Diffusive Influence Systems *

Bernard Chazelle †

Abstract

Influence systems seek 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 widely used family of multiagent systems, we introduce a general renormalization method for the bifurcation analysis of multiagent systems.

1 Introduction

Consider a set of "agents" on a line, each one represented by a point. Starting out at an arbitrary position on the line, every agent moves to the mass center of the agents within unit distance. The agents move simultaneously and repeat this procedure indefinitely. Does the system eventually come to a halt? This multiagent model, known as a *Hegselmann-Krause* (*HK*) system [28] is by now reasonably well-understood [30, 37, 42, 51]: in particular, the motion is known to converge in polynomial time [5, 11, 19]. This remains true if the agents are modeled as points in higher-dimensional Euclidean space. The model lends itself to natural extensions. For example, the agents might have different distance thresholds. These so-called *heterogeneous HK* systems break the symmetry of the neighbor relations and imply a *directed* communication graph. Naturally, the dynamics becomes more complicated. Experiments suggest that the system still converges but a proof has remained elusive; in fact, even chaos cannot be ruled out.

Further extensions of the model seek to enrich the communication dynamics. For instance, instead of claiming as a neighbor any agent lying within a certain distance, ie, satisfying a single inequality, one could require the satisfaction of several constraints: in the *HK* system, a neighbor could be defined as any agent at distance between 0.5 and 1. Climbing up the expressivity ladder, one could allow any sentence in first-order logic over the reals. This level of generalization might seem far-fetched but it is actually quite common. A popular choice of communication graph in rendezvous problems in

*This work was supported in part by NSF grants CCF-0832797, CCF-0963825, and CCF-1016250. A preliminary version of this work appeared in Proc. 53rd FOCS 2012.

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

robotics is the constrained Delaunay graph [11]: to specify an edge in this case requires a first-order sentence with two levels of quantification.[1] Dozens of systems of this kind have been studied in fields as diverse as biology, zoology, economics, and sociology. We introduce a model of multiagent dynamics, called *influence systems*, to bring all of these special cases under a common umbrella. As long as the dynamic rule involves moving each agent to the mass center of its neighbors in the communication graph, the system is called *diffusive*. Not all multiagent dynamics follow this pattern, however: flocking and swarming models are classical examples of *nondiffusive* influence systems.

With rare exceptions, the precise dynamics of specific influence systems remains a mystery. We prove the surprising result that, under random perturbations, diffusive influence systems are almost always asymptotically periodic. Furthermore, if the communication stays bidirectional at all times then the system converges to a fixed-point attractor.[2] What makes this unexpected is that, with proper finetuning, a diffusive influence system with even just a few agents can be chaotic or even simulate a Turing machine. The challenge is to show that this behavior is extremely rare. The broad sweep of our result suggests the value of a more general, abstract approach to multiagent dynamics. The proof introduces a broadly applicable recursive technique for bifurcation analysis, somewhat reminiscent of the renormalization group in statistical physics.

Next, we define the model formally and state our main result. We then bring this introduction to an end with a brief discussion of the significance and historical background of our contribution.

The model. An *influence system* consists of n agents labeled $1, \ldots, n$. Agent i is modeled by a point $x_i \in \mathbb{R}^d$ ($d > 0$). We denote the state of the system by $\mathbf{x} = (x_1, \ldots, x_n)$. By convention, each x_i is a row vector with d coordinates, so that $\mathbf{x} \in \mathbb{R}^{n \times d}$ can be thought of as an n-by-d matrix whose rows indicate the positions of the agents. The dynamics of the system is specified by a function f: if $\mathbf{x}$ is the system's state at time t, then $f(\mathbf{x})$ indicates its state at time $t+1$. To understand an influence system is to understand the geometry of the orbit $\mathbf{x}$, $f(\mathbf{x})$, $f^2(\mathbf{x})$, $f^3(\mathbf{x}), \ldots$, for any initial state $\mathbf{x}$ in $\mathbb{R}^{n \times d}$. The defining feature of an influence system is that the map f relies on an intermediate function $\mathcal{G}$ specifying the communication channels:

- The function $\mathcal{G}$ maps a state $\mathbf{x}$ to a *communication graph* whose nodes are in bijection with the agents. The existence of an edge (i, j) is determined by the truth condition of a first-order sentence over the reals:

$$\Phi_{i,j}(\mathbf{x}) := (\square\, y_1) \cdots (\square y_k)\ \mathcal{L}_{i,j}(\mathbf{x}, y_1, \ldots, y_k), \tag{1}$$

  where $\square \in \{\exists, \forall\}$ and $\mathcal{L}_{i,j}$ is a quantifier-free logical proposition whose clauses are (strict and nonstrict) algebraic inequalities with rational coefficients. The free

[1] Given a boundary B, an edge from p to q is expressed as: "$(\exists c \in \mathbb{R}^d)\ (\forall x \in B)\ \|cx\|_2 \geq \|cp\|_2 = \|cq\|_2$ & $\|cr\|_2 \geq \|cp\|_2$ (for each agent r)."

[2] The result suggests (at least in theory) that bidirectional influence systems like FaceBook lead to stable communities while adding one-way communication (eg, talk shows, newscasts) may create oscillations.

variables are given by the entries of $\mathbf{x}$ and the bound variables consist of $y_1, \ldots, y_k$. The pair (i, j) is an edge of $\mathcal{G}(\mathbf{x})$ if and only if $\Phi_{i,j}(\mathbf{x})$ is true.

- For $i = 1, \ldots, n$, the i-th row of the n-by-d matrix $f(\mathbf{x})$ depends only on the pairs (j, x_j) for all edges (i, j) of $\mathcal{G}(\mathbf{x})$. In other words, the next position of an agent can be computed if we know both the label and the position of each one of its neighbors.

- The influence system is called *diffusive* if the next position of any agent i is a convex combination of its neighbors' positions. Stated formally, $f(\mathbf{x}) = P(\mathbf{x})\mathbf{x}$ for any state $\mathbf{x} \in \mathbb{R}^{n \times d}$, where $P(\mathbf{x})$ is a stochastic matrix such that $P(\mathbf{x})_{ij} > 0$ if and only if (i, j) is an edge of $\mathcal{G}(\mathbf{x})$. We assume that the diagonal entries of $P(\mathbf{x})$ are positive and that the i-th row of $P(\mathbf{x})$ is uniquely determined by $\{\, j \mid (i, j) \in \mathcal{G}(\mathbf{x}) \,\}$.[3] The system is called *bidirectional* if the graph $\mathcal{G}(\mathbf{x})$ is undirected for all $\mathbf{x}$.

We illustrate these definitions with a few examples. The classical *HK* system is bidirectional: $d = 1$ and $\mathcal{G}(\mathbf{x})$ is the n-node graph linking any two agents at distance one or less. The graph is undirected and the "transition" matrix $P(\mathbf{x})$ is the adjacency matrix of $\mathcal{G}(\mathbf{x})$, with each row scaled down so it sums up to 1. The *truth-seeking systems* [29,35] from social epistemology add to the *HK* model a fixed agent to which a subset of the agents are always connected. This adds one row $(1, 0, \ldots, 0)$ to the adjacency matrix as well as one column of mixed zeroes and ones; the system is not bidirectional. While the *HK* model captures the mode of social interaction among agents engaged in two-way communication, the truth-seeking variant adds a one-way source of information such as TV or radio.

For a simple example of a nondiffusive influence system, we can pick a standard flocking model [18, 23, 31, 58].[4] Each bird i has position $p_i = (p_i^x, p_i^y, p_i^z)$ and velocity $v_i = (v_i^x, v_i^y, v_i^z)$. At the next step, $p_i \leftarrow p_i + v_i$ and $v_i \leftarrow (1/N_i) \sum_j v_j$, where the sum extends over the N_i birds j within a fixed distance of bird i. The state $\mathbf{x}$ is of the form $(x_1, \ldots, x_n)$, where x_i is the row vector $(p_i^x, p_i^y, p_i^z, v_i^x, v_i^y, v_i^z)$. It is obvious that the map of the dynamical system cannot be expressed by a stochastic matrix. If we rearrange the n-by-6 matrix $\mathbf{x}$ by stacking the leftmost half on top of the rightmost half, then $\mathbf{x}$ becomes the $2n$-by-3 matrix whose first (resp. last) n rows are (p_i^x, p_i^y, p_i^z) (resp

[3] The positivity assumption, the same one found in lazy random walks, adds self-loops to the communication graphs. It is a minor technicality whose motivation is to avoid trivial periodicity, such as two agents switching positions endlessly. The second assumption is mostly for notational convenience and can be relaxed in several ways.

[4] Examples abound. The states of an influence system can be: opinions [5,7,8,25,28,29,33,35–39,41], Bayesian beliefs [1], neuronal spiking sequences [17], animal herd locations [22], consensus values [14,42], chemotactic responses [47] swarming trajectories [27], cell populations [49], schooling fish velocities [45, 47], sensor networks data [11], synchronization phases [13, 26, 48, 52], heart pacemaker cell signals [56, 60], cricket chirpings [59], firefly flashings [40], yeast cell suspensions [49], microwave oscillator frequencies [56], or flocking headings [4, 18, 23, 30, 31, 50, 58].

$(v_i^x, v_i^y, v_i^z))$, for $i = 1, \ldots, n$, and

$$P(\mathbf{x}) = \begin{pmatrix} \mathbf{I}_n & \mathbf{I}_n \\ \mathbf{0} & Q_{\mathbf{x}} \end{pmatrix},$$

where $Q_{\mathbf{x}}$ is a stochastic matrix and $\mathbf{I}_n$ is the n-by-n identity matrix. It follows that the velocity part of the system is itself diffusive. To express the graph on the basis of velocities alone, however, would require integration over the past, which would produce a non-Markovian system.

The results. Our main theorem requires a notion of perturbation. To perturb an *HK* system, we replace the distance threshold of 1 by $1 + \delta$, where δ is picked randomly between 0 and an arbitrarily small constant $\delta_0 > 0$. More generally, we perturb an arbitrary influence system similarly by subjecting its constraints to a tiny random shift. To explain how to do this, it is useful to rephrase the formulation in (1) in terms of semi-algebraic sets. By a celebrated result of Collins [21], the quantifiers of any first-order sentence over the reals can be eliminated. In particular, there exists a finite set of dn-variate polynomials $\mathcal{P}_1, \ldots, \mathcal{P}_N$ such that whether $\Phi_{i,j}(\mathbf{x})$ is true or false is entirely a function of the N-tuple of signs of $\mathcal{P}_1(\mathbf{x}), \ldots, \mathcal{P}_N(\mathbf{x})$, ie, the *sign-conditions* of $\mathbf{x}$. We call the $\mathcal{P}_i$'s the *constraint polynomials*. This process is best explained with a simple example. Consider a two-agent influence system in one dimension: $n = 2$ and $d = 1$. An edge joins the two agents 1 and 2 at positions x_1 and x_2, respectively, if and only if the sentence $\Phi(x_1, x_2)$ is true, where $\Phi(X, Y)$ is defined as

$$(\exists\, u)(\exists v)\ (u > 0)\ \&\ (v > 0) \\ \&\ \big(X^3 + XY^2 + uX^2 + uY^2 - (v+1)X - u(v+1) = 0\,\big).$$

The polynomial factorizes into $(X + u)(X^2 + Y^2 - v - 1)$, so we easily verify that the sentence is logically equivalent to the quantifier-free proposition:

$$X < 0 \quad \text{or} \quad X^2 + Y^2 - 1 > 0. \tag{2}$$

It follows that $N = 2$, $\mathcal{P}_1 = X$, and $\mathcal{P}_2 = X^2 + Y^2 - 1$. Quantifier elimination decomposes the phase space into regions (semi-algebraic sets) with the same sign-conditions. Each such region c is associated with a communication graph and a stochastic transition matrix: in other words, $\mathcal{G}(\mathbf{x})$ and $P(\mathbf{x})$ remain the same for all $\mathbf{x} \in c$. Perturbing the system means picking a random δ in an arbitrarily small interval near 0 and replacing the semi-algebraic set (2) by

$$X + \delta < 0 \quad \text{or} \quad X^2 + Y^2 - 1 + \delta > 0.$$

In general, one would take all the polynomials $\mathcal{P}_1, \ldots, \mathcal{P}_N$ from Collins's quantifier elimination and replace them by $\mathcal{P}_1 + \delta, \ldots, \mathcal{P}_N + \delta$. A separate perturbation is needed for each initial state since no single value of δ is likely to work over the entire phase space;

