Toward a Theory of Markov Influence Systems and their Renormalization*

Bernard Chazelle

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

Abstract

Nonlinear Markov chains are probabilistic models commonly used in physics, biology, and the social sciences. In *Markov influence systems* (*MIS*), the transition probabilities of the chains change as a function of the current state distribution. This work introduces a renormalization framework for analyzing the dynamics of *MIS*. It comes in two independent parts: first, we generalize the standard classification of Markov chain states to the dynamic case by showing how to "parse" graph sequences. We then use this framework to carry out the bifurcation analysis of a few important *MIS* families. In particular, we show that irreducible *MIS* are almost always asymptotically periodic. We also give an example of "hyper-torpid" mixing, where a stationary distribution is reached in super-exponential time, a timescale that cannot be achieved by any Markov chain.

1998 ACM Subject Classification G.2.2 Graph Theory (F.2.2)

Keywords and phrases Markov influence systems, nonlinear Markov chains, dynamical systems, renormalization, graph sequence parsing

Digital Object Identifier 10.4230/LIPIcs.ITCS.2018.58

1 Introduction

Nonlinear Markov chains are popular probabilistic models in the natural and social sciences. They are commonly used in interacting particle systems, epidemic models, replicator dynamics, mean-field games, etc. [8, 12, 13, 15, 18]. They differ from the linear kind by allowing transition probabilities to vary as a function of the current state distribution.[1] For example, a traffic network might update its topology and edge transition rates adaptively to alleviate congestion. The traditional formulation of these models comes from physics and relies on the classic tools of the trade: stochastic differential calculus, McKean interpretations, Feynman-Kac models, Fokker-Planck PDEs, etc. [3, 5, 13, 18]. These techniques assume all sorts of symmetries that are typically absent from the "mesoscopic" scales of natural algorithms. They also tend to operate at the thermodynamic limit, which rules out genuine agent-based modeling. Our goal is to initiate a theory of discrete-time Markov chains whose topologies vary as a function of the current probability distribution. Of course, the entire theory of finite Markov chains

* The Research was sponsored by the Army Research Office and the Defense Advanced Research Projects Agency and was accomplished under Grant Number W911NF-17-1-0078. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office, the Defense Advanced Research Projects Agency, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

[1] The systems are Markovian in that the future depends only on the present: in this case the current state distribution rather than the single state presently visited.

9th Innovations in Theoretical Computer Science (ITCS 2018).
Editor: Anna R. Karlin; Article No. 58; pp. 58:1–58:18
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

should be recoverable as a special case. Our contribution comes in two parts (of independent interest), which we discuss informally in this introduction.

Renormalization

The term refers to a wide-reaching approach to complex systems that originated in quantum mechanics and later expanded to statistical mechanics and dynamics. Whether in its exact or coarse-grained form, the basic idea is intuitively appealing: break down a complex system into a hierarchy of simpler parts. The concept seems so simple—isn't it what divide-and-conquer is all about?—one can easily be deceived and miss the point. When we slap a dynamics on top of the system (think of interacting particles moving about) then the hierarchy itself creates its own dynamics between the layers. This new "renormalized" dynamics can be entirely different from the original one. Crucially, it can be both easier to analyze and more readily expressive of global properties. For example, second-order phase transitions in the Ising model correspond to fixed points of the renormalized dynamics.[2]

What is the relation to Markov chains? You may have noticed how texts on the subject often dispatch absorbing chains quickly before announcing that from then on all chains will be assumed to be irreducible (and then, usually a few pages later, ergodic). This is renormalization at work! Indeed, although rarely so stated, the standard classification of the states of a Markov chain is a prime example of exact renormalization. Recall that the main idea is to express the chain as an acyclic directed graph, its *condensation*, whose vertices correspond to the strongly connected components. This creates a two-level hierarchy: a tree with a root (the condensation) and its children (the strongly connected components). Now get the chain going and watch what happens at the root: the probability mass flows entirely into the sinks of the condensation. Check the leaves of the tree for a detailed understanding of the motion. The renormalized dynamics (visible only in the condensation) has an attracting manifold that tells much of the story. If the story lacks excitement it is partly because the hierarchy is flattish: only two levels. Time-varying Markov chains, as we shall soon see, do not suffer from that problem.

