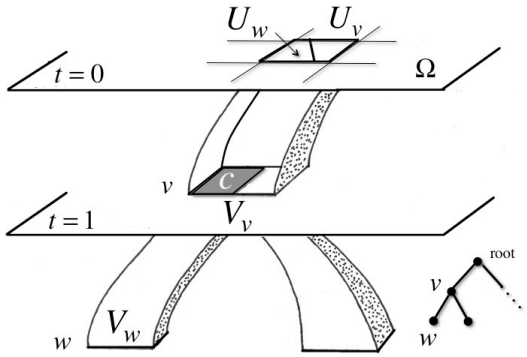

Figure 2: A phase tube (U_w, V_w) of length two: $V_w = f(c) = f^{t_w}(U_w)$.

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 (Fig.2). 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, $U_v, f(U_v), \ldots, f^{t_v}(U_v) = V_v$.

Building $\mathcal{T}$

[1] The root v has depth $t_v = 0$; set $U_v \leftarrow V_v \leftarrow \Omega$.

[2] Repeat forever:

[2.1] For each newly created node v:

- For each cell c of $V_v \setminus \mathcal{M}_\delta$, create a child w of v and set $P_w \leftarrow f_{|c}$; $V_w \leftarrow P_w\, c$; $U_w \leftarrow U_v \cap f^{-t_v}(c)$.

Let $ww'w''\cdots$ denote the upward, t_w-node path from w to the root (but excluding the root). Using the notation $P_{\le w} = P_w P_{w'} P_{w''} \cdots$, we have the identities $V_w = P_{\le w} U_w$ 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 $\bigcap_{t\ge 0} U_{v_t}$ 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;[7] $\#\mathcal{S}_\nu \le 2^{h(\mathcal{T})}$.

We need additional parameters, such as the attraction rate and the augmented word-entropy, but we postpone their introduction. Later, we will randomize δ within a small interval Δ, 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 $\mathbb{R}^{n+1}$.

3.3 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 $P_w = P_u \oplus P_v$.

[7] All logarithms are to the base 2.

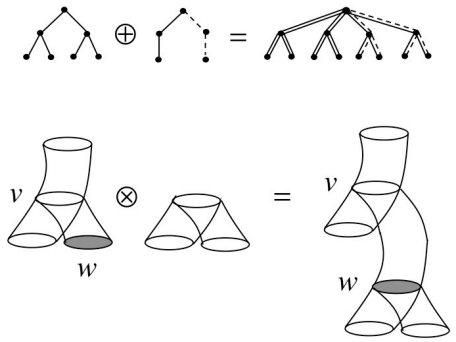

Figure 3: The two tensor operations.

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 Ω. Given an arbitrary region $\Lambda \subset \Omega$, define the hybrid system S with the dynamics of S_2 over Λ 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: in Fig.3, for example, $U_{\text{root}}(\mathcal{T}_2)$ must be clipped to match $V_w(\mathcal{T}_1)$, which in turn might involve pruning $\mathcal{T}_2$.

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 $m \parallel n - m$.

3.4 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 $0 < m \leq n$ 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_ℓ 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.

EXAMPLE 3.4: The third column below lists a graph sequence $G_0, \ldots, G_{11}$ in chronological order, with the superscript w indicating the edges through which water propagates to dry nodes. The system is block-directional with three A-agents labeled a, b, c and one B-agent labeled d. For clarity, we spell out the agents as subscripts. Note that we use $\mathcal{T}_{ab \to cd}$ and not $\mathcal{T}_{abd \to c}$. The latter would be correct but inductively unsound: a system of type $3 \to 1$ should not be calling a system of the same type recursively; that is, unless other structure is put in place—as will be done later. After all, with a, b, d already wet, the phase awaits the wetting of c. This strategy would be inductively unsound, however, because it would resolve a system $\mathcal{T}_{abc \to d}$ by means of another one, $\mathcal{T}_{abd \to c}$, of the same type $3 \to 1$. Renormalization, which is denoted by underlining, compresses into single time units all the time intervals during which wetness does not spread to dry agents. With the subscripts (resp. superscript) indicating the time compression rates (resp. tree height), the 11-node path of $\mathcal{T}_{abc \to d}$ matching the graph sequence above can be expressed as

$$\underline{\mathcal{T}_{d \,\|\, abc}}_{|3} \otimes \mathcal{T}^{|1}_{abcd} \otimes \underline{\mathcal{T}_{a \to bcd}}_{|2} \otimes \mathcal{T}^{|1}_{abcd} \otimes \underline{\mathcal{T}_{ab \to cd}}_{|3} \otimes \mathcal{T}^{|1}_{abcd}\,.$$

Flow tracking			
	$W_0 = \{d\}$ $W_1 = \{d\}$ $W_2 = \{d\}$	$d \quad a \to b \to c$ $d \quad a \leftarrow b \to c$ $d \quad a \to b \leftarrow c$	$\mathcal{T}_{d \,\|\, abc}$
$t_1 = 3$	$W_3 = \{d\}$	$d \overset{w}{\leftarrow} a \leftarrow b \leftarrow c$	$\mathcal{T}_{abcd}$
	$W_4 = \{a, d\}$ $W_5 = \{a, d\}$	$d \leftarrow a \to b \to c$ $d \quad a \to b \to c$	$\mathcal{T}_{a \to bcd}$
$t_2 = 6$	$W_6 = \{a, d\}$	$d \leftarrow a \overset{w}{\leftarrow} b \leftarrow c$	$\mathcal{T}_{abcd}$
	$W_7 = \{a, b, d\}$ $W_8 = \{a, b, d\}$ $W_9 = \{a, b, d\}$	$d \leftarrow a \to b \to c$ $d \leftarrow a \leftarrow b \quad c$ $d \leftarrow a \to b \to c$	$\mathcal{T}_{ab \to cd}$
$t_3 = 10$	$W_{10} = \{a, b, d\}$	$d \leftarrow a \to b \overset{w}{\leftarrow} c$	$\mathcal{T}_{abcd}$
	$W_{11} = \{a, b, c, d\}$	$d \quad a \leftarrow b \quad c$	$\mathcal{T}_{d \,\|\, abc}$

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.

4 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 $(dn)^{\mathfrak{d}}$ 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.[8] 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 ρ is the smallest nonzero entry among the ground matrices. We conclude that water propagation to all the agents

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

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

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

After ρ^{-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 Δ, 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:

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

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,

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

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 ε, where

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

Removing from Δ 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 $\Delta \setminus \underline{\Delta}$ 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:

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

hence, for ε small enough,

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

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

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

Setting ε 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 δ. 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 δ, the number of times the communication graph has at least one edge of length λ or more is called the *communication count* C_λ: it has been shown, using the *total s-energy* [13], that $C_\lambda \le \lambda^{-1}\rho^{-O(n)}$. 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. □

5 Nonbidirectional Systems

We prove 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.[9] 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 $\mathcal{G}(f^t(\mathbf{x}))_{|t\ge 0}$. 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

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

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 δ, 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: in §5.1 we bound the rate at which phase tubes thin out; in §5.2 we argue that, deep enough in the coding tree, perturbations keep the expected (mean) degree below one; in §5.3, we show how perturbed phase tubes avoid being split by *SP* discontinuities at high depths; finally, in §5.4, 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 ρ for a different purpose.