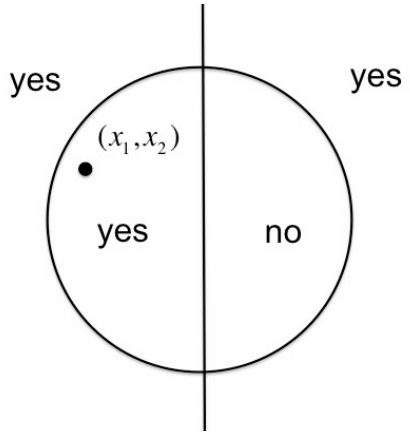

Figure 1: The existence of the edge $(1, 2)$ is determined by the location of the point (x_1, x_2).

note that there is no need to perturb the initial state $\mathbf{x}$ of the system. The probabilistic language below refers to the random perturbation of the constraint polynomials described above. We now state our main result:

THEOREM 1.1. *Given any initial state, the orbit of a perturbed influence system is asymptotically periodic almost surely. Without perturbation, the model can be chaotic and even Turing-complete. In the bidirectional case, the system is always attracted to a fixed point: if ρ is the lowest positive entry of any $P(\mathbf{x})$ then, with probability arbitrarily close to 1, any initial state $\mathbf{x}$ in the unit cube evolves to within distance ε of its fixed-point attractor in $\rho^{-O(n)}|\log \varepsilon|$ steps, for any $\varepsilon > 0$.*

The theorem should be understood in the spirit of smoothed analysis [55]: it says that, while an influence system can be arbitrarily complex in the worst case, the tiniest perturbation will nudge it into a limit cycle. We do not investigate the precise geometric nature of the critical region (ie, the values of δ for which the system becomes chaotic) but we show how to cover it by a fractal structure (a type of Cantor set). To prove that perturbation is indeed required, we design influence systems that can produce chaos and simulate Turing machines even in dimension $d = 1$, using constraint polynomials $\mathcal{P}_i$ of degree 1. A significant difference with standard smooth analysis is that even small influence systems can be extremely hard to analyze. This makes it all the more surprising that a dynamic classification of influence systems can be carried out for all n.

In addition to shifting each constraint randomly by δ, we also apply a *perturbation rule*: (i) the status of an edge (i, j) cannot change while agents i, j stay infinitesimally close to each other; and (ii) no edge that disappears indefinitely can return. This last rule (which is not needed in the bidirectional case) prevents an edge from reappearing after an absence exceeding an arbitrarily large threshold. This does not keep agents from shedding edges whenever the function $\mathcal{G}$ calls for it, so it is not a roundabout way

to enforce connectivity. Nor is it a heuristic assumption about the behavior of the system: the perturbation rule can be enforced by the agents themselves. We show in the next section that some form of perturbation rule is required for Theorem 1.1 to hold. Our formulation can probably be weakened but it cannot be removed altogether. (We're currently in the process of pinning down the minimum set of conditions required.) In the bidirectional case, the convergence time is a function of the perturbation probability; the bound, which matches the performance of bidirectional multiagent agreement dynamics [19], is provably false without perturbation.

We assume that all the relevant parameters (matrix entries, number and coefficients of polynomials) can be encoded as rationals over $O(\log n)$ bits and that the dimension d and the degrees of the polynomials are bounded by constants: we do this mostly for notational convenience and some of these assumptions can be relaxed if necessary.

Background and previous results. An overarching ambition of social dynamics is to understand and predict the collective behavior of agents influencing one another across an endogenously changing network [11, 13, 15, 45, 47]. Influence diffusive systems provide a versatile platform for such investigations and unify into a common framework many variants found in the literature (eg, bounded-confidence, bounded-influence, truth-seeking, Friedkin-Johnsen type, deliberative exchange) [6, 20, 37, 42].

The fixed-point attraction of bidirectional diffusive influence systems follows from known results [19, 25, 30, 37, 42] but convergence times are known only in a few cases [11, 19]. Without bidirectionality, known convergence results are conditional [14, 16, 17, 31, 41–43, 46, 57].[5] 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. Our main result allows us to remove all such assumptions. It can be seen as an extension of Bruin and Deane's unconditional resolution of planar piecewise contractions [10], which are special kinds of influence systems with a single mobile agent.

In the context of social dynamics, our results might be disconcerting. Influence systems can be used to 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.

A technical remark. Following their introduction by Sontag [54], piecewise-linear systems have become the subject of an abundant literature, which we do not attempt to review here. They are known to be Turing-complete [2, 9, 32, 53]. A typical simulation relies on the existence of Lyapunov exponents of both signs, negative ones to move the head of the Turing machine in one direction and positive ones to move it the other way. Diffusive influence systems have no positive exponents, so the same construction cannot work. Chaos is equally problematic, being typically associated with positive topological

[5] As they should be, since convergence is not assured. An exception is *truth-seeking HK systems*, which have been shown to converge unconditionally [19, 29, 35].

entropy, hence positive Lyapunov exponents. Piecewise linearity blurs the picture. With only null exponents (over any ergodic invariant measure), piecewise isometries actually have zero topological entropy [12]. Oddly, contractions can be chaotic [34], even though they have only negative exponents. Diffusive influence systems sit in the middle, having null and negative exponents but no positive ones. Yet they can be chaotic. Plainly, the spectral lens breaks down in the face of piecewise linearity. To resolve this confusion requires a different approach. We provide one in the form of an algorithmic brand of bifurcation analysis.

2 Chaos, Turing, and Perturbation

We show that diffusive influence systems can have periodic orbits of length exponential in the number of agents and that such long orbits are resistant to perturbation; next, we build a conjugation with the baker's map to exhibit chaos; finally, we show how to simulate a Turing machine. All three constructions use linear constraints. This section gives us a chance to discuss simple examples of diffusive influence systems and give the reader a feel for some of the issues ahead including the necessity of perturbations.

Exponential 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 constant bit-length per agent. This is noteworthy because complex behavior in dynamical systems usually requires infinite precision. We simulate a counter in binary notation. The system uses n agents and has periodic points of period 2^{n-2}. Initially, $x_1 = -1$, $x_2 = 2$, and $x_3 = \cdots = x_n = 0$. Agents 1 and 2 are fixed; for $k = 3, \ldots, n$,

$$x_k \leftarrow \begin{cases} \frac{1}{2}(x_1 + x_k) & \text{if } x_i > 0.5 \ (2 < i \leq k) \\ \frac{1}{2}(x_2 + x_k) & \text{if } x_k \leq 0.5 \text{ and } x_i > 0.5 \ (2 < i < k) \\ x_k & \text{else.} \end{cases}$$

Agent 3 oscillates between 0 and 1 at each step. Agent 4 does likewise but switches only with $x_3 = 1$. Likewise, agent 5 switches between 0 and 1 only when $x_3 = x_4 = 1$; etc. This produces an $(n-2)$-bit binary counter. The only constraints are given by the polynomials $x_i - 0.5$: replacing them by $x_i - 0.5 + \delta$ perturbs the system but has no effect on the counter as long as δ is not too large. The exponentially long period requires only a constant number of bits per agent and is resistant to perturbation. The system is not bidirectional. We summarize our result:

> *A diffusive influence system with n agents can have periodic points of period exponential in n, while using only a constant number of bits per agent. The long periodic orbit is robust under perturbation.*

Chaos. We describe a four-agent diffusive influence system in one dimension that exhibits chaotic behavior robust even under perturbation. This is caused by the microscopic dynamics of infinitesimally close agents, which points to the necessity of "perturbation rules" to avoid this pathological kind of chaos. The first two agents stay on opposite sides of the origin (assuming they start that way) with agent 1 to the left and agent 2 to the right. The rule is that the agent further from the origin moves toward it while the other one stays put:

$$(x_1, x_2) \;\leftarrow\; \tfrac{1}{2} \begin{cases} (\,2x_1, x_1 + x_2\,) & \text{if } x_1 + x_2 \geq 0 \\ (\,x_1 + x_2, 2x_2\,) & \text{else.} \end{cases} \tag{3}$$

The two agents converge toward 0 but the order in which they proceed (ie, their symbolic dynamics) is chaotic. To see why, 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:

$$u \leftarrow \begin{cases} \frac{1}{2}(u+1) & \text{if } u \leq -1 \\ \frac{2u}{u+1} & \text{if } -1 < u \leq 0. \end{cases}$$

Writing $u = (v+1)/(v-1)$ gives $v \leftarrow 2v + 1$ if $v < 0$, and $v \leftarrow 2v - 1$ otherwise. The system v escapes for $|v(0)| > 1$ and otherwise conjugates with the baker's map [24]. To turn this into actual chaos, we fix agent 4 at $x_4 = 1$ and we initialize agents 1 and 2 close enough to 0. We let agent 3 oscillate in $[x_1, x_4] \approx [0, 1]$, depending on the order in which the first two agents move:

$$x_3 \leftarrow \begin{cases} \frac{1}{3}(x_3 + 2x_1) & \text{if } x_1 + x_2 \geq 0 \\ \frac{1}{3}(x_3 + 2x_4) & \text{else.} \end{cases}$$

Agent 3, initially at 0, stays either lower than 0.4 or higher than 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. The system is not bidirectional but the subsystem formed by the first two agents is.

> *A diffusive influence system with only four agents can have chaotic behavior that is robust under perturbation.*

Turing completeness. Given an 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.[6] 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 $\overline{\mathbf{x}} = (\mathbf{x}, -\mathbf{x}, x_{-1}, x_0)$. To cope with scaling, we rewrite any hyperplane $\mathbf{a}^T\mathbf{x} = a_0$ (projectively) 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. The purpose of this construction is to show that, up to scaling, a piecewise-linear system with hyperplane constraints can be simulated by a diffusive influence system with linear constraints. Because the conjugated system is scaled down by $r \leq 1$ at each step, the absence of any affine term from these constraints is crucial. The reduction fails if we apply a random shift to them; hence it is not robust under perturbation.

Koiran et al [32] have shown how to simulate a Turing machine with a single agent in two dimensions using piecewise-affine maps. We can turn the construction into a one-dimensional piecewise-linear system with a number of agents proportional to the number of states. We then use the previous construction to turn it into a diffusive influence system. We only sketch the idea, which is a minor modification of Koiran et al [32]. One agent is positioned at the value given by the contents of the tape (interpreted as a number in binary or some other radix notation) to the left of the head. The other agent models the tape's contents to the right of the head, while saving the higher-order bits to encode the current state. A transition of the Turing machine translates into affine transformations $z \mapsto az + b$ applied to the two agents' positions. We linearize them by writing b as bx_0 and creating an agent 0 permanently stationed at $x_0 = 1$. If instead of using a single agent for all the affine terms, we introduce a distinct one, always positioned at 1, for each affine transformation in the system, we thus satisfy our assumption that the row of each transition matrix be uniquely specified by the corresponding node labels in the communication graph. We omit the details, which are routine. Predicting nontrivial state properties of a diffusive influence system (such as basic connectivity properties of the communication graph) is therefore undecidable.

> *A diffusive influence system can be Turing-complete. The construction is not robust under perturbation.*

[6] 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.

The role of perturbation. The fragility of the Turing machine simulation by a diffusive influence system echoes the main message of Theorem 1.1, which is that the slightest perturbation will nudge any orbit into a limit cycle. The chaotic four-agent system we designed earlier belies that fact, however. Replacing $x_1 + x_2 \geq 0$ in (3) by $x_1 + x_2 + \delta \geq 0$ does not alter the chaotic behavior of the system. Chaos survives large perturbations. To prevent it therefore requires the *perturbation rule* discussed in §1 or some variant of it. Once x_1 and x_2 differ by less than an arbitrarily small amount $\varepsilon_0 > 0$, the current edge joining them stays in place as long as the agents remain within ε_0 of each other (in this case, ever after): the evolution of $x_1 + x_2$ becomes predictable and chaos is broken.

We briefly mention two additional reasons for perturbing the system. The first is that the convergence time bound of Theorem 1.1 is provably false without perturbation. To see why, consider this three-agent system in one dimension. The first two agents move toward each other according to the rule:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leftarrow \tfrac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{4}$$

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. The system is bidirectional so this particular feature does not help in this case.