5.1 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 $\mathbf{z}_i \in \mathbb{R}^{n_i}$ $(i = 1,\dots,r)$, such that, for any $t_v \geq t_c := n^{cnt_o}$ and a large enough constant c,*

$$\text{(i)}\ \ \|A_{\leq v}\mathbf{1}_m\|_\infty \leq e^{-\gamma t_v} \quad \textit{and} \quad \text{(ii)}\ \ \left\| B_{\leq v} - \operatorname{diag}\left(\mathbf{1}_{n_1}\mathbf{z}_1^T,\dots,\mathbf{1}_{n_r}\mathbf{z}_r^T\right)\right\|_{\max} \leq e^{-\gamma' t_v},$$

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 ℓ_∞-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.[10] 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

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

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 §4, 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 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 $D_u = \operatorname{diag}(\,\mathbf{1}_{n_1}\mathbf{z}_1(u)^T, \ldots, \mathbf{1}_{n_r}\mathbf{z}_r(u)^T\,)$.

5.2 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 δ or on the agent positions.[11] With $y_1, \ldots, y_r$ understood, a *gap* of type ω denotes an interval of the form $a + \omega\,\mathbb{I}$, where $a = \mathrm{Aff}\,[y_1, \ldots, y_r]$. We define the set

$$\mathbb{C}[y_1, \ldots, y_r] = \Big\{ (\,\xi\,,\, \overbrace{y_1, \ldots, y_1}^{n_1}, \ldots, \overbrace{y_r, \ldots, y_r}^{n_r}\,) \mid \xi \in (0,1)^m \Big\}.$$

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 ρ 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 ρ 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*

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

This implies that $\mathbf{1}_{|I|}$ is not in the column space of $V_{|I}$. Although unrelated, we can pick the same constant c as the one used in Lemma 5.1. Since $\alpha \geq N^{-O(1)}$, β can be assumed much less than 1. To prove Lemma 5.3, we first consider the case where the

[11] 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 $y = \delta + \mathrm{Aff}\,[x_1, x_2]$.

splitting nodes are well separated along their path, which allows for Lemma 5.1 to be used; then we extend this result to all cases. Given a valid matrix sequence $P_0, \ldots, P_{D_0}$, pick a sequence of $D+1$ integers $0 = s_0 < \cdots < s_D \le D_0$ such that

$$D \ge 2^{(1/\beta)^{m+1}} \qquad \text{and} \qquad 1/\gamma \le s_k - s_{k-1} \le 3/\gamma, \tag{9}$$

for $k = 1, \ldots, D$: we identify the matrix A_k of Lemma 5.4 with the m-by-m upper left principal submatrix of $P_{s_k} P_{s_k-1} \cdots P_{s_{k-1}+1}$; using the notation of (8), $A_k = A_{\le w}$, for some node w (not necessarily an ancestor of v) of depth $t_w = s_k - s_{k-1} \ge 1/\gamma$. Thus, by Lemma 5.1, for $k > 0$, the maximum row-sum of any A_k satisfies $\alpha \le 1/e$: each A_k is a submatrix of a product of at most $3/\gamma$ transition matrices, so each entry is an $O(\log N)$-bit rational, with $N = n^{n^2/\gamma}$. What is the row vector a_k? For $k = 0, \ldots, D$, pick any one of the $n^{O(1)}$ margin hyperplanes and denote by a_k the m-dimensional vector consisting of the $O(\log n)$-bit rational coefficients indexed by the A-agents.[12] Fix $\delta \in \mathbb{I}$ and pick I in Lemma 5.4 to be of size $\lceil D^{1-\beta^{m+1}} \rceil$. Assume that, given $\mathbf{x} \in \mathbb{C}[y_1, \ldots, y_r]$, the phase tube formed by the box $\mathbf{x} + \rho\, \mathbb{I}^n$ and $P_{s_0}, \ldots, P_{s_D}$ splits at every index of I along the chosen hyperplane. In other words, for each $k \in I$, there exist a node z_k of depth $t_{z_k} = s_k$ and $\rho_i = \rho_i(k)$, for $i = 1, \ldots, n$, such that $|\rho_i| < \rho$ and

$$(a_k, b_k) \begin{pmatrix} A_{\le z_k} & C_{z_k} \\ 0 & B_{\le z_k} \end{pmatrix} (x_1 + \rho_1, \ldots, x_n + \rho_n)^T = 1 + \delta,$$

where the selected hyperplane is of the form $a_k(x_1, \ldots, x_m)^T + b_k(x_{m+1}, \ldots, x_n)^T = 1 + \delta$, with $b_k \in \mathbb{Q}^{n-m}$. Since $a_k A_{\le z_k} = v_k$ and $\mathbf{x} \in \mathbb{C}[y_1, \ldots, y_r]$, it follows that

$$v_k(x_1 + \rho_1, \ldots, x_m + \rho_m)^T + \operatorname{Lin}[\, y_1, \ldots, y_r, \rho_{m+1}, \ldots, \rho_n \,] = 1 + \delta\,, \tag{10}$$

where the coefficients in the linear form are of magnitude $n^{O(1)}$.

Lemma 5.4 allows us to eliminate the variables $x_1, \ldots, x_m$: we premultiply $V_{|I}$ by the unit vector u to find that $\operatorname{Lin}[\, y_1, \ldots, y_r \,] + \chi = (1 + \delta)\xi$, where $|\chi| \le \rho D n^{O(1)}$, $\xi \ge N^{-cm^3 D}$, and the coefficients of the linear form are bounded by $D n^{O(1)}$; hence,

$$|\, \delta + \operatorname{Aff}[\, y_1, \ldots, y_r \,]\,| \le \rho N^{O(cm^3 D)}, \tag{11}$$

where the coefficients of the affine form are bounded by $N^{O(cm^3 D)}$. (We leave the constant c in the exponent to highlight its influence.) The whole point of the exercise is that the variable δ never vanishes during the elimination. Thus, as long as it remains outside a gap of type $\rho N^{O(cm^3 D)}$, the phase tube formed by $\mathbf{x} + \rho\, \mathbb{I}^n$ and $P_0, \ldots, P_D$ cannot split at every index of I. Counting the number of possible choices of hyperplanes per node raises the number of gaps to $n^{O(|I|)}$. The argument assumes that δ has the same value in each of $|I|$ inequalities. It need not be so: each δ in (10) can be replaced

[12] With $m = 3$, $x_1 - x_3 = 1 + \delta$ gives $a_k = (1, 0, -1)$ and $x_2 - x_4 = 1 + \delta$ produces $a_k = (0, 1, 0)$.