Another reason for perturbing the constraint polynomials is that, without it, a system can converge *physically* but not *combinatorially*. This is the case with heterogeneous *HK* systems [28,29], as we show now: the agents converge toward a fixed point but their communication graph keeps switching forever without perturbation. Heterogeneous *HK* systems are diffusive influence systems where each agent i is associated with a threshold distance r_i and the communication graph links i to any agent j such that $|x_i - x_j| \leq r_i$. We design a periodic five-agent system with period 2. We start with a two-agent system with $r_1 = r_2 = 2$. When the two agents are linked, they move toward each other as follows:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leftarrow \tfrac{1}{3} \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{5}$$

If agents 1 and 2 are initially positioned at -1 and 1, respectively, they oscillate around 0 with $x_i = (-1)^{i+t}\, 3^{-t}$ at time t. Now place a copy of this system with its center at $X = 4$ and a copy at $X = -4$; finally, place a fifth agent at 0 and set $r_5 = 4$. By symmetry, agent 5 never moves, while agents 1 and 2 (resp. 3 and 4) converge to a fixed point. Because agents 1 and 2 oscillate around their common fixed point (at $X = -4$), agent 5 joins to a different one at each step (same with agents 3 and 4). The agents converge but their communication graph does not. Of course, the tiniest perturbation resolves this pathology: replacing $r_5 = 4$ by $r_5 = 4 + \delta$ implies convergence of the communication graph in $O(|\log \delta|)$ steps.

To prevent chaos, randomly shifting the constraint polynomials is necessary but not sufficient: an additional perturbation rule is needed. Perturbation is also necessary for any uniform bound on the relaxation time.

3 Linearization

We show how to linearize the constraints of an influence system. We begin with the one-dimensional case $d = 1$. Let $\mathtt{d}$ be the maximum total degree of the constraint polynomials.[7] We can always assume the existence of an agent confined to position 1 with no in/out-link, which we use to homogeneize the polynomials and in the process make every monomial of degree $\mathtt{d}$. For any $1 \le k_1, \ldots, k_{\mathtt{d}} \le n$, we define the monomial $z_{k_1,\ldots,k_{\mathtt{d}}} = \prod_{i=1}^{\mathtt{d}} x_{k_i}$ and, listing them in lexicographic order, we form $\mathbf{z} = (z_{k_1,\ldots,k_{\mathtt{d}}}) \in \mathbb{R}^m$, where $m = n^{\mathtt{d}}$; note that the point $\mathbf{z}$ lies on a (real) algebraic variety $\mathcal{V}$ smoothly parametrized injectively by $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. The map $\mathbf{x} \mapsto f(\mathbf{x}) = P(\mathbf{x})\,\mathbf{x}$ induces the lifted map $\mathbf{z} \mapsto g(\mathbf{z})$, where $g(\mathbf{z}) = P(\mathbf{x})^{\otimes \mathtt{d}}\,\mathbf{z}$ 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 $\rho^{\mathtt{d}}$, where $\rho < 1$ is a uniform lower bound on the positive entries of $P(\mathbf{x})$. Its associated graph, whose edges track 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 the rows of $\mathbf{x}$. The constraint polynomials for the ground agents, being homogeneous, can be expressed as linear forms in variables taken from the coordinates of $\mathbf{z}$; for example, $x_1x_2^3 - 5x_1^2x_3^2$ becomes $z_{1,2,2,2} - 5z_{1,1,3,3}$. This produces hyperplanes which form an arrangement in $\mathbb{R}^m$. The cells c that intersect the variety $\mathcal{V}$ are thus assigned a unique matrix Q_c of the form $P(\mathbf{x})^{\otimes \mathtt{d}}$; the matrix is unique since $P(\mathbf{x})$ depends only on the sign-conditions for the constraint polynomials $\mathcal{P}_1, \ldots, \mathcal{P}_N$ evaluated at $\mathbf{x}$. The other cells can be assigned the identity matrix $\mathbf{I}_m$. Whereas a random shift produces constraint hyperplanes of the form

$$\delta + \sum_{k_1,\ldots,k_{\mathtt{d}}} a_{k_1,\ldots,k_{\mathtt{d}}}\, z_{k_1,\ldots,k_{\mathtt{d}}} = 0,$$

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 sometimes want to exploit. Given $K \subseteq \{1, \ldots, n\}$, the *cluster* C_K refers to the subset of $|K|^{\mathtt{d}}$ (lifted) 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 $(x_i - x_j)^{\mathtt{d}}$ and use the fact that the monomials almost cancel out. By the perturbation rule, therefore, the induced subgraph of the cluster cannot change until it is pulled apart by outside

[7] Not to be confused with the dimension d.

agents. We revisit this point below in greater detail. The general case $d > 1$ can be easily reduced to the one-dimensional case. We express the matrix $\mathbf{x}$ as a single vector $(x_{1,1}, \ldots, x_{1,d}, \ldots, x_{n,1}, \ldots, x_{n,d})$ and we replace $P(\mathbf{x})$ by $P(\mathbf{x}) \otimes \mathbf{I}_d$. We omit the details, which are straightforward.

4 An Algorithmic Calculus

Dynamical systems differ from standard algorithms in the way they are analyzed. Statements regarding periodicity or chaos are global in nature and require looking at *all* the inputs at once. Unlike the worst-case analysis of, say, quicksort, the search for periodicity cannot focus on a hypothetical initial state and then track its orbit in the hope of finding it intersecting itself at some point. The issue is not that the periodicity is asymptotic anyway so that perfect self-intersection might never occur: the difficulty is that two orbits might come very close only to veer off from each other much later. Theorem 1.1 states that this can indeed happen. The challenge is to show that such occurrences are rare. Unlike the execution of a standard algorithm, the orbit of a dynamical system is typically infinite. In fact, the number of combinatorially distinct orbits is itself infinite, so that any probabilistic statement about them involves infinite sums.

Individual orbits are not the place to look. Instead, one must investigate how f transforms the *whole* of phase space at once. As we saw in §2, diffusive influence systems can already be chaotic for four agents, so obviously one will need powerful tools to handle the case of a thousand agents. What we need is an algorithmic language that allows us to decompose systems recursively into subsystems while keeping track of their space-transforming features. Unfortunately, even the simplest part of Theorem 1.1, the convergence bound for bidirectional systems, requires this infrastructure, so this is where we begin. By linearization, we can assume that all constraint polynomials are of degree 1 and the dimension d is also equal to 1. By adding an extra stationary agent if necessary, we can further assume without loss of generality that the constraints are of the form

$$\mathbf{a}^T \mathbf{x} = 1.$$

The *switching partition* (*SP*) refers to the set of these hyperplanes or, equivalently, the arrangement they form in $\mathbb{R}^n$. To perturb the system, we define the *margin*

$$\mathcal{M}_\delta = \bigcup_{SP} \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T \mathbf{x} = 1 + \delta \,\}. \tag{6}$$

The shift δ is a small real number; the presence of the affine term 1 is convenient to make the perturbation additive and not multiplicative. The *SP* and the margin are combinatorially equivalent (all cells have the same sign-conditions) if we assume that $\delta > -1$. In this way, the function f, which is defined by assigning a stochastic matrix P_c to each cell of the *SP*, extends naturally to the margin since the cells are in bijection. As we shall see, the focus will be on the *atoms* of the margin, which refer to the open n-cells.

4.1 The intuition

The main difficulty of the analysis can be articulated in a few words, so a short digression might help. Figure 2 shows the margin for a two-dimensional system. Its atoms are each assigned a stochastic matrix, which together form the piecewise-linear map f. Consider the iterates of a small ball b taken inside one of the atoms: $b, f(b), f^2(b)$, etc. Suppose that it never intersects the margin (unlike in the figure). If the ball does not decrease in volume too much as it bounces around (more on this below), it must self-intersect at some point, ie, $f^t(b) \cap f^u(b) \neq \emptyset$, for some $t < u$. This is an easy consequence of the pigeonhole principle and the boundedness of the orbits. Since $f^t(b) \cup f^u(b)$ lies in the same cell, so will its iterates: this follows from the continuity of a linear map and the assumed lack of intersection between the margin and any $f^s(b)$. This shows that, for any $k = 0, 1, 2$, etc, $f^{t+k}(b)$ visits the same sequence of atoms in cyclical fashion. This in turn implies that the sequence of stochastic matrices corresponding to the iterates of b is periodic, from which it is easy to infer the evolution of the orbits from b into limit cycles.

This rosy scenario may fail for two reasons: one is that the ball b might shrink so fast as to avoid self-intersection; the other is that it might bump against the margin. In Fig.2, for example, $f^3(b)$ overlaps with the margin and splits into two (or more) smaller bodies: the map f is locally discontinuous, with a distinct stochastic matrix sending each smaller body to a different place. This phenomenon is illustrated in the figure by showing the image of an atom c being split into two and mapped to different locations. This sort of splitting could recur indefinitely, creating fragmentation obeying the exponential growth of a branching process. Although the process is deterministic, we can show that it can behave in a quasi-random fashion. This is the sign of chaos. Working against this is the fact that the splitting fragments might shrink exponentially fast, so that the total volume of all the fragments at any given time might actually decrease. If so, periodicity might be in the cards.

It follows from this discussion that the key to the dynamics resides in the shape of b and its iterates. These are ellipsoids whose dimensions can only decrease with time, if at all. Because the matrices are stochastic, the ellipsoids shrink along the axes corresponding to the nondominant eigenvectors and stay unchanged along the dominant eigenvector axes (since the eigenvalue is 1). The difficulty is that the matrices change constantly and so does their dominant eigenspace. To make sense of this, we go back to the dynamic network underpinning the influence system and design a recursive method for breaking it into subsystems and update the decomposition dynamically. This can be viewed as a grand generalization of renormalization-group techniques to highly irregular structures. Indeed, the physical analogy is deep and compelling. The tension we find here between a branching process and a shrinking "force" has its counterpart in statistical mechanics. In the Ising model, for example, a phase transition corresponds to a balanced tension between the entropy of the system (the "branching process") and its energy (the "shrinking force"). In our case, the former originates in the discontinuities of the system and the latter in its consensus-seeking nature.

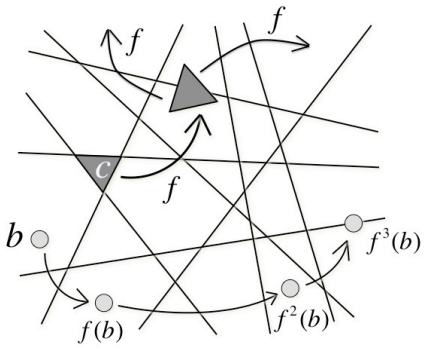

Figure 2: The atom c maps via $f = P_c$ to a cell intersecting two atoms. The third iterate of the ball b intersects the margin and splits into two or more fragments.

4.2 Phase space refinement

The open unit box $\Omega = (0,1)^n$ maps to itself via f, so obviously we may confine the phase space to it. 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 $t > 0$,

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

Our assumption $\delta > -1$ keeps the preimages of the hyperplanes of $\mathcal{M}_\delta$ via any of the linear restrictions of f either empty or of codimension 1: this 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 feasible label, then obviously so is $k-1$. The following facts follow easily from that simple but crucial observation.

LEMMA 4.1. *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 $\mathcal{S}_\nu = \mathcal{S}_\infty$.*

We define the directed graph F with one node per cell c of $\mathcal{S}_\nu$, with the edge (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 any point $\mathbf{x}$ in Ω 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 to this approach is that diffusive influence systems are rarely nesting. Here is an example: this three-agent system is initialized with $x_1 = 0$ and $x_2 = 1$; the agent 1 is fixed and the map f is defined by

$$\begin{cases} x_2 \leftarrow \frac{1}{2}(x_1 + x_2) \\ x_3 \leftarrow \frac{1}{2}(x_1 + x_3) \text{ if } x_3 - x_2 \geq 1 \text{ else } x_3 \leftarrow x_3. \end{cases}$$

The margin $\mathcal{M}_\delta$ consisting of the line $x_3 - x_2 = 1 + \delta$. When initialized at $x_3 = 1 + \delta + 2^{-t}$, for any $t > 0$, agent 3 takes time t to reach the margin. The cell decomposition $\mathcal{S}_\infty$ is infinite, and this holds for any δ, so perturbing the constraint seems to be of no help. There are two solutions: one is to thicken the margin by a tiny amount; the other, the one we choose, is to identify the "good" perturbations for a given initial state. This leads us to add δ as a variable to the system itself.

4.3 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. The root is associated with $\Omega = U_{\text{root}}$, and each child v of the root with an atom U_v. The *phase tube* (U_v, V_v) of 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.

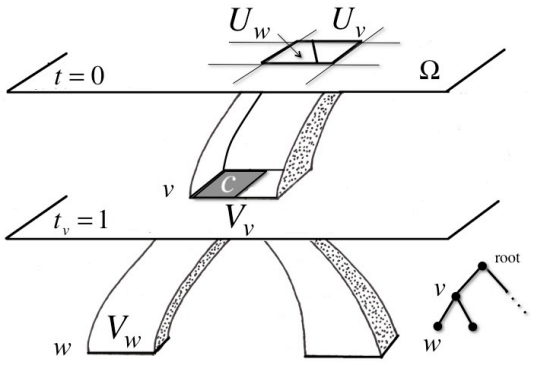

Figure 3: A phase tube (U_w, V_w) of length two: $V_w = f(c) = f^{t_w}(U_w)$.

The infinite 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.3). Whereas U_v is always an open n-cell, V_v and c can be of lower dimension. By $\delta > -1$ and simple linear algebra, 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 transition matrix of the restriction to c of the map f. The phase tube (U_v, V_v) consists of all the cylinders whose cross-sections at $t = 0, \ldots, t_v$ are, respectively, $U_v, f(U_v), \ldots, f^{t_v}(U_v) = V_v$. We note that all these iterates of f are linear over U_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 4.1); 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;[8] $\#\mathcal{S}_\nu \le 2^{h(\mathcal{T})}$.

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 $\mathbb{R}^{n+1}$.

[8] All logarithms are to the base 2.

4.4 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. All paths are infinite, so matching lengths is not an issue. The direct sum is commutative and associative; furthermore, $U_w = U_u \times U_v$, $V_w = V_u \times V_v$, and $P_w = P_u \oplus P_v$, where $w = (u, v)$ and $A \oplus B$ is the matrix obtained by placing A and B as blocks along the diagonal.

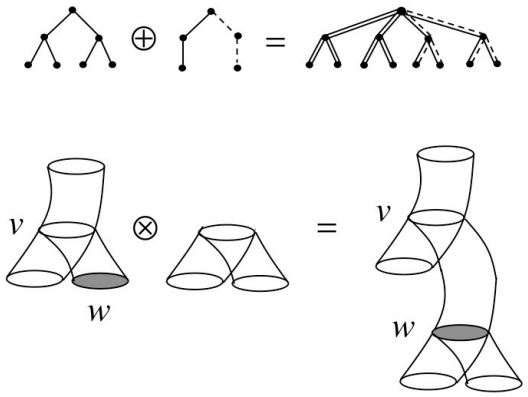

Figure 4: The two tensor operations.

Direct product. We begin with a few words of intuition. Consider two diffusive influence systems S_1 and S_2 with the same margin but possibly different maps f_1 and f_2, respectively. Let Λ denote an atom of the margin. We define the hybrid system S by requiring that a point should follow the orbit defined by f_1 until it first lands in Λ at which point it switches to f_2 and sticks to it from then on. 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, being chronological, is not commutative. It begins by identifying each node v of $\mathcal{T}_1$ such that $V_v \cap \Lambda \neq \emptyset$ and marking as *absorbed* its child w such that $V_w = f_1(V_v \cap \Lambda)$. For each such v and w, the subtree of $\mathcal{T}_1$ below w is removed and w is turned into a leaf. This operation is called *absorption* by analogy with the absorbing states of a Markov chain: any orbit reaching an absorbed leaf comes (conceptually) to a halt, broken only after we attach a copy of $\mathcal{T}_2$ to that leaf. The copy must be properly cropped: after we redefine V_w as $f_2(V_v \cap \Lambda)$, we clip $U_{\text{root}}(\mathcal{T}_2)$ to match V_w (Fig.4), which, in turn, might involve pruning $\mathcal{T}_2$.

Renormalization. Direct 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 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 $m \parallel n-m$.

4.5 The flow tracker

A little imagery might help. Suppose that $m < n$. Pour water on the $n-m$ B-agents while keeping the m A-agents dry. Whenever an edge of the (time-varying) communication graph links a dry agent to a wet one, the former agent gets wet; note how the water flows in the *reverse* direction of the edges. As soon as all the agents become wet (if ever), dry them but leave the B-agents wet; repeat forever. The case $m = n$ is identical, with one agent (any one will do) designated wet once and for all. The sequence of times at which water spreads or drying occurs plays a central role in the arborator's construction of the coding tree.

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 $m \to n-m$. Note that 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 $\mathcal{T}_{m \parallel n-m}$ but the converse is not true.[9] When the initial state $\mathbf{x}$ is understood, 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 agents.[10]

[9] The lack of edges between two subgraphs does not imply that the two subsystems are decoupled: the function $\mathcal{G}$ can depend on all the agents.

[10] All references are to the ground agents: information exchanges among the lifted agents are implied.

Flow tracker

[1] $t_0 \leftarrow 0$.

[2] Repeat:

[2.1] If $m < n$ then $W_{t_0} \leftarrow \{m+1, \ldots, n\}$ else $W_{t_0} \leftarrow \{1\}$.

[2.2] For $t = t_0, t_0+1, \ldots, \infty$

$$W_{t+1} \leftarrow W_t \cup \{\, i \mid \exists\, j \in W_t : (i,j) \in G_t \,\}.$$

[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. It is implicit in the infinite loop of step [2.2] that we break out of it as soon $|W_{t+1}| = n$, if ever; same with the repeat statement of [2], out of which "stop" breaks. 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 4.5: 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.

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

Note that in row 3 ($t = 4, 5$) we write $\mathcal{T}_{a\to bcd}$ and not $\mathcal{T}_{ad\to bc}$. Likewise, in row 5 ($t = 7, 8, 9$), we write $\mathcal{T}_{ab\to cd}$ and not $\mathcal{T}_{abd\to c}$. The reason is inductive soundness. The original system is of the form $3 \to 1$: it would be absurd to use $\mathcal{T}_{abd\to c}$ in row 5, since it is the same type $3 \to 1$ as the original system. 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 indicating the time compression rates, the 11-node path of $\mathcal{T}_{abc\to d}$ matching the graph sequence above can be expressed as

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

If we define the renormalization scale $w_k = |W_{t_k+1}| - n + m$ for $k = 1, \ldots, \ell - 1$, and let superscripts denote tree height, then any path of the coding tree can be expressed as

$$\mathcal{T}_{m\to n-m} \implies \underline{\mathcal{T}_{m\,\|\,n-m}}_{|\,t_1} \otimes \underline{\mathcal{T}_{n}}_{|1} \otimes \Big\{ \bigotimes_{k=1}^{\ell-1} \Big(\underline{\mathcal{T}_{w_k\to n-w_k}}_{|\,t_{k+1}-t_k-1} \otimes \underline{\mathcal{T}_{n}}_{|1} \Big) \Big\} \otimes \mathcal{T}_{m\to n-m}\,. \tag{7}$$

Note that the factors are optional: for example, $\mathcal{T}_{m\,\|\,n-m}$ is empty if the first step shows an edge linking an A-agent to a B-agent. The case $m = n$ (no B-agent) gets started by designating as wet any one of the A-agents. Recall that we cannot write these factors as direct sums because the communication rules for the various subgraphs are not decoupled—only the action rules are. The rewriting rule for $\mathcal{T}_{m\,\|\,n-m}$ is similar to (7), with the factors $\mathcal{T}_{w_k\to n-w_k}$ becoming

$$\mathcal{T}_{(w_k\to m-w_k)\,\|\,(w'_k\to n-m-w'_k)}.$$

Proceeding in this vein, the general factor takes on the more complicated form

$$\mathcal{T}_{(w_{k,1}\to z_{k,1})\,\|\,(w_{k,2}\to z_{k,2})\,\|\cdots\|\,(w_{k,l}\to z_{k,l})},$$

where $\sum_{i=1}^{l}(w_{k,i} + z_{k,i}) = n$. Expression (7) describes a single path of the coding tree. The coupling times t_k and renormalization scales w_k must be understood as variables since their values depend on the paths with which they are associated. The recursive derivations extend easily to the global coding trees and we need not elaborate on the details.

4.6 A linguistic analogy

Consider the infinite set of all English sentences. By sorting it alphanumerically, we can represent the set by an infinite tree, with each edge denoting a letter, number, or punctuation mark. Each path of the tree, representing a correct sentence, can be itself parsed into a tree. This induces a recursive representation of any path, similar to

the expression given in (7). All the parse trees are heavily correlated, with overlapping paths, for example, expected to contribute similar features to their respective trees. The nodes of the parse trees are naturally labeled by their linguistic functions: noun phrase, verb phrase, determiner, etc. Removing all the nodes of the parse tree below any node with a given label "renormalizes" the sentence to a coarseness level determined by the chosen label.

If we carry over this same operation to all the paths in the coding tree, we similarly renormalize the whole tree to the level of that chosen label. This enables the use of recursive analytical tools. The recursion operates in a fundamentally different way from the norm. Usually, recursion in a tree identifies a level of nodes and submits the subtrees rooted at them to a recursive treatment. Here the node identification takes place among the parse trees of the paths themselves; renormalization thus is akin to recursively applied contractions of graph minors.

5 Bidirectional Systems

We prove Theorem 1.1 for the case of bidirectional systems. Recall from §3 that d-dimensional systems of degree $\mathtt{d}$ are modeled as one-dimensional systems with linear constraints by transforming the n ground agents into $(dn)^{\mathtt{d}}$ lifted agents. For this reason, we may as well assume that $d = 1$. Running the flow tracker with respect to the ground agents and their communication graphs induces wetness among the lifted agents in the obvious way: if W_t is the set of ground agents that are wet at time t, the wet lifted agents consists of the cluster C_{W_t} of size $|W_t|^{\mathtt{d}}$: no other lifted agent is wet. The crux of the argument resides with the ground agents, so this is where we confine our discussion. We set the perturbation space to $\Delta = (0, n^{-b})$, where b is a suitably large constant (the higher b the smaller the perturbation). We need only part (i) of the perturbation rule, which consists of adding the constraints $x_i - x_j = \pm n^{-b}$ to the set $\mathcal{P}_1, \ldots, \mathcal{P}_N$ and keeping $\Phi_{i,j}(\mathbf{x})$ constant between the two hyperplanes. In other words, whenever $|x_i - x_j| \leq n^{-b}$, we need no further positioning information to infer the status of (i, j) as an edge of the communication graph. This is to keep infinitesimally close agents from changing their mode of interaction endlessly.[11]

The contraction property. It has been shown that the diameter of W_{t_k+1} is bounded by $1 - \rho^{O(k)}$ (see (14) in [19]), where ρ is the smallest nonzero entry among the ground matrices. Since the system is one-dimensional, this is merely the length of the smallest enclosing interval. It follows that water propagation to all the agents entails the shrinking of the system's diameter by at least a factor of $1 - \rho^{O(n)}$. Let $diam(s)$ be the diameter of the set of ground agents right after the s-th epoch, ie, as soon as all the agents have been made wet s times. Since an epoch witnesses the wetting of all the agents, repeated applications of this principle yields the contractivity bound:

$$diam(s) \leq e^{-s\rho^{O(n)}}. \tag{8}$$

[11] There is no relation in the two uses of the constant b; we reuse symbols to minimize their number.

It being a diffusive system, the smallest interval enclosing the agents evolves in downward inclusion. If, for large enough c, ρ^{-cn} epochs elapse, then the agents end up lying within a fixed interval of length n^{-b}. By the perturbation rule (i), the communication subgraph is now frozen and can no longer change. Consider the global coding tree with Δ as the perturbation space. We fix the initial (ground) state $\mathbf{x} \in \Omega = (0,1)^n$ once and for all. For simplicity, we let $\mathcal{T}_n^\Delta$ refer not to the whole global coding tree, but just to the "portion" corresponding to this fixed $\mathbf{x}$, with the rest of the tree pruned out of the way.[12] This implies that U_{root} is equal to the line segment $\mathbf{x} \times \Delta$, as opposed to $\Omega \times \Delta$. More generally, the set U_v and V_v are segments of the form $\mathbf{x} \times I_v$ and $f^{t_v}(\mathbf{x}) \times I_v$, respectively, where I_v is an interval in Δ.