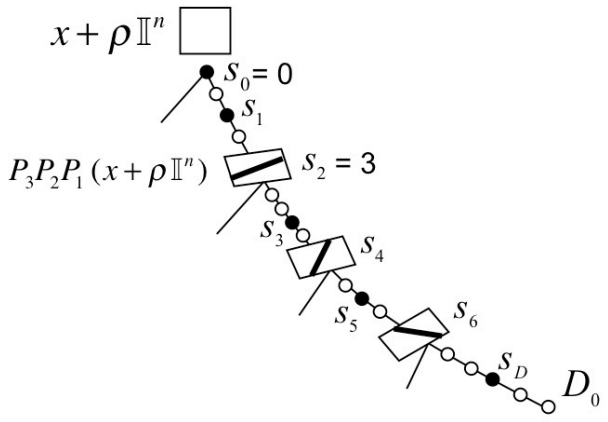

Figure 4: The phase tube splits at the nodes indexed by $I = \{2, 4, 6\}$. The nodes of depth s_k for $k \notin I$ are represented as black dots: s_0, s_1, s_3, s_5, s_7 ($D = 7$). The other nodes in the paths are the white dots.

by $\delta + \nu_k$ ($k \in I$), for $|\nu_k| \le \rho$, and the new system of inequalities will still imply (11); this will be crucial for the randomization. We summarize our results, using the bound $|\log \alpha| \ge \log e > 1$.

LEMMA 5.5. *Let $N = n^{n^2/\gamma}$ and $\beta = 1/(cm^3 \log N)$, where c is the constant used in Lemma 5.4. Fix a path in $\mathcal{T}^{\mathbb{I}}_{m\to n-m}$ from the root and pick $D+1$ nodes on it of depth $0 = s_0 < \cdots < s_D$ satisfying (9); out of these nodes, choose a subset I of size $\lceil D^{1-\beta^{m+1}} \rceil$. There exists an exclusion zone W consisting of the union of at most $n^{O(|I|)}$ gaps of type $\rho N^{O(cm^3 D)}$, such that, for any interval $\Delta \subseteq \mathbb{I} \setminus W$ of length ρ and any $\mathbf{x} \in \mathbb{C}[y_1, \ldots, y_r]$, the phase tube formed by $\mathbf{x} + \rho\,\mathbb{I}^n$ cannot split at all the nodes of I in $\mathcal{T}^{\Delta}_{m\to n-m}$ (assuming they exist).*

To prove Lemma 5.3, we need to extend the previous lemma to all the paths of the coding tree of the prescribed length and remove from (9) the lower bound of $1/\gamma$ on the distance between consecutive splitting nodes. Fix $D_0 \ge 2^{(1/\gamma)^{n+1}}$, and let v be a node of $\mathcal{T}^{\mathbb{I}}_{m\to n-m}$ of depth $t_v = D_0$. Since the path is fixed, we can uniquely identify the node v and its ancestors by their depths and denote by P_t the transition matrix of the node at depth t. Define the node set $J = \{1/\gamma, 2/\gamma, \ldots, D_0\}$, with $|J| = \lceil \gamma D_0 \rceil$; recall that $1/\gamma = t_c$ is an integer. Let K be the set of ancestors of v at which the phase tube formed by $\mathbf{x} + \rho\,\mathbb{I}^n$ and $P_0, \ldots, P_{D_0}$ splits (with respect to $\mathcal{T}^{\mathbb{I}}_{m\to n-m}$); assume that

$$|K| \ge D_0^{1-\gamma^{n+1}}. \tag{12}$$

We define I to be the largest subset of K with no two elements of $I \cup \{0\}$ at a distance less than $1/\gamma$; obviously, $|I| \ge \lfloor \gamma |K| \rfloor - 1$. To define $s_1, \ldots, s_D$, we add all of J to I (to keep distances between consecutive nodes small enough) and then clean up the set to avoid distances lower than allowed: we define J' to be the smallest subset of J such that

$L = I \cup (J \setminus J')$ contains no two elements at a distance less than $1/\gamma$. Each element of I can cause the disappearance of at most two elements in J for the addition of one into L, hence $|J|/2 \le |L| \le \gamma D_0 + 1$. By construction, consecutive elements of L are at most $3/\gamma$ away from each other, so we can identify L with the sequence $s_1 < \cdots < s_D$. By $m < n$ and the specifications of γ in Lemma 5.1 and N, β in Lemma 5.5, we can verify that

$$\text{(i)}\ \ D_0 \ge 2^{(1/\gamma)^{n+1}} \ge \gamma^{-1} 2^{(1/\beta)^{m+1}+1} \quad \text{and} \quad \text{(ii)}\ \ D_0^{1-\gamma^{n+1}} \ge \tfrac{2}{\gamma}(\gamma D_0 + 1)^{1-\beta^{m+1}}. \quad (13)$$

Part (i) ensures (9). By Lemma 5.5, keeping δ outside the union W of at most $n^{O(|I|)}$ gaps of type $\rho N^{O(m^3 D)}$ prevents I from witnessing a phase tube split at each of its nodes, and hence keeps $K \supseteq I$ from being, as claimed, made entirely of "splitting" nodes. For this, we need to ensure that $|I| \ge D^{1-\beta^{m+1}}$, which follows from: (12); $|I| \ge \lfloor \gamma |K| \rfloor - 1$; $D = |L| \le \gamma D_0 + 1$; and part (ii) of (13).

We conclude that, as long as we choose an interval $\Delta \subseteq \mathbb{I} \setminus W$ of length ρ, the coding tree $\mathcal{T}^{\Delta}_{m \to n-m}$ cannot witness splits at all of the nodes of K (if they exist: their existence is ensured only in $\mathcal{T}^{\mathbb{I}}_{m \to n-m}$) for the phase tube formed by any box $\mathbf{x} + \rho\, \mathbb{I}^n$, where $y_1, \ldots, y_r$ are fixed and $\mathbf{x} \in \mathbb{C}[y_1, \ldots, y_r]$. Note the order of the quantifiers: first, we fix the coordinates y_k and the target length D_0, and we pick a large enough candidate splitting node set K in $\mathcal{T}^{\mathbb{I}}_{m \to n-m}$; these choices determine the exclusion zone W; next, we pick a suitable Δ and then claim an impossibility result for *any* $\mathbf{x}$ in $\mathbb{C}[y_1, \ldots, y_r]$. To complete the proof of Lemma 5.3, we bound, by 2^{D_0} and $n^{O(nD_0)}$ respectively, the number of ways of choosing K (hence I, L) and the number of nodes v in $\mathcal{T}^{\mathbb{I}}_{m \to n-m}$ of depth $t_v = D_0$. □