Why renormalizing? We show why a naive approach to perturbation fails. We can conclude on the basis of (8) alone that the system is always attracted to a fixed point. As we saw in (4), however, this will not allow us to bound the convergence time uniformly. To do this requires perturbation and the use of the coding tree. Before we delve into the details, it is helpful to develop some intuition and, in particular, explain why a simple argument does not work.

We could try to argue that, by (8), the uncertainty ε about the position of the agents decreases at a certain exponential rate with time, ie, $\varepsilon(t) = C^{-t}$, for C barely above 1. (This is not quite true but let's pretend that it is for the moment.) To prevent long waits of the kind discussed in (4), one will want to exclude from Δ "risky" intervals of length commensurate with the uncertainty. If the coding tree has average degree D, this would imply an exclusion zone from Δ of a length roughly $D^t\varepsilon(t)$ at time t. Unfortunately, a node v can have a number of children polynomial in n, so we must expect D to be much larger than C: the tree branches out faster than the system contracts. As a result, $\sum_t D^t\varepsilon(t)$ does not converge and the exclusion zone is unbounded. The solution is to look more closely at the degree distribution in the tree in order to bound the size of the exclusion zone. Renormalization thus proves essential even for the bidirectional case. We rewrite (7) as

$$\mathcal{T}_n^\Delta \implies \Big\{ \bigotimes_{s=1}^{E} \bigotimes_{k=1}^{\ell_s - 1} \Big(\underline{\mathcal{T}^\Delta_{w_k \,\|\, n - w_k}}_{\,|\, t_{k+1} - t_k - 1} \otimes \underline{\mathcal{T}_n}_{\,|1} \Big) \Big\} \otimes \mathcal{T}_n^* \,, \tag{9}$$

where the outer product enumerates the epochs leading to the combinatorial "freezing" of the system. The tree $\mathcal{T}_n^*$ models a system whose communication graph is fixed. Without restriction on Δ, the number E of epochs cannot be bounded. We get around this difficulty by defining an exclusion zone for Δ, which we bound by deriving recurrence relations on the word-entropy. This leads to a relaxation time of $\rho^{-O(n^2)}|\log \varepsilon|$. Rather than going over the technical details of this result, we describe a slight variant of the proof based on the s-energy which allows us to derive a better bound.

[12] The full coding tree will be needed for the nonbidirectional case.

The renormalization equations. We call a node v of $\mathcal{T}_n^\Delta$ *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 greater than λ is called the *communication count* C_λ: we have shown elsewhere, by appealing to a certain generating function called the *s-energy* [19], that $C_\lambda \leq \lambda^{-1}\rho^{-O(n)}$. It follows that, along any path of $\mathcal{T}_n^\Delta$, the number of heavy nodes is $\rho^{-O(n)}$. Light nodes have an interesting structure, which we explore next.

Consider the agents at any light node of $\mathcal{T}_n^\Delta$. The connected components of their communication graph have diameter at most $n^{1-2b} < n^{-b}$. By convexity, the smallest enclosing interval of any component encloses the one at the next step; therefore, by the perturbation rule, the induced subgraph of the component remains unchanged. This also shows that agents move by less than n^{1-2b} in one step. If a new edge (i, j) appears, by the perturbation rule, the two agents must have been more than n^{-b} apart during at least one of these two steps. Whichever one it is, the agents are at a distance greater than $n^{-b} - 2n^{1-2b} > n^{-2b}$ in the second step, so the newly appearing edge makes the corresponding node of the coding tree heavy. In other words, along downward paths of light nodes, the communication graph stays the same and so does the matrix $P = P_1 \oplus \cdots \oplus P_r$ corresponding to f: the n agents appear in clusters formed by their r connected components, each one enclosed in an interval of length less than n^{-b}. These intervals evolve in descending nested fashion (downward inclusion). By (8) and basic Markov chain theory, there exists a matrix $\Pi = \Pi(P)$ whose rank is the number of summands in $\oplus_i P_i$, such that, for any $k \geq 0$,

$$\|P^k - \Pi\|_{\max} = e^{-k\rho^{-O(n)}}. \tag{10}$$

Fix a small positive ε_0. A light node is called *ethereal* if each one of its connected components has diameter at most ε_0. By (10) and our previous discussion, in any downward path of light nodes, all the non-ethereal light nodes appear in a prefix of length $\rho^{-O(n)}|\log \varepsilon_0|$. Let $\mathcal{H}^\Delta$, $\mathcal{E}^\Delta$, and $\mathcal{L}^\Delta$ denote any maximal subtree of $\mathcal{T}_n^\Delta$ consisting of heavy, ethereal, and non-ethereal light nodes, respectively. It follows that

$$\mathcal{T}_n^\Delta \implies \bigotimes \left\{\mathcal{H}^\Delta, \mathcal{L}^\Delta, \mathcal{E}^\Delta\right\}. \tag{11}$$

The analysis rests on three simple observations:

[1] The height of any subtree $\mathcal{H}^\Delta$ and the number of times it appears in a direct product (11) are both bounded by $\rho^{-O(n)}$.

[2] The height of any subtree $\mathcal{L}^\Delta$ is at most $\rho^{-O(n)}|\log \varepsilon_0|$. The intervals I_v corresponding to a single level of the tree are disjoint; because the map f is the same linear function across the level, the number of intervals can vary by at most $n^{O(1)}$ between levels. This proves that the size of any $\mathcal{L}^\Delta$ is itself bounded by $\rho^{-O(n)}|\log \varepsilon_0|$.

[3] Given any node v of any $\mathcal{E}^\Delta$, V_v is a segment $f^{t_v}(\mathbf{x}) \times I_v$, where I_v is an interval in Δ. Furthermore, there is an n-cube of side-length ε_0 that contains the points $f^{t_v}(\mathbf{x})$

for all $v \in \mathcal{E}^{\Delta}$. This shows that in any of the margin equations (6), $\mathbf{a}^T\mathbf{x} = 1 + \delta$, the left-hand side can vary by at most $\varepsilon_0 n^{O(1)}$. As a result, the intervals I_v only differ at the ends over a length of the same bound; so, one can remove from Δ some intervals of total length $\varepsilon_0 n^{O(1)}$ to make I_v constant for all v in $\mathcal{E}^{\Delta}$. We carry out this removal for all trees of ethereal nodes and denote by $\underline{\Delta}$ what is left from Δ after all the interval removals. All of the trees $\mathcal{E}^{\underline{\Delta}}$ are now path-like and absorption-free, so we can rewrite (11) as

$$\mathcal{T}_n^{\underline{\Delta}} \Longrightarrow \bigotimes \{\mathcal{H}^{\underline{\Delta}}, \mathcal{L}^{\underline{\Delta}}\} \otimes \mathcal{E}^{\underline{\Delta}}.$$

From observations [1,2], it follows readily that the nesting time for $\mathcal{T}_n^{\underline{\Delta}}$ is at most $\rho^{-O(n)}|\log \varepsilon_0|$. By (10), the system evolves within distance ε of its fixed-point attractor after $\rho^{-O(n)}|\log \varepsilon|$ steps, for any $\varepsilon < \varepsilon_0$, which establishes the desired convergence bound. We now show that $\underline{\Delta}$ differs from Δ by an arbitrarily small amount. The trick is to renormalize the non-ethereal light nodes by contracting their trees into single nodes, as in

$$\mathcal{T}_n^{\underline{\Delta}} \Longrightarrow \bigotimes_{k=1}^{\rho^{-O(n)}} \{\, \mathcal{H}^{\underline{\Delta}}, \underline{\mathcal{L}}^{\underline{\Delta}} \,\} \otimes \mathcal{E}^{\underline{\Delta}}.$$

By observation [2], the contracted tree $\underline{\mathcal{L}}^{\underline{\Delta}}$ is a node of degree $\rho^{-O(n)}|\log \varepsilon_0|$. Recall that the word-entropy is the logarithm of the number of shallow nodes. From the inequalities

$$\begin{cases} h(\mathcal{T} \otimes \mathcal{T}') \le h(\mathcal{T}) + h(\mathcal{T}') + \log \texttt{max-degree}\,(\mathcal{T}) \\ h(\mathcal{T}) \le \nu(\mathcal{T}) \times \log \texttt{max-degree}\,(\mathcal{T}), \end{cases} \tag{12}$$

it is immediate that

$$h(\mathcal{T}_n^{\underline{\Delta}}) \le \rho^{-O(n)} \big(h(\mathcal{H}^{\underline{\Delta}}) + \log(\rho^{-O(n)}|\log \varepsilon_0|) \big) \le \rho^{-O(n)} \log|\log \varepsilon_0|.$$

We can now bound the Lebesgue measure of the exclusion zone of "bad" perturbations:

$$\text{Leb}\,(\Delta \setminus \underline{\Delta}\,) \le \varepsilon_0 n^{O(1)}\, 2^{\,h(\mathcal{T}^{\underline{\Delta}})} \le \varepsilon_0 |\log \varepsilon_0|^{\rho^{-O(n)}} < \sqrt{\varepsilon_0},$$

for ε_0 small enough, which allows us to set a probability of success as high as we want. This concludes the proof of the bidirectional case of Theorem 1.1, as we have proven that, given any initial state $\mathbf{x}$ in the unit n-cube and a random perturbation δ of the constraints, with probability arbitrarily close to 1, the system evolves to within distance ε of its fixed-point attractor in $\rho^{-O(n)}|\log \varepsilon|$ steps, for any positive ε small enough. □

6 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

persistent system and show later how to reduce any influence system to the persistent case. Let t_o be the timing threshold of the perturbation rule (ii) and let H be a directed n-node graph.[13] 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, the system comes to a halt. 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, with high probability over the margin's perturbation, either the orbit of any point is attracted to a limit cycle or its path in the coding tree reaches an absorbed leaf. To relate this special system to the original one, we should think of H as our guess for the *persistent graph*, which is defined as the graph consisting of the edges that appear infinitely often in the communication graph.

Consider the directed graph derived from H by contracting every strongly connected component into a single node. Let $B_1, \ldots, B_r$ be the strongly connected components of H whose corresponding nodes in the contracted graph are sinks (ie, with no outgoing edges) 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 $\mathcal{T}_{m\to n-m}$. Because no communication between two A-agents can transit through a B-agent (and conversely),

$$P_{\leq v} = \begin{pmatrix} A_{\leq v} & C_v \\ \mathbf{0} & B_{\leq v} \end{pmatrix}. \tag{13}$$

We break down the bifurcation analysis in four stages: in §6.1 we bound the rate at which phase tubes thin out; in §6.2 we argue that, deep enough in the coding tree, perturbations keep the expected (mean) degree below one; in §6.3, we show how perturbed phase tubes avoid being split by the margin at high depths; finally, in §6.4, we show to reduce any influence system to the "persistent" case. Our assumption that ρ is a positive rational encoded over $O(\log n)$ bits is not necessary but, by implying that $\rho > n^{-O(1)}$, it simplifies the calculations and the notation.

6.1 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 6.1. *Given a node v of $\mathcal{T}_{m\to n-m}$, there exist vectors $\mathbf{z}_i \in \mathbb{R}^{n_i}$ $(i = 1, \ldots, r)$, such that, for any $t_v \geq t_c := n^{cnt_o}$ and a large enough constant c,*

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

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

[13] We can pick t_o as large as we want.