Proof of Lemma 5.4. We can make the assumption that I includes 0, since all cases easily reduce to it. Indeed, let l be the smallest index in I. If $l > 0$, subtract l from the indices of I to define $I' \supseteq \{0\}$. Form the matrix $V'_{|I'}$ of vectors v'_k, where $v_{k+l} = v'_k A_l \cdots A_0$. Rewriting $V_{|I}$ as $V'_{|I'} A_l \cdots A_0$ takes us to the desired case (padding V' to bring the size up back D): we observe that, if $u^T V'_{|I'} = \mathbf{0}$, then so does $u^T V_{|I}$. We may also assume that all v_k are nonzero since the lemma is trivial otherwise. All the coordinates of v_k can be expressed as $O(m^2(k+1)\log N)$-bit rationals sharing a common denominator; therefore,

$$N^{-O((k+1)m^2)} \le \|v_k\|_1 \le 2^{-k|\log \alpha| + O(\log n)}. \quad (14)$$

The *affine hull* of $V_{|I}$ is the flat defined by $\{\, z^T V_{|I} \,:\, z^T \mathbf{1} = 1 \,\}$: its dimension is called the *affine rank* of $V_{|I}$. Let $g(D, r)$ be the maximum value of $|I|$, for $\{0\} \subseteq I \subseteq \{0, \ldots, D\}$, such that $V_{|I}$ has affine rank at most r and its affine hull does not contain the origin. Lemma 5.4 follows from this inequality, whose proof we postpone: for $r = 0, \ldots, m-1$,

$$g(D, r) < D^{1-\beta^{m+1}}, \quad \text{for any } \ D \ge 2^{(1/\beta)^{m+1}}, \quad (15)$$

where $\beta = |\log \alpha|/(cm^3 \log N)$, for constant c large enough. Indeed, given any $\{0\} \subseteq I \subseteq \{0, \ldots, D\}$ of size at least $D^{1-\beta^{m+1}}$, we have $|I| > g(D, m-1)$, so the affine hull of

$V_{|I}$ contains the origin. If r is its affine rank, then there exists $J \subseteq I$ of size $r+1$ such that the affine rank of $V_{|J}$ is r and its affine hull contains the origin, hence coincides with the row space of $V_{|J}$,[13] which is therefore of dimension r. This implies the existence of r independent columns in $V_{|J}$ spanning its column space: add a column of $r+1$ ones to the right of them to form the $(r+1)$-by-$(r+1)$ matrix M. Since the affine hull of $V_{|J}$ contains the origin, there exists z such that $z^T V_{|J} = \mathbf{0}$ and $z^T \mathbf{1} = 1$, which in turn shows that $\mathbf{1}_{r+1}$ lies outside the column space of $V_{|J}$; therefore M is nonsingular. Since each one of its rows consists of $O(m^2 D \log N)$-bit rationals with a common denominator,

$$|\det M| \geq N^{-O(m^3 D)}. \tag{16}$$

Let ξ be the $(r+1)$-dimensional vector whose k-th coordinate is the cofactor of the k-th entry in the last column of ones in M. Determinant cofactor expansions yield

$$\xi^T M = (\overbrace{0, \ldots, 0}^{r}, \det M).$$

Since the first r columns of M span the column space of $V_{|J}$, it follows that

$$\xi^T (\, V_{|J}, \mathbf{1}_{r+1}\,) = (\overbrace{0, \ldots, 0}^{m}, \det M).$$

By Hadamard's inequality and (14), each coordinate of ξ is at most $n^{O(m)}$ in absolute value; so, by (16), straightforward rescaling and padding with zeroes turns ξ into a suitable vector u such that $u^T V_{|I} = \mathbf{0}$ and $u^T \mathbf{1} \geq N^{-c_1 m^3 D}$, for an absolute constant c_1 that does not depend on c. Replacing c by $\max\{c, c_1\}$ establishes Lemma 5.4.

It suffices now to prove (15), which we do by induction on r. If $V_{|I}$ has affine rank $r = 0$ and its affine hull does not contain the origin, then all the rows of $V_{|I}$ are equal and nonzero. Since $V_{|I}$ has the row v_0, it follows from (14) that $|I| \leq 1 + \max\{k \in I\} = O(|\log \alpha|^{-1} m^2 \log N)$, hence

$$g(D, 0) \leq \beta^{-1}. \tag{17}$$

Assume now that $r > 0$ and that $V_{|I}$ has affine rank exactly r and its affine hull does not contain the origin. Put $I = \{k_0, k_1, \ldots, k_i\}$, with $k_0 = 0$, and consider the smallest j such that $V_{|J}$ has affine rank r, where $J = \{k_0, k_1, \ldots, k_j\} \subseteq I$. Since the origin is not in the affine hull of $V_{|I}$ hence of $V_{|J}$, we can always pick a subset $K \subseteq J$ consisting of $r+1$ independent rows: let $M = V_{|K \cup \{k_i\}}$ denote the $(r+2)$-by-m matrix formed by adding the row v_{k_i} at the bottom of $V_{|K}$.[14] Since $V_{|I}$ has affine rank r, its rank is $r+1$ (using once again the noninclusion of O in the affine hull of $V_{|I}$), hence so is the rank of

[13] Because any $y^T V_{|J}$ can be written as $(y + (1 - y^T \mathbf{1}) z)^T V_{|J}$, where $z^T V_{|J} = \mathbf{0}$ and $z^T \mathbf{1} = 1$.

[14] It may be the case that $i = j$ or $k_i \in K$. Since $r > 0$, we have $k_i \geq k_j \geq 1$ and $j > 0$.

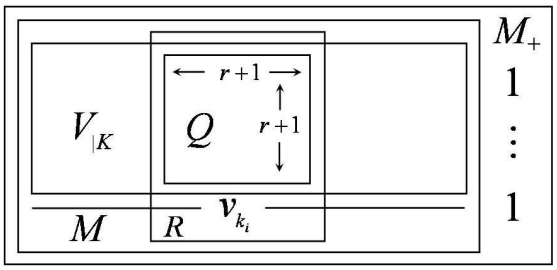

Figure 5: Why a large value of k_i implies that the affine hull of $V_{|I}$, hence of M, contains the origin.

M; in other words, adding v_{k_i} does not increase the rank. We show that if k_i is large enough, the system below is feasible in $\xi \in \mathbb{R}^{r+2}$:

$$\xi^T M_+ = (\overbrace{0, \ldots, 0}^{m}, 1), \tag{18}$$

where M_+ is the $(r+2)$-by-$(m+1)$ matrix $(M, \mathbf{1}_{r+2})$, which leads to a contradiction. This is the crux of the argument and makes essential use of the rapid decay of the vectors v_k. Assume that $k_i > ck_j|\log\alpha|^{-1}m^3 \log N$, for a large enough constant c. We first show that M_+ is of rank $r+2$. Pick $r+1$ independent columns of $V_{|K}$, which is possible since the latter has rank $r+1$, to form the full-ranked $(r+1)$-by-$(r+1)$ matrix Q. Add a new row to it by fitting the relevant part of v_{k_i} (the last row of M) and call R the resulting $(r+2)$-by-$(r+1)$ matrix (Fig.5); consistent with our notation, R_+ will denote the matrix $(R, \mathbf{1})$. A cofactor expansion of the determinant of R_+ along the bottom row shows that

$$|\det R_+| \geq |\det Q| - \Delta \|v_{k_i}\|_1,$$

where Δ is an upper bound on the absolute values of the cofactors other than $\det Q$. In view of (14), the matrix entries involved in these cofactors are all in $n^{O(1)}$; by Hadamard's inequality, this shows that we can set $\Delta = n^{O(m)}$. Likewise, we find that

$$\|v_{k_i}\|_1 \leq 2^{-k_i|\log\alpha| + O(\log n)}.$$

Since Q is nonsingular, we can adapt (16) to derive $|\det Q| \geq N^{-O(m^3 k_j)}$, hence $|\det R_+| > 0$. It follows that the linear system (18) is feasible if we replace M_+ by R_+. As it happens, there is no need to do so since every column of M missing from R lies in the column space of the latter: thus the missing homogeneous equalities are automatically satisfied by the solution ξ. The feasibility of (18) contradicts our assumption that the origin is outside the affine hull of $V_{|I}$; therefore

$$k_j \geq \beta k_i > 0, \tag{19}$$

where $\beta = |\log\alpha|/(cm^3 \log N)$. By definition of j, the affine rank of $V_{|\{k_0,\ldots,k_{j-1}\}}$ is $r-1$ and its affine hull does not contain the origin; therefore $j \leq g(k_{j-1}, r-1)$, with

$g(0, r-1) = 1$. Let $w_0 = a_{k_j}$ and, for $k > 0$, $w_k = a_{k_j+k} A_{k_j+k} \cdots A_{k_j+1}$, thus ensuring that $v_{k_j+k} = w_k A_{k_j} A_{k_j-1} \cdots A_1 A_0$. Since the affine hull of $V_{|I}$ does not contain the origin, neither does that of the matrix W with rows $w_0, w_{k_{j+1}-k_j}, \ldots, w_{k_i-k_j}$. It follows that the affine rank of W is less than m, so $i - j + 1 \leq g(k_i - k_j, m-1)$, hence[15] $i \leq g(k_{j-1}, r-1) + g(k_i - k_j, m-1) - 1$. By (19) and $i = |I| - 1$, we derive, by monotonicity,

$$|I| \leq g(k, r-1) + g(D-k, m-1),$$

where $\beta D \leq k \leq D$; hence, by (17), for $m > 0$ and $D \geq 0$:

$$g(D,r) \leq \begin{cases} 1 & \text{if } D = 0 \\ \beta^{-1} & \text{if } r = 0 \\ g(n_1, m-1) + \cdots + g(n_r, m-1) + \beta^{-1} & \text{if } 0 < r < m, \end{cases}$$

where $n_1 + \cdots + n_r \leq (1-\beta^s)D$, with $s = |\{\, i \mid n_i > 0 \,\}|$. Setting $\eta = \beta^m$, we check that, for all $D, m > 0$,

$$g(D, m-1) \leq \beta^{-2}(2D^{1-\eta} - 1). \tag{20}$$

The case $m = 1$ follows from $g(D, 0) \leq \beta^{-1}$. For $m > 1$, we begin with the case $s = 0$, where

$$g(D, m-1) \leq m - 1 + \beta^{-1} \leq \beta^{-2}(2D^{1-\eta} - 1),$$

which follows from $\alpha \geq N^{-O(1)}$. For $s = 1$, by induction,

$$\begin{aligned} g(D, m-1) &\leq \beta^{-2}(2(1-\beta)^{1-\eta} D^{1-\eta} - 1) + m - 2 + \beta^{-1} \\ &\leq 2\beta^{-2} D^{1-\eta} - (2\beta^{-1}(1-\eta) - O(1)) D^{1-\eta} - \beta^{-2} + \beta^{-1} + m - 2 \\ &\leq \beta^{-2}(2D^{1-\eta} - 1). \end{aligned}$$

Assume that $s > 1$. Being concave and nonnegative, the function $x \mapsto x^{1-\eta}$ is subadditive for $x \geq 0$; therefore,

$$n_1^{1-\eta} + \cdots + n_r^{1-\eta} \leq (1-\beta^s)^{1-\eta} D^{1-\eta}.$$

Setting $r = m - 1$, relation (20) follows from the inequality,

$$\begin{aligned} g(D, m-1) &\leq \beta^{-2}(2(1-\beta^s)^{1-\eta} D^{1-\eta} - s) + m - s - 1 + \beta^{-1} \\ &\leq 2\beta^{-2}(1-\beta^{m-1})^{1-\eta} D^{1-\eta} - \tfrac{3}{2}\beta^{-2} \leq 2\beta^{-2} D^{1-\eta} - \beta^{-2}, \end{aligned}$$

which proves (20), hence (15) and Lemma 5.4. □

[15] It would be nice to bound the affine rank as a function of r, but since we never perturb the transition matrices it is unclear how to do that.

5.3 The degree structure

We decompose the global coding tree into three layers: the top one has no degree constraints; the second has mean degree less than two; and the third has no branching. Consider an initial placement of the B-agents such that the diameter of each B_i is less than n^{-b}. By the perturbation rule (a), the communication subgraph induced by the B-agents is frozen and its transition matrix Q is fixed and independent of the future placement of the B-agents.[16] By the proof of Lemma 5.1, we derive the existence of a rank-r stochastic matrix

$$\widetilde{Q} = \operatorname{diag}\left(\mathbf{1}_{n_1}\mathbf{z}_1^T, \ldots, \mathbf{1}_{n_r}\mathbf{z}_r^T\right)$$

such that $\mathbf{z}_i \in \mathbb{R}^{n_i}$ and

$$\|Q^k - \widetilde{Q}\|_{\max} \le e^{-kn^{-O(n)}}. \tag{21}$$

The B-agents find themselves attracted to the fixed point $\mathbf{y} = \widetilde{Q}\xi$, where $\xi \in \mathbb{R}^{n-m}$ is their initial state vector and

$$\mathbf{y} = (\overbrace{y_1, \ldots, y_1}^{n_1}, \ldots, \overbrace{y_r, \ldots, y_r}^{n_r}).$$

Define $\Upsilon = (0,1)^m \times ((0,1)^{n-m} \cap \Upsilon_B)$, where

$$\Upsilon_B = \mathbf{y} + (n^{-2b}\,\mathbb{I}^{n-m}) \cap \ker \widetilde{Q}.$$

If $\mathbf{x} \in \Upsilon$, the diameter of any group B_i is at most $2n^{-2b} < n^{-b}$ so the communication graph induced by their agents is frozen and remains so. The B-agents are attracted to $\mathbf{y}$.[17] This follows easily, as does the next lemma, whose proof we omit, from the stochasticity of Q and the identities: $\widetilde{Q}Q = Q\widetilde{Q} = \widetilde{Q}^2 = \widetilde{Q}$.