Proof. We note the faster convergence rate in (ii). Beginning with (i), we 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. (Note that the path need not track the orbit of $\mathbf{x}$.) Referring to the arborator (7), let's analyze the factor

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

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}, which consists of the wet agents at time $t_k + 1$, lie in $(0, 1 - \sigma]$. Because their distance to 1 can decrease by at most a polynomial factor at each step, they all lie in $(0, 1 - \sigma n^{-O(t_o)}]$ between times t_k and t_{k+1}. The agents newly wet at time $t_{k+1} + 1$, ie, those in $W_{t_{k+1}+1} \setminus W_{t_{k+1}}$, move to a weighted average of up to n numbers in $(0, 1)$, at least one of which is in $(0, 1 - \sigma n^{-O(t_o)}]$. This implies that the agents of $W_{t_{k+1}+1}$ lie in $(0, 1 - \sigma n^{-O(t_o)}]$. Since $\sigma \leq 1$, when all the A-agents are wet, which happens within nt_o steps, their positions are confined within $(0, 1 - n^{-O(nt_o)}]$. Every nt_0 steps, the agents thus find themselves confined to an increasingly smaller interval; hence,

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

which proves (i). We establish (ii) along similar lines. The 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 (i), as stated in §5, 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 $H_{|B}$. The margin is defined by a polynomial number of hyperplanes in an n-dimensional phase space, so any node of the coding tree has only $n^{O(n)}$ children. 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 $B_{\leq v} = Q^{t_v - t_c} B_{\leq u_i}$ 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 $e^{-kn^{-O(n)}}$. (Note that we save 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

$$\|Q^k - \mathbf{1}_{n_1}\mathbf{z}^T\|_{\max} \leq e^{-kn^{-O(n)}}.$$

Since $t_v = k + t_c$, it suffices to double the constant c to derive part (ii) of the lemma and complete the proof. □

The lemma points to C_v as the key to the dynamics and the necessary focus of our attention. We restate the previous lemma (in slightly weaker form) in terms of the nodes u_i at which the communication graph among the B-agents freezes. This freezing is immune to the choice of perturbation so it applies to the global coding tree as well. We enlarge the perturbation space to $\mathbb{I} = (-1, 1)$.

LEMMA 6.2. *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*

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

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

6.2 Sparse branching

Bruin and Deane [10] 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 inspired by theirs but the lack of contractivity of stochastic matrices creates serious difficulties: this conceptual obstacle, in fact, is the reason for the use of renormalization. The main source of complication is easy to grasp. Recall from Figure 2 that the ball b is mapped to an ellipsoid whose possible intersection with the margin causes branching in the coding tree. Whereas in the case of contractions, the ball shrinks at each step, the use of stochastic matrices entails contractions only along certain privileged directions, which themselves might change with time. Along the principal eigenvectors, distances remain unchanged. Analyzing the branching structure involves keeping track of the geometry of the modes, which is much more challenging and, in fact, the main reason for using renormalization.

Let $\operatorname{Lin}[x_1, \ldots, x_n]$ denote any real linear form over $x_1, \ldots, x_n$, with $\operatorname{Aff}[x_1, \ldots, x_n]$ designating the affine version; in neither case may the coefficients depend on δ or on the agent positions.[14] With $y_1, \ldots, y_r$ understood, a *gap of width* $\omega \in \mathbb{R}$ denotes an interval of the form $a + \omega\,\mathbb{I}$, where $a = \operatorname{Aff}[y_1, \ldots, y_r]$. We define the set

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

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 $P_0, \ldots, P_{t_v}$

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

form what we call a *valid matrix sequence*. Fix a parameter $\varrho > 0$ (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 n-cube $\mathbf{B} = \mathbf{x} + \varrho\,\mathbb{I}^n$ 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 $(P_k \cdots P_0\,\mathbf{B}) \setminus \mathcal{M}_\delta$ is disconnected. The following lemma is the key to sparse branching. It states that suitably chosen phase tubes split rarely and independently of $\mathbf{x}$; only $(y_1, \ldots, y_r)$ needs to be fixed. The uniformity over $\mathbf{x}$ is crucial. In the following, recall that $\gamma = 1/t_c = n^{-cnt_o}$.

LEMMA 6.3. *Fix $\varrho > 0$, $D_0 \geq 2^{(1/\gamma)^{n+1}}$, 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 width $\varrho n^{O(n^5 D_0)}$ 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 $\mathbf{x} + \varrho\,\mathbb{I}^n$ 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. We begin with a technical lemma. For an integer $D > 0$ and $k = 0, \ldots, D$, 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 $N > n$. (No relation with the number N of constraint polynomials.) 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 and achieve the uniformity claimed above. This is the key condition for eliminating $\mathbf{x}$ from the phase tube splitting equations.

LEMMA 6.4. *Given any integer $D \geq 2^{(1/\beta)^{m+1}}$ and any $I \subseteq \{0, \ldots, D\}$ 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 $u \in \mathbb{R}^{|I|}$ such that*

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

The lemma implies that $\mathbf{1}_{|I|}$ is not in the column space of $V_{|I}$. Why should we care? Branching in the tree reflects the splitting of the phase tube, which in turn signifies that the orbit comes crashing into the margin. We can express this by a linear equation: $\sum_i c_i x_i + \sum_j c'_j y_j = 1 + \delta$. In this way, the many splittings present along any given path of the global coding tree can be expressed by a single linear system of the form

$$M\mathbf{x} + M'\mathbf{y} = (1 + \delta)\mathbf{1}. \tag{14}$$

To eliminate $\mathbf{x}$ from the system above, we need to find a vector u such that $u^T M = \mathbf{0}$, so that we are left with conditions on the perturbation alone: $u^T M' \mathbf{y} = (1 + \delta) u^T \mathbf{1}$. Since M is typically much higher than it is wide, such a vector is certain to exist. The danger is that they might all be orthogonal to $\mathbf{1}$, in which case premultiplying each side of (14)

by u yields $u^T M'\mathbf{y} = 0$, which renders perturbation ineffectual. For this to happen, the matrix M must be heavily constrained: geometrically, its rows must all lie in the same affine subspace. To show this is impossible, we must rely on the structure of M, about which unfortunately we know rather little: the rows decay quickly and the number of bits does not grow too fast. Although roughly accurate, this explanation fails to capture a number of subtleties. For example, the collisions into the margin need not correspond to the same $\mathbf{x}$ and the same δ, so we must add slack variables to (14).

Proof. We give a rough sketch of the argument here and refer the reader to the Appendix for the full proof. Suppose that $m = 3$ and I is very large. In this way the rows of $V_{|I}$ form a long sequence of points $p_1, \ldots, p_{|I|}$ in $\mathbb{R}^3$. Because $\alpha < 1$, the points get close to the origin exponentially fast. Assume that p_1 and p_2 are distinct. If the line L through p_1, p_2 passes through the origin, then the lemma follows easily: L is parametrized as $p = u_1 p_1 + u_2 p_2$, with $u_1 + u_2 = 1$, so that setting p to the origin provides the desired vector u consisting of u_1, u_2 with the other coordinates set to 0. Otherwise, enumerate the points $p_3, p_4, \ldots$ until we find one, p_i, that is not on the line. The key observation is that we do not need to wait long for this to happen: indeed, the line L, whose coefficients require only "few" bits (determined by p_1 and p_2), cannot approach the origin too closely. This is the gap argument we exploit to infer the existence of p_i. Next, we consider the plane π formed by p_1, p_2, p_i and note that, for the same reasons discussed earlier, the proof is complete if π passes through the origin. If it does not then the distance from the plane to the origin is bounded from below by a function of the number of bits required to encode p_1, p_2, p_i. Although this number grows with i, there must be some j such that p_j is too close to the origin to be on the plane π. The proof iterates in this fashion until the affine subspace spanned by the current prefix of points p_k contains the origin, at which point the vector u of the lemma can be identified. See the complete proof in the Appendix. □

Although unrelated, we can pick the same constant c as the one used in Lemma 6.1. Since $\alpha \geq N^{-O(1)}$, β can be assumed much less than 1. To prove Lemma 6.3, we first consider the case where the splitting nodes are well separated along their path, which allows for Lemma 6.1 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 $0 = s_0 < \cdots < s_D \leq D_0$ such that

$$D \geq 2^{(1/\beta)^{m+1}} \qquad \text{and} \qquad 1/\gamma \leq s_k - s_{k-1} \leq 3/\gamma, \tag{15}$$

for $k = 1, \ldots, D$: we identify the matrix A_k of Lemma 6.4 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 (13), $A_k = A_{\leq w}$, for some node w (not necessarily an ancestor of v) of depth $t_w = s_k - s_{k-1} \geq 1/\gamma$.[15] Thus, by Lemma 6.1, for $k > 0$, the maximum row-sum of any A_k satisfies

$$\alpha \leq \frac{1}{e}\,. \tag{16}$$

[15] This relies on the Markovian nature of the dynamics: any subsequence of a valid matrix sequence is itself a prefix of a valid matrix sequence.

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.[16] Fix $\delta \in \mathbb{I}$ and pick I in Lemma 6.4 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 $\mathbf{x} + \varrho\,\mathbb{I}^n$ and $P_{s_0}, \ldots, P_{s_k}$ splits at every node indexed by s_k ($k \in I$) along the chosen hyperplane (Fig.5). In other words, for each $k \in I$, there exist a node z_k of depth $t_{z_k} = s_k$ and $\varrho_i = \varrho_i(k)$, for $i = 1, \ldots, n$, such that $|\varrho_i| < \varrho$ and

$$(a_k, b_k) \begin{pmatrix} A_{\leq z_k} & C_{z_k} \\ \mathbf{0} & B_{\leq z_k} \end{pmatrix} (x_1 + \varrho_1, \ldots, x_n + \varrho_n)^T = 1 + \delta,$$

where the selected margin 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 $v_k = a_k A_{\leq z_k}$ (by definition) and $\mathbf{x} \in \mathbb{C}[y_1, \ldots, y_r]$ (by assumption), it follows that

$$v_k(x_1 + \varrho_1, \ldots, x_m + \varrho_m)^T + \mathrm{Lin}\,[\, y_1, \ldots, y_r, \varrho_{m+1}, \ldots, \varrho_n \,] = 1 + \delta\,, \qquad (17)$$

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

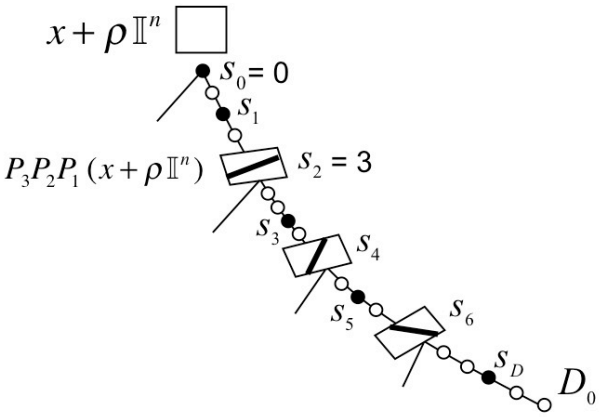

Figure 5: The phase tube splits at the nodes indexed by $I = \{2, 4, 6\}$. The nodes of depth s_k for $k \notin I$ are represented as black dots: s_0, s_1, s_3, s_5, s_7 ($D = 7$). The other nodes in the paths are the white dots.

Lemma 6.4 allows us to eliminate the variables $x_1, \ldots, x_m$: we premultiply $V_{|I}$ by

[16] With $m = 3$, $x_1 - x_3 = 1 + \delta$ gives $a_k = (1, 0, -1)$ and $2x_2 - x_4 = 1 + \delta$ produces $a_k = (0, 2, 0)$.

the unit vector u to find that

$$\sum_{i=1}^{|I|} u_i \operatorname{Lin}_i [\, y_1, \ldots, y_r, \varrho_{m+1}, \ldots, \varrho_n \,] = (1+\delta) u^T \mathbf{1}_{|I|} \;\geq\; (1+\delta) N^{-cm^3 D}.$$

Since the coefficients of Lin_i are $n^{O(1)}$, $\|u\|_2 = 1$, $|\varrho_i| < \varrho$, it follows that

$$\operatorname{Lin} [\, y_1, \ldots, y_r \,] + \chi = (1+\delta)\xi,$$

where $|\chi| \leq \varrho D n^{O(1)}$, $\xi \geq N^{-cm^3 D}$, and the coefficients of the linear form are bounded by $Dn^{O(1)}$; hence,

$$|\, \delta + \operatorname{Aff} [\, y_1, \ldots, y_r \,] \,| \leq \varrho N^{O(cm^3 D)}, \tag{18}$$

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 does not vanish during the elimination. Thus, as long as it remains outside a gap of width $\varrho N^{O(cm^3 D)}$, the phase tube formed by $\mathbf{x} + \varrho\, \mathbb{I}^n$ 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. But, in the global coding tree, the splitting can occur for different values of δ. We note, however, that each δ in (17) can be replaced by $\delta + \nu_k$ ($k \in I$), for $|\nu_k| \leq \varrho$, and the new system of inequalities will still imply (18); this will be crucial for the randomization. We summarize our results, using the bound derived from (16): $|\log \alpha| \geq \log e > 1$.