LEMMA 5.6. *The set Υ is forward-invariant. Furthermore, any $\xi \in \mathbf{y} + n^{-2b}\,\mathbb{I}^{n-m}$ belongs to Υ_B if and only if $\widetilde{Q}\xi = \mathbf{y}$.*

We set ρ, D_0 as in Lemma 5.3 and call an interval Δ *free* if it does not intersect the exclusion zone $W = W(\mathbf{y})$. For counting purposes, it is convenient to partition the perturbation space $n^{-b}\,\mathbb{I}$ into so-called *canonical* intervals of length ρ (with possibly a single smaller one). A gap of W can keep only $n^{O(n^5 D_0)}$ canonical intervals from being free, so the Lebesgue measure of the free ones satisfies:

$$\operatorname{Leb}\left\{\bigcup \text{free canonical intervals}\right\} \ge 2n^{-b} - \rho n^{O(n^5 D_0)}. \tag{22}$$

[16] We return to the rule used in §4 for convenience; we could use an arbitrarily small threshold instead.

[17] Although the B-agents in Υ_B have been essentially immobilized around $\mathbf{y}$, they are not decoupled from the rest. Indeed, while the increasingly microscopic movements of the B-agents can no longer affect their own communication graph, they can still influence the communication among the A-agents, even if none of the latter link to any B-agent.

Fixing the B-agent attractor. With $\mathbf{y}$ fixed, we pick a free canonical interval Δ and focus on the global coding tree $\mathcal{T}^{\Delta|\Upsilon}_{m\to n-m}$, with the superscripts indicating the perturbation and phase spaces, respectively. For any node v of depth $t_v \geq t_c$, the limit matrix D_u in Lemma 5.2 is the same for all nodes u of depth t_c. Indeed,

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

Pick v of depth $t_v \geq 3t_c$ and let w be its ancestor at depth $t_w = \lfloor t_v/2 \rfloor$. Given $\mathbf{x} \in U_v \subseteq \Upsilon$,

$$\begin{aligned} \mathbf{x}' = f^{t_w}(\mathbf{x}) = P_{\leq w}\,\mathbf{x} &\in \begin{pmatrix} C_w \\ \widetilde{Q} \end{pmatrix} (x_{m+1}, \ldots, x_n)^T + ne^{-\gamma t_w}\,\mathbb{I}^n \\ &\in \begin{pmatrix} C_w (x_{m+1}, \ldots, x_n)^T \\ \mathbf{y} \end{pmatrix} + ne^{-\gamma t_w}\,\mathbb{I}^n. \end{aligned}$$

We claim that $\mathbf{x}' \in \Upsilon$. Let $\underline{\mathbf{x}}'$ be the vector formed by the last $n-m$ coordinates of $\mathbf{x}'$. By (21),

$$\|\underline{\mathbf{x}}' - \mathbf{y}\|_\infty \leq ne^{-t_w n^{-O(n)}} < n^{-2b}.$$

Our claim follows then from the fact that $\widetilde{Q}(\underline{\mathbf{x}}' - \mathbf{y}) = \mathbf{0}$. By Lemma 5.6 and the Markovian property of the system, therefore, there exists a node v' of depth $t_{v'} = t_v - t_w \geq t_c$ such that,

$$\begin{aligned} f^{t_v}(\mathbf{x}) = f^{t_{v'}}(\mathbf{x}') &= P_{\leq v'}\,\mathbf{x}' \\ &\in \begin{pmatrix} C_{v'} \\ \widetilde{Q} \end{pmatrix} (\mathbf{y} + ne^{-\gamma t_w}\,\mathbb{I}^{n-m}) + ne^{-\gamma t_{v'}}\,\mathbb{I}^n \subseteq \begin{pmatrix} C_{v'}\mathbf{y} \\ \mathbf{y} \end{pmatrix} + 2ne^{-\gamma t_v/3}\,\mathbb{I}^n. \end{aligned}$$

It is important to note that v' depends only on v and not on $\mathbf{x} \in U_v$: indeed, the phase tube from U_v between time t_w and t_v does not split; therefore $f^{t_w}(U_v) \subseteq U_{v'}$. It follows that, for $t_v \geq 3t_c$ and $v' = v'(v)$,

$$V_v \subseteq \begin{pmatrix} C_{v'}\mathbf{y} \\ \mathbf{y} \end{pmatrix} + 2ne^{-\gamma t_v/3}\,\mathbb{I}^n. \tag{23}$$

The A-agents evolve toward convex combinations of the B-agents, which themselves become static. The weights of these combinations (ie, the barycentric coordinates of the A-agents), however, might change at every node, so there is no assurance that the orbit is always attracted to a limit cycle. The layer decomposition of the coding tree, which we describe next, allows us to bound the nesting time while exhibiting weak yet sufficient conditions for periodicity.

To stratify the coding tree $\mathcal{T}^{\Delta|\Upsilon}_{m\to n-m}$ into layers, we set up three parameters D_0, D_1, and D_2: the first targets the topological entropy; the second specifies the height of the first layer; the third indicates the nesting time. We examine each one in turn and indicate their purpose and requirements.

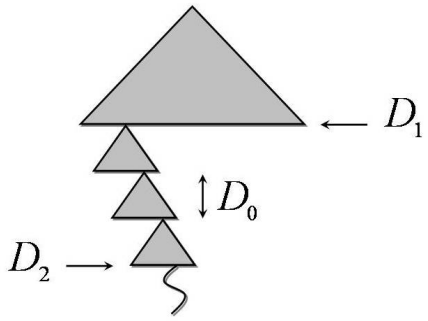

Figure 6: The global coding tree is stratified into three layers, with decreasing branching rates.

First layer. By (23), the phase tubes get thinner over time at a rate of roughly $e^{-\gamma/3}$, while the tree is branching at a rate of $n^{O(n)}$. To ensure that the topological entropy is zero, the product of these two rates should be less than 1: with $\gamma < 1$, this is far from being the case, so we need a sparsification mechanism. This is where Lemma 5.3 comes in. Indeed, deep enough in $\mathcal{T}^{\Delta|\Upsilon}_{m\to n-m}$, the size of a subtree of height D_0 will be shown to be at most

$$D_0(n^{O(n)})^{D_0^{1-\gamma^{n+1}}},$$

while the tubes get thinner at a rate of $2ne^{-\gamma D_0/3}$ for every consecutive D_0 nodes: the choice of D_0 below ensures that the product is less than 1, as desired. We justify this choice formally below.

$$D_0 \geq 2^{(1/\gamma)^{n+2}} \qquad [\ D_0 \textit{ big enough for thinning to outpace branching}\ ]. \tag{24}$$

Second layer. Technically, Lemma 5.3 addresses only the branching of the phase tube formed by a small box $\mathbf{x} + \rho\,\mathbb{I}^n$, for $\mathbf{x} \in \mathbb{C}[y_1, \ldots, y_r]$, whereas we are concerned here with phase tubes originating at some cell V_v of $\mathcal{T}^{\Delta|\Upsilon}_{m\to n-m}$. To make V_v thin enough, we choose a node v deep in the tree.[18] By (23), $V_v \subseteq \mathbf{x} + \rho\,\mathbb{I}^n$, for $\mathbf{x} \in \mathbb{C}[y_1, \ldots, y_r]$, provided that $t_v \geq D_1$ and

$$D_1 \geq \frac{3}{\gamma} \log \frac{2n}{\rho} \qquad [\ D_1 \textit{ big enough for tree branches to be thinner than } \rho\ ]. \tag{25}$$

Note that the requirement in (23) that $t_v \geq 3t_c = 3/\gamma$ is implied by $t_v \geq D_1$. In view of Lemma 5.3, the number of nodes in $\mathcal{T}^{\Delta|\Upsilon}_{m\to n-m}$ of depth no greater than $t \geq D_1$, is bounded by

$$\underbrace{n^{O(nD_1)}}_{\texttt{depth D1}} \times \underbrace{n^{O(nD_0^{1-\gamma^{n+1}}\lfloor (t-D_1)/D_0 \rfloor)}}_{\texttt{from D1 to t in chunks of D0}} \times \underbrace{n^{O(nD_0)}}_{\texttt{truncated chunk}} \times \underbrace{D_0}_{\texttt{single paths}}.$$

[18] Factoring out the B-agents gives us the sort of fixed-point attraction that is required by Lemma 5.3: it is a dimension reduction device in attractor space.