Lemma 6.5. *Let $N = n^{n^2/\gamma}$ and $\beta = 1/(cm^3 \log N)$, where c is the constant used in Lemma 6.4. Fix $(y_1, \ldots, y_r) \in \mathbb{R}^r$ and a path in $\mathcal{T}^{\mathbb{I}}_{m \to n-m}$ from the root. Pick $D+1$ nodes of the path of depth $0 = s_0 < \cdots < s_D$ satisfying (15); 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 width $\varrho N^{O(cm^3 D)}$, 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 $\mathbf{x} + \varrho\, \mathbb{I}^n$ cannot split at all the nodes of I in $\mathcal{T}^{\Delta}_{m \to n-m}$ (assuming they exist).*[17]

To prove Lemma 6.3, we need to extend the previous lemma to all the paths of the coding tree of the prescribed length and remove from (15) the lower bound of $1/\gamma$ on the distance between consecutive splitting nodes. Fix $D_0 \geq 2^{(1/\gamma)^{n+1}}$, 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 $\mathbf{x} + \varrho\, \mathbb{I}^n$ and $P_0, \ldots, P_{D_0}$ splits (with respect to $\mathcal{T}^{\mathbb{I}}_{m \to n-m}$); assume that

$$|K| \geq D_0^{1-\gamma^{n+1}}. \tag{19}$$

[17] The issue is that a path in $\mathcal{T}^{\mathbb{I}}_{m \to n-m}$ might no longer exist once we replace $\mathbb{I}$ by the small interval Δ.

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, $|I| \geq \lfloor \gamma |K| \rfloor - 1$. 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 element into L, hence $|J|/2 \leq |L| \leq \gamma D_0 + 1$. 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 6.1 and N, β in Lemma 6.5, we can verify that

$$\text{(i)}\ \ D_0 \geq 2^{(1/\gamma)^{n+1}} \geq \gamma^{-1} 2^{(1/\beta)^{m+1}+1} \quad \text{and} \quad \text{(ii)}\ \ D_0^{1-\gamma^{n+1}} \geq \tfrac{2}{\gamma}(\gamma D_0 + 1)^{1-\beta^{m+1}}. \quad (20)$$

Part (i) ensures (15). By Lemma 6.5, keeping δ outside the union W of at most $n^{O(|I|)}$ gaps of width $\varrho N^{O(m^3 D)}$ 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 $|I| \geq D^{1-\beta^{m+1}}$, which follows from: (19); $|I| \geq \lfloor \gamma |K| \rfloor - 1$; $D = |L| \leq \gamma D_0 + 1$; and part (ii) of (20).

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 $\mathbf{x} + \varrho\, \mathbb{I}^n$, 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 6.3, 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 $t_v = D_0$. □

6.3 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 (i), 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.[18] By the proof of Lemma 6.1, 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

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

[18] We return to the rule used in §5 for convenience; we could use an arbitrarily small threshold instead.

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 $\mathbf{y}$.[19] This follows easily, as does the next lemma, from the stochasticity of Q and the identities: $\widetilde{Q}Q = Q\widetilde{Q} = \widetilde{Q}^2 = \widetilde{Q}$ and $\widetilde{Q}\mathbf{y} = \mathbf{y}$.

LEMMA 6.6. *The set Υ is forward-invariant. Furthermore, any $\bar{\xi} \in \mathbf{y} + n^{-2b}\,\mathbb{I}^{n-m}$ belongs to Υ_B if and only if $\widetilde{Q}\bar{\xi} = \mathbf{y}$.*

Proof. If $\mathbf{x} \in \Upsilon$, the diameter of any group B_i is below n^{-b} so the communication graph induced by their agents remains forever frozen. The B-agents are attracted to the fixed point $\mathbf{y}$. Indeed, we note that, by stochasticity, $Q\mathbf{y} = \widetilde{Q}\mathbf{y} = \mathbf{y}$. To prove that Υ is forward-invariant, it then suffices to show (by convexity) that if $\bar{\xi}$ lies in Υ_B then so does $Q\bar{\xi}$. We can write $\bar{\xi}$ as $\mathbf{y} + w$, where $w \in \ker \widetilde{Q}$ and $\|w\|_\infty \leq n^{-2b}$. Because Q is stochastic, this implies that

$$\|Q\bar{\xi} - \mathbf{y}\|_\infty = \|Qw\|_\infty \leq \|w\|_\infty \leq n^{-2b},$$

hence $Q\bar{\xi} \in \mathbf{y} + n^{-2b}\,\mathbb{I}^{n-m}$. It now suffices to show that $Q\bar{\xi} \in \mathbf{y} + \ker \widetilde{Q}$, which follows from the fact that $\widetilde{Q}Q = \widetilde{Q}$, hence $\widetilde{Q}Qw = \widetilde{Q}w = \mathbf{0}$. The second part of the lemma follows from two simple observations. First, if $w = \bar{\xi} - \mathbf{y} \in \ker \widetilde{Q}$, then $Q^k w = \widetilde{Q}w + (Q^k - \widetilde{Q})w = (Q^k - \widetilde{Q})w \to \mathbf{0}$, as $k \to \infty$; hence $Q^k\bar{\xi} \to \mathbf{y}$. Second, if $\bar{\xi}$ is arbitrary but $Q^k\bar{\xi} \to \mathbf{y}$, then $w = \bar{\xi} - \mathbf{y}$ satisfies $Q^k w \to \mathbf{0}$, hence $w \in \ker \widetilde{Q}$. □

We set ϱ, D_0 as in Lemma 6.3 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 an extra, smaller one if need be). 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:

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

[19] 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. This is because the communication function $\mathcal{G}$ is global.

Fixing the B-agent attractor. With $\mathbf{y} = \widetilde{Q}\xi$ 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. (Note how Lemma 6.6 allows us this choice of phase space.) For any node v of depth $t_v \geq t_c$, the limit matrix D_u in Lemma 6.2 is the same for all nodes u of depth t_c. Indeed,

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

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 $\mathbf{x} \in U_v \subseteq \Upsilon$, the orbit from $\mathbf{x}$ leads to w (resp. v) at time t_w (resp. t_v) and

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

By the Markovian nature of the system, 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.

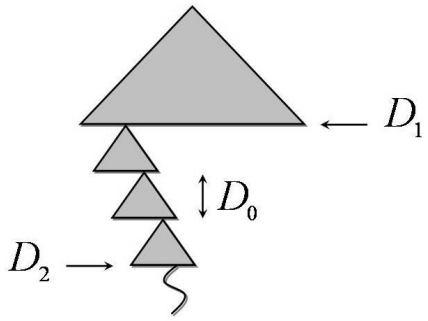

Figure 6: The global coding tree is stratified into three layers, with decreasing branching rates.

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 6.3 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. We enlarge D_0 slightly from its lower bound of Lemma 6.3.

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

Second layer. Technically, Lemma 6.3 addresses only the splitting of a phase tube originating from the root, 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.[20] By (23), for any node v of depth $t_v \geq D_1$, there exists $\mathbf{x}(v) \in \mathbb{C}[y_1, \ldots, y_r]$ such that $V_v \subseteq \mathbf{x}(v) + \varrho\,\mathbb{I}^n$, provided that

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

Note that the requirement in (23) that $t_v \geq 3t_c = 3/\gamma$ is implied by $t_v \geq D_1$. The inclusion $V_v \subseteq \mathbf{x}(v) + \varrho\,\mathbb{I}^n$ implies that Lemma 6.3 holds with respect to the subtree of $\mathcal{T}^{\Delta|\Upsilon}_{m\to n-m}$ rooted at any node v of depth $t_v \geq D_1$. We apply it to every D_0-th level in the tree below depth D_1 and find that the number of nodes in $\mathcal{T}^{\Delta|\Upsilon}_{m\to n-m}$ of depth at most $t \geq D_1$ is bounded by

[20] Factoring out the B-agents gives us the sort of fixed-point attraction that is required by Lemma 6.3: it is a dimension reduction device in attractor space.

$$\underbrace{n^{O(nD_1)}}_{\texttt{depth D}_1} \times \underbrace{n^{O(nD_0^{1-\gamma^{n+1}}\lfloor (t-D_1)/D_0 \rfloor)}}_{\texttt{from D}_1 \texttt{ to t in chunks of D}_0} \times \underbrace{n^{O(nD_0)}}_{\texttt{truncated chunk}} \times \underbrace{D_0}_{\texttt{single paths}} .$$

To see why, treat each path of single-child nodes as a single edge (hence the factor of D_0). Thus, for any $t \geq D_1$,

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

The strength of this bound is that the exclusion zone W needed to make it true can be inferred from a single subtree of height D_0: this is why the uniformity over $\mathbf{x}$ in Lemma 6.3 is so crucial. The whole argument would, indeed, collapse if we had to include a different exclusion zone for each such subtree.

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 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 branching at depth D_2, provided that $\Delta' = \Delta \setminus W'$, where W' consists of gaps of width $n^{O(1)}e^{-\gamma D_2/3}$ 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 will set 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 $\kappa = \kappa(b, c)$; recall that $\gamma = n^{-cnt_o}$. We set the parameters $\varrho = n^{-\kappa n^5 D_0}$, where, rounding up to the nearest integer,

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

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) becomes $\mathrm{Leb}\,(W') \leq \varrho\, 2^{-D_0}$. Since, as a free canonical interval, Δ has length ϱ, we find 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, the nesting time is at most D_2 and asymptotic periodicity ensues (details below). 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 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

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

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 this upper bound on the word-entropy:

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

Below depth D_2, 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[21]

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

Asymptotic periodicity. We fill in some of the details of the periodicity claim made earlier. We show that the limit cycles converge exponentially fast with a characteristic time proportional to the period. By Lemma 4.1 (and the discussion that followed), we know that following a path of the tree below depth D_2 will produce a periodic sequence of transition matrices (infinite or stopping at an absorbed node). For any $t_v > D_2$ with a fixed residue modulo the period, we have $P_{\le v} = Q^s Q_0$, where s is roughly $t_v - D_2$ divided by the period. Taking multiples of the period if need be, we can ensure that Q covers at least c/γ steps, so that writing

$$Q = \begin{pmatrix} A & C \\ \mathbf{0} & B \end{pmatrix}, \tag{33}$$

we can adapt Lemma 6.1 to see to it that $\|A\,\mathbf{1}\|_\infty \le \frac{1}{2}$ and $\|B^s - \widetilde{B}\|_{\max} \le 2^{-s}$. Expressing Q^s as

$$\begin{pmatrix} A^s & C_s \\ \mathbf{0} & B^s \end{pmatrix},$$

we use standard properties of a Markov chain's fundamental matrix to derive

$$C_s - (I - A^s)(I - A)^{-1} C\widetilde{B} = \sum_{k=0}^{s-1} A^{s-k-1} C B^k - \sum_{k=0}^{s-1} A^k C \widetilde{B} = \sum_{k=0}^{s-1} A^{s-k-1} C D_k,$$

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

for $s > 0$, where $D_k = B^k - \widetilde{B}$ and $\sum_{k\geq 0} A^k = (I-A)^{-1}$. Since C is substochastic, $\|CD_k\|_{\max} \leq \|D_k\|_{\max} \leq 2^{-k}$. From $\|A^k \mathbf{1}\|_\infty \leq 2^{-k}$, we find that

$$\|A^{s-k-1} C D_k\|_{\max} \leq 2^{1-s};$$

hence

$$\|C_s - (I-A^s)(I-A)^{-1} C\widetilde{B}\|_{\max} \leq s 2^{1-s}.$$

It follows that, if we write

$$\widetilde{Q} = \begin{pmatrix} \mathbf{0} & (I-A)^{-1}C\widetilde{B} \\ \mathbf{0} & \widetilde{B} \end{pmatrix},$$

then, by $\|A^s\|_{\max} \leq 2^{-s}$ and $\|(I-A)^{-1}\|_{\max} \leq 2$, we find the approximation bound

$$\|Q^s - \widetilde{Q}\|_{\max} = O(sn2^{-s}).$$

This implies an exponential convergence rate with a characteristic time proportional to the period.