To see why, treat each path of single-child nodes as a single edge. Thus, for any $t \geq D_1$,

$$\left| \{ v \in \mathcal{T}^{\Delta|\Upsilon}_{m \to n-m} \,|\, t_v \leq t \} \right| \leq n^{O(nD_0 + nD_1 + ntD_0^{-\gamma^{n+1}})}. \tag{26}$$

THIRD LAYER. The bottom layer of the stratified global coding tree begins at a depth $D_2 \geq D_0 + D_1$. If the node v of depth $t_v \geq D_2$ has more than one child, then, by continuity, V_v contains a point right on the boundary of the global margin. By (23), this implies the existence of $\zeta \in \mathbb{R}^n$ such that $\|\zeta\|_\infty \leq 2ne^{-\gamma D_2/3}$ and $\mathrm{Lin}\,[\mathbf{y} + \zeta] = 1 + \delta$, where the coefficients of the linear form are of magnitude $n^{O(1)}$ and depend only on the node v and a margin's hyperplane. It then follows from (26) that $\mathcal{T}^{\Delta'|\Upsilon}_{m \to n-m}$ has no nontrivial branching at depth D_2, provided that $\Delta' = \Delta \setminus W'$, where W' consists of gaps of type $n^{O(1)}e^{-\gamma D_2/3}$ numbering at most

$$\underbrace{n^{O(nD_0 + nD_1 + nD_2 D_0^{-\gamma^{n+1}})}}_{\#\,\texttt{nodes at depth D}_2} \times \underbrace{n^{O(1)}}_{\#\,\texttt{margin hyperplanes}}.$$

This sets a bound of D_2 on the nesting time. It follows that

$$\mathrm{Leb}\,(W') \leq e^{-\gamma D_2/3} n^{O(nD_0 + nD_1 + nD_2 D_0^{-\gamma^{n+1}})}. \tag{27}$$

Pick a large enough constant $d = d(b, c)$; recall that $\gamma = n^{-cnt_o}$. We set the parameters $\rho = n^{-dn^5 D_0}$, where, rounding up to the nearest integer,

$$\begin{cases} D_0 = 2^{d(1/\gamma)^{n+2}} \\ D_1 = d^2 n^6 D_0/\gamma \\ D_2 = dn^2 D_1/\gamma. \end{cases} \tag{28}$$

We verify that conditions (24, 25) are both satisfied and that

$$D_1 \geq D_2 D_0^{-\gamma^{n+1}}. \tag{29}$$

Thus the measure bound (27) implies that $\mathrm{Leb}\,(W') \leq \rho 2^{-D_0}$. Since Δ has length ρ, this implies that, with probability at least $1 - 2^{-D_0}$, subjecting the system's margin to a perturbation δ chosen randomly in Δ makes phase tube splitting impossible at time D_2: we call this *success*. In these conditions, by Lemma 3.1 and the fact that absorption is determined only by time differences, splitting past time D_2 is ruled out, too: any infinite orbit (ie, nonvanishing or nonabsorbing) from $\mathbf{x} \in \Upsilon$ is attracted to a limit cycle.[19] The full perturbation space is not Δ but $n^{-b}\,\mathbb{I}$, so we apply the previous result to each free canonical interval and argue as follows. If Λ is the measure of the union of all the free

[19] Vanishing occurs with probability zero.

canonical intervals, then the perturbations that do not guarantee success have measure at most $(2n^{-b} - \Lambda) + 2^{-D_0}\Lambda$. Dividing by $2n^{-b}$ and applying (22) shows that

$$\mathrm{Prob}_\delta\,[\,\text{failure in } \mathcal{T}_{m\to n-m}^{n^{-b}\mathbb{I}\,|\,\Upsilon}\,] \le 1 - (1-2^{-D_0})(\,1 - \rho n^{O(n^5 D_0)}\,) \le 2^{1-D_0}. \tag{30}$$

Let $\mathbb{I}_s$ denote the set of successful perturbations within $n^{-b}\,\mathbb{I}$. The nesting time is at most D_2, which, by (26, 29), implies that

$$h(\mathcal{T}_{m\to n-m}^{\mathbb{I}_s\,|\,\Upsilon}) \le O(D_1 n \log n) \le \gamma^{-1} n^{O(1)} D_0\,. \tag{31}$$

At depths D_2 and higher, the coding tree can no longer branch but its paths can still end in absorbing leaves. The period, preperiod, number of absorbing leaves, and maximum depth of an absorbing leaf are all bounded by[20]

$$2^{h(\mathcal{T}_{m\to n-m}^{\mathbb{I}_s\,|\,\Upsilon})} + t_o\,. \tag{32}$$

Freeing the B-agents. Set $D_3 = \lceil 3b\gamma^{-1}\log n\rceil$ and fix $\mathbf{x}$ in $\Omega = (0,1)^n$. Let ξ denote the projection of $f^{D_3}(\mathbf{x})$ onto the last $n-m$ coordinate axes. By Lemma 5.2, the coding tree $\mathcal{T}_{m\to n-m}^{n^{-b}\mathbb{I}}$ has $n^{O(nt_c)}$ nodes u such that $t_u = t_c$ and

$$\xi \in \mathbf{y} + ne^{-\gamma D_3}\,\mathbb{I}^{n-m} \subseteq \mathbf{y} + n^{-2b}\,\mathbb{I}^{n-m},$$

where $\mathbf{y} = D_u(x_{m+1},\ldots,x_n)^T$. The state vector for the B-agents is ξ at time D_3 and $Q^{t-D_3}\xi$ at $t > D_3$, where Q is the transition matrix of the frozen communication subgraph joining the B-agents at time D_3. By taking t to infinity, it follows that $\mathbf{y} = \widetilde{Q}\xi$ (note that $\widetilde{Q}$ may not be the same as D_u) and, by Lemma 5.6, $\xi \in \Upsilon_B$ hence $f^{D_3}(\mathbf{x}) \in \Upsilon$. We can then apply the previous result. Since $\mathbf{x}$ is fixed, only the choice of random perturbation δ can change which path in $\mathcal{T}_{m\to n-m}^{n^{-b}\mathbb{I}}$ the orbit will follow. The failure probability of (30) needs to be multiplied by the number of nodes u, which yields an upper bound of $n^{O(nt_c)}2^{1-D_0}$; hence

$$\mathrm{Prob}_\delta\,[\,\text{failure in } \mathcal{T}_{m\to n-m}^{n^{-b}\mathbb{I}}\,] \le 2^{-D_0/2}. \tag{33}$$