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 6.2, 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 $\widetilde{Q}$ is not the same as D_u) and, by Lemma 6.6, $\xi \in \Upsilon_B$ 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

$$\mathrm{Prob}_\delta\,[\,\text{failure in } \mathcal{T}_{m\to n-m}^{n^{-b}\mathbb{I}}\,] \leq 2^{-D_0/2}. \tag{34}$$

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),

$$\begin{cases} \nu(\mathcal{T}_{m\to n-m}^{\mathbb{I}'_s}) \leq D_2 + D_3 \\ h(\mathcal{T}_{m\to n-m}^{\mathbb{I}'_s}) \leq h(\mathcal{T}_{m\to n-m}^{\mathbb{I}'_s \,|\, \Upsilon}) + O(D_3\, n \log n) \leq \gamma^{-1} n^{O(1)} D_0\,. \end{cases} \tag{35}$$

6.4 Removing persistence

The perturbation rule (ii) stipulates that any edge that fails to appear in the communication graph during t_o consecutive steps is to remain permanently absent.[22] This is modeled by an absorption in the coding tree. Since we do not know the persistent 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,[23]

$$\mathcal{T} \Longrightarrow \bigotimes_{k=1}^{k_0} \mathcal{T}_{m_k \to n-m_k}, \tag{36}$$

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 $\phi_0 = 0$. By (34, 35) and a union bound, for $k > 0$,

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

for some large enough 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. □

6.5 Arbitrary dimension and algebraic degree

The analysis readily extends to any dimension d and degree $\mathtt{d}$ via the tensor lifting construction of §3. Recall that the two essential ingredients are: (i) the thinning rate

[22] Note that the margin does not change. Only its labeling does: having assumed that any row of a transition matrix is uniquely determined by its distribution of positive entries, the loss of an edge entails an unambiguous updating of the corresponding row.

[23] Of course, we cannot require that the actual influence systems should reset to 0 the timeout counts of their edges at each direct product in (36). There is no need to do that anyway: not resetting them can only mean earlier absorptions (ie, smaller t_o) so the analysis can ignore this point and assume resetting.

(Lemma 6.1); and (ii) sparse branching (Lemma 6.3). We easily check that both conditions still hold (though with different rates). The matrix Q_c associated with a cell c of the margin is of the form $(P \otimes \mathbf{I}_d)^{\otimes \mathtt{d}}$, with $P = P(\mathbf{x})$ whenever $\mathbf{z} = \mathbf{z}(\mathbf{x})$ for some $\mathbf{z}$ on the variety $\mathcal{V}$. We retain the coding tree type $\mathcal{T}_{m \to n-m}$ over the m (resp. $n-m$) ground A-agents (resp. B-agents). This induces a block-directional decomposition of the lifted system. Recall that, given $\mathbf{z} \in \mathcal{V}$, the (lifted) agent l is at position $\mathbf{z}_l = \prod_{i=1}^{\mathtt{d}} x_{k_i,j_i}$, 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 $\underline{B}$. The power graph edge 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 wet lifted agents consist of the $\mathtt{d}$-tuples of wet grounded agents. The extension of Lemma 6.1 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 $\mathbf{z}$ lay on the algebraic variety $\mathcal{V}$, i.e., $\mathbf{z} = \mathbf{z}(\mathbf{x})$ for some $\mathbf{x}$. This need not be case. A simple fix is to 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 Nalini Anantharaman, Pascal Koiran, John Tsitsiklis, and the anonymous referees for their useful comments and suggestions.

References

[1] Acemoglu, D., Dahleh, M.A., Lobel, I., Ozdaglar, A. *Bayesian learning in social networks*, LIDS report 2780, 2009.

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

[3] Acemoglu, D., Como, G., Fagnani, F., Ozdaglar, A. *Opinion fluctuations and disagreement in social networks*, manuscript, 2011.

[4] 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.

[5] Bhattacharyya, A., Braverman, M., Chazelle, B., Nguyen, H.L. *On the convergence of the Hegselmann-Krause system*, Proc. 4th ITCS, 2013.

[6] 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.

[7] Blondel, V.D., Hendrickx, J.M., Tsitsiklis, J.N. *On the 2R conjecture for multi-agent systems*, Proc. European Control Conference 2007 (ECC 2007), July 2007, 874-881, Kos (Greece), 2007.

[8] Blondel, V.D., Hendrickx, J.M., Tsitsiklis, J.N. *On Krause's multi-agent consensus model with state-dependent connectivity*, IEEE Transactions on Automatic Control 54, 11 (2009), 2586–2597.

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

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

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

[12] Buzzi, J. *Piecewise isometries have zero topological entropy*, Ergodic Theory and Dynamical Systems 21 (2001), 1371–1377.

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

[14] 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

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

[16] 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.

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

[18] Chazelle, B. *The convergence of bird flocking*, arXiv:0905.4241v1, 2009. Prelim. version in Proc. SIAM SODA 2009, with improvements in Proc. ACM SoCG 2010. To appear in J.ACM.

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

[20] Chazelle, B. *Natural algorithms and influence systems*, CACM 55 (2012), 101–110.

[21] 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.

[22] Conradt, L., Roper, T.J. *Group decision making in animals*, Nature 421 (2003), 155–158.

[23] Cucker, F., Smale, S. *Emergent behavior in flocks*, IEEE Trans. Automatic Control 52 (2007), 852–862.

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

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

[26] Earl, M.G., Strogatz, S.H. *Synchronization in oscillator networks with delayed coupling: a stability criterion*, Phys. Rev. E, 67 (2003), 036204(1–4).

[27] Gazi, V., Passino, K.M. *Stability analysis of swarms*, IEEE Transactions on Automatic Control 48 (2003), 692–697.

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

[29] 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).

[30] 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.

[31] 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.

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

[33] Krause, U. *A discrete nonlinear and non-autonomous model of consensus formation*, Communications in Difference Equations (2000), 227236.

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

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

[36] Lorenz, J. *Multidimensional opinion dynamics when confidence changes*, Economic Complexity, Aix-en-Provence, May 2003.

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

[38] Lorenz, J. *Bounds of confidence: Meet, discuss and find consensus!*, Complexity, 4 (2010), 43–52.

[39] Martinez, S., Bullo, F., Cortés, J., Frazzoli, E. *On synchronous robotic networks – Part II: Time complexity of rendezvous and deployment algorithms*, IEEE Transactions on Automatic Control 52 (2007), 2214–2226.

[40] Mirollo, R.E., Strogatz, S. *Synchronization of pulse-coupled biological oscillators*, SIAM J. Applied Math 50 (1990), 1645–1662.

[41] Mirtabatabaei A., Bullo, F. *Opinion dynamics in heterogeneous networks: convergence conjectures and theorems*, SIAM Journal on Control and Optimization, to appear (arxiv preprint, March 2011).

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

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

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

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

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

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

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

[49] Pikovsky, A., Rosenblum, M., Kurths, J. Synchronization: A Universal Concept in Nonlinear Sciences, Cambridge University Press, 2001.

[50] Reynolds, C.W. *Flocks, herds, and schools: A distributed behavioral model*, Computer Graphics 21 (1987), 25–34.

[51] Roozbehani, M., Megretski, A., Frazzoli, E. *Lyapunov analysis of quadratically symmetric neighborhood consensus algorithms*, IEEE Conf. on Decision and Control (2008), 2252–2257, 2008.

[52] Scardovi, L., Sarlette, A., Sepulchre, R. *Synchronization and balancing on the N-torus*, Systems & Control Letters 56 (2007), 335–341.

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

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

[55] Spielman, D.A., Teng, S. *Smoothed analysis: why the simplex algorithm usually takes polynomial time*, J. ACM, 51 (2004), 385–463.

[56] Strogatz, S.H. *From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators*, Physica D 143 (2000), 1–20.

[57] 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.

[58] Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O. *Novel type of phase transition in a system of self-driven particles*, Physical Review Letters 75 (1995), 1226–1229.

[59] Walker, T.J. *Acoustic synchrony: two mechanisms in the snowy tree cricket*, Science 166 (1969), 891–894.

[60] Winfree, A.T. *Biological rhythms and the behavior of populations of coupled oscillators*, J. Theoret. Bio. 16, 1 (1967), 15–42.

Appendix

Proof of Lemma 6.4. 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 $I' \ni 0$. 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 ($k \in I$) 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,

$$N^{-O((k+1)m^2)} \leq \|v_k\|_1 \leq 2^{-k|\log\alpha| + O(\log n)}. \tag{37}$$

The *affine hull* of $V_{|I}$ is the flat defined by $\{\, z^T V_{|I} \,:\, z^T \mathbf{1}_{|I|} = 1 \,\}$: its dimension is called the *affine rank* of $V_{|I}$; not that if it is equal to m then the affine hull must contain the origin. 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 6.4 follows from this inequality, whose proof we postpone: for $r = 0, \ldots, m-1$,

$$g(D, r) < D^{1-\beta^{m+1}} \quad \text{for any} \quad D \geq 2^{(1/\beta)^{m+1}}, \tag{38}$$

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 $V_{|J}$,[24] 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 $z^T \mathbf{1}_{r+1} = 1$, which in turn shows that $\mathbf{1}_{r+1}$ lies outside the column space of $V_{|J}$; therefore M is nonsingular, hence of rank $r+1$. Since each one of its rows consists of $O(m^2 D \log N)$-bit rationals with a common denominator,

$$|\det M| \geq N^{-O(m^3 D)}. \tag{39}$$

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

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

[24] 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 $z^T \mathbf{1} = 1$.

By Hadamard's inequality and (37), each coordinate of ξ is at most $n^{O(m)}$ in absolute value; so, by (39), 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 6.4.

It suffices now to prove (38), 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 (because $0 \in I$), it follows from (37) that

$$|I| \leq 1 + \max\{k \in I\} = O(|\log \alpha|^{-1} m^2 \log N),$$

hence

$$g(D, 0) \leq \beta^{-1}. \tag{40}$$

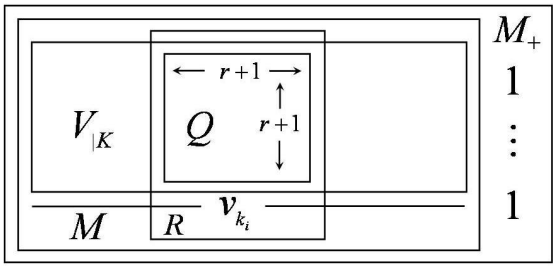

Figure 7: Why a large value of k_i implies that the affine hull of $V_{|I}$ contains the origin.

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 $V_{|K}$.[25] 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 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 $\xi \in \mathbb{R}^{r+2}$:

$$\xi^T M_+ = (\overbrace{0, \ldots, 0}^{m}, 1), \tag{41}$$

where M_+ is the $(r + 2)$-by-$(m + 1)$ matrix $(M, \mathbf{1}_{r+2})$, which proves that the affine hull of M, hence of $V_{|I}$, contains the origin, leading 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

$$k_i > c k_j |\log \alpha|^{-1} m^3 \log N, \tag{42}$$

[25] 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 $j > 0$.

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.7); consistent with our notation, R_+ will denote the square 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 (37), the matrix entries involved in these cofactors are all in $n^{O(1)}$; by Hadamard's inequality, this shows that we can set $\Delta = n^{O(m)}$. By (37), we also find that

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

Since Q is nonsingular, we can adapt (39) to derive $|\det Q| \geq N^{-O(m^3 k_j)}$; hence, by (42), $|\det R_+| > 0$. It follows that the linear system (41) is feasible if we replace M_+ by R_+ and readjust the right-hand side accordingly. 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 (41) contradicts our assumption that the origin is outside the affine hull of $V_{|I}$; therefore

$$k_j \geq \beta k_i > 0, \tag{43}$$

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, for $k \geq 0$, $v_{k_j+k} = w_k A_{k_j} A_{k_j-1} \cdots A_1 A_0$. 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[26] $i \leq g(k_{j-1}, r-1) + g(k_i - k_j, m-1) - 1$. By (43) 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 (40), given $m > 0$ and $D \geq 0$, for $r = 0, \ldots, m-1$,

$$g(D, r) \leq \begin{cases} 1 & \text{if } D = 0 \\ \beta^{-1} & \text{if } r = 0 \\ g(n_1, m-1) + \cdots + g(n_r, m-1) + \beta^{-1} & \text{if } 0 < r < m, \end{cases} \tag{44}$$

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 $D, m > 0$,

$$g(D, m-1) \leq \beta^{-2}(2D^{1-\eta} - 1). \tag{45}$$

[26] It would be nice to bound the affine rank as a function of r and not m, but since we never perturb the transition matrices it is unclear how to do that.

The case $m = 1$ follows from (40). For $m > 1$, we begin with the case $s = 0$, where

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

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

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

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

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

Setting $r = m - 1$ in (44),

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

which proves (45), hence (38) and Lemma 6.4. □