If $\mathcal{T}^*$ denotes the part of the global coding tree extending to depth D_3 and $\mathbb{I}'_s$ the new "success" perturbation set, then

$$\mathcal{T}_{m\to n-m}^{\mathbb{I}'_s} = \mathcal{T}^* \otimes \mathcal{T}_{m\to n-m}^{\mathbb{I}'_s\,|\,\Upsilon}\,.$$

The upper bound on the number of absorbing leaves given in (32) still holds with $\mathbb{I}_s$ replaced by $\mathbb{I}'_s$. The tree $\mathcal{T}^*$ has at most $n^{O(nD_3)}$ nodes; therefore, by (31),

$$\begin{cases} \nu(\mathcal{T}_{m\to n-m}^{\mathbb{I}'_s}) \le D_2 + D_3 \\ h(\mathcal{T}_{m\to n-m}^{\mathbb{I}'_s}) \le h(\mathcal{T}_{m\to n-m}^{\mathbb{I}'_s\,|\,\Upsilon}) + O(D_3\, n\log n) \le \gamma^{-1} n^{O(1)} D_0\,. \end{cases} \tag{34}$$

[20] The additive term t_o is only needed for the depth bound.

5.4 Removing persistence

Since we do not know the graph H ahead of time, we initialize it to the complete directed graph and update it at each absorbing node by removing the edge(s) whose missing presence causes the absorption. This yields the rewriting rule[21]

$$\mathcal{T} \Longrightarrow \bigotimes_{k=1}^{k_0} \mathcal{T}_{m_k \to n-m_k}, \tag{35}$$

where $k_0 \leq n(n-1)$. To keep the failure probability from being amplified by each product, we reset D_0 in (28) at every iteration: to do so, we define C_k as its suitable value for a persistent graph consisting of k (nonloop) directed edges and let ϕ_k denote the maximum failure probability for such a graph: $C_{n(n-1)} \geq D_0$ and $\phi_0 = 0$; and, by (33, 34), for $k > 0$,

$$\phi_k \leq 2^{-C_k/2} + 2^{\gamma^{-1} n^a C_k} \phi_{k-1},$$

for some constant $a > 0$. Setting $C_{n(n-1)-j} = \gamma^{-j} n^{2aj} D_0$, for $j = 0, \ldots, n(n-1)$, we verify by induction that $\phi_k \leq 2^{1-C_k/2}$, for $k = 0, \ldots, n(n-1)$; hence,

$$\mathrm{Prob}_\delta\,[\,\text{failure}\,] \leq \phi_{n(n-1)} \leq 2^{1-D_0/2}.$$

The period and preperiod are bounded by

$$\prod_{k=0}^{n(n-1)} \left(2^{\gamma^{-1} n^a C_k} + t_o\right) \leq 2^{C_0 \gamma^{-1} n^{O(1)}},$$

which grows polynomially in 2^{D_0}, hence in the reciprocal of the failure probability (which can be made arbitrarily small); the dependency on n grows much faster, of course. This completes the proof of the case $d = \mathtt{d} = 1$ of the nonbidirectional case of Theorem 1.1. □

5.5 Arbitrary dimension and algebraic degree

The analysis readily extends to any dimension d and degree $\mathtt{d}$ via the tensor lifting construction of §2.4. Recall that the two essential ingredients are: (i) the thinning rate (Lemma 5.1); and (ii) sparse branching (Lemma 5.3). We easily check that both conditions still hold (though with different rates). The matrix Q_c associated with a cell c of the *SP* is of the form $(P \otimes \mathbf{I}_d)^{\otimes \mathtt{d}}$, with $P = P(\mathbf{x})$ whenever $\mathbf{y} = \mathbf{y}(\mathbf{x})$ for some $\mathbf{y} \in \mathcal{V}$. We retain the coding tree type $\mathcal{T}_{m \to n-m}$ over the m (resp. $n-m$) ground A-agents

[21] We cannot require that the influence system reset to 0 the timeout counts of its edges at each direct product. There is no need to do so anyway: no resetting means that absorption (and hence the corresponding direct product) might come earlier. Since our calculations do not depend on the actual number of absorbing leaves but only the number of nodes of depth no greater than the nesting time, adding earlier absorptions is of no consequence; delaying them, of course, would be a different matter.

(resp. B-agents). This induces a block-directional decomposition of the lifted system. Recall that, given $\mathbf{y} \in \mathcal{V}$, the (lifted) agent l is at position $\mathbf{y}_l = \prod_{i=1}^{\mathtt{d}} x_{k_i,j_i}$, where l is the lexicographic rank of $(k_1, j_1, \ldots, k_{\mathtt{d}}, j_{\mathtt{d}})$. The lifted agents fall into two groups: $\underline{A}$ consists of the agents l with at least one constitutive ground agent k_i in A; the others form the group $\underline{B}$. An edge in the power graph from $(k_1, j_1, \ldots, k_{\mathtt{d}}, j_{\mathtt{d}})$ to $(k'_1, j'_1, \ldots, k'_{\mathtt{d}}, j'_{\mathtt{d}})$ requires the presence in the ground graph of the $\mathtt{d}$ edges $(k_1, k'_1), \ldots, (k_{\mathtt{d}}, k'_{\mathtt{d}})$. This shows that the lifted coding tree is of type $\mathcal{T}_{\underline{A} \to \underline{B}}$. Any transition matrix entry corresponding to an edge from $\underline{A}$ to $\underline{A}$ is of the form $p_{i_1 j_1} \cdots p_{i_{\mathtt{d}} j_{\mathtt{d}}}$, with at least one factor $p_{i_k j_k}$ such that i_k and j_k are both ground A-agents. The extension of Lemma 5.1 to the lifted system, with its transition matrices of the form $(P(\mathbf{x}) \otimes \mathbf{I}_d)^{\otimes \mathtt{d}}$, follows immediately.

By lifting the thinning rate argument as we just did, we implicitly assumed that the agent position $\mathbf{y}$ lay on the algebraic variety $\mathcal{V}$, i.e., $\mathbf{y} = \mathbf{y}(\mathbf{x})$ for some $\mathbf{x}$. This need not be case. A simple fix is absorb any orbit that strays from $\mathcal{V}$. Specifically, we turn into an absorbed leaf any node v whose parent w is such that $V_v \cap f(\mathcal{V} \cap V_w) = \emptyset$. Note that these absorbed leaves are terminal and not the contact points of direct products. This completes the proof of Theorem 1.1. □

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.

[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.