Consider an infinite sequence $(g_k)_{k>0}$ of digraphs over the same set of vertices. A *temporal* random walk is defined in the obvious way by picking a starting vertex in g_1, moving to a random neighbor, and then repeating this step in g_2, g_3, etc [6, 7, 14, 19]. The walk is called temporal because it traverses one edge from g_t at time $t = 1, 2, \ldots$. How might one go about classifying the states of this "dynamic" Markov chain? Repeating the condensation decomposition at each step makes no sense, as it carries zero information about the temporal walks. The key insight is to monitor when and where temporal walks are *extended*. The *cumulant* graph collects all extensions and, when this process stalls, reboots the process while triggering a deepening of the hierarchy. To streamline this process, we define a grammar with which we can parse the sequence $(g_k)_{k>0}$. The (exact) renormalization framework introduced in this work operates along two tracks: time and network. The first track summarizes the past to anticipate the future while the second one clusters the graphs hierarchically. The method is very general and we expect it to be used elsewhere.

[2] The idea is very powerful: Ken Wilson won the 1982 Nobel prize in physics and Artur Avila the 2014 Fields medal for their (very different) breakthroughs in the use of renormalization: finding new critical exponents; proving the weak mixing of interval exchange transformations, etc.

Markov influence systems

All finite Markov chains oscillate periodically or mix to a stationary distribution. The key fact about their dynamics is that the timescales never exceed a single exponential in the number of states. Allowing the transition probabilities to fluctuate over time at random does not change that basic fact [1, 9, 10]. Markov influence systems are an entirely different beast. Postponing formal definitions, let us think of an *MIS* for now as a dynamical system defined by iterating the map $f: \mathbf{x}^\top \mapsto \mathbf{x}^\top S(\mathbf{x})$, where $\mathbf{x}$ is a probability distribution and $S(\mathbf{x})$ is a stochastic matrix that is piecewise-constant as a function of $\mathbf{x}$. We assume that the discontinuities are linear (ie, hyperplanes). The assumption is not restrictive in any meaningful sense: we explain why with a simple example.

Consider a random variable ξ over the distribution $\mathbf{x}$ and fix two stochastic matrices A and B. Define $S(\mathbf{x}) = A$ (resp. B) if $\mathrm{var}_{\mathbf{x}}\, \xi > 1$ (resp. else); in other words, the Markov chain picks one of two stochastic matrices at each step depending on the variance of ξ with respect to the current state distribution $\mathbf{x}$. This clearly violates our assumption because the discontinuity is quadratic in $\mathbf{x}$; hence nonlinear. This is not an issue because we can linearize the variance: here, we begin with the identity $\mathrm{var}_{\mathbf{x}}\, \xi = \sum_{i,j} (\xi_i - \xi_j)^2 x_i x_j$ and the fact that $\mathbf{y} := (x_i x_j)_{i,j}$ is a probability distribution. We form the Kronecker square $T(\mathbf{x}) = S(\mathbf{x}) \otimes S(\mathbf{x})$ and lift the system into the (n^2-1)-dimensional unit simplex to get a brand-new *MIS* defined by the map $\mathbf{y} \mapsto T(\mathbf{y})$. We now have linear discontinuities. This same type of tensor lift can be used to linearize any algebraic constraints.[3] Using ideas from [4], one can go much further than that and base the step-by-step Markov chain selection on the outcome of any first-order logical formula we may fancy (with the x_i's as free variables).[4] What all of this shows is that the assumption of linear discontinuities is not restrictive.

We prove that irreducible *MIS* are almost always asymptotically periodic. (This assumes that $S(\mathbf{x})$ forms an irreducible chain for each $\mathbf{x}$.) We extend this result to larger families of Markov influence systems. We also give an example of "hyper-torpid" mixing: an *MIS* that converges to a stationary distribution in time equal to a tower-of-twos in the size of the chain. The emergence of timescales far beyond the reach of standard Markov chains is a distinctive feature of Markov influence systems. We note that the long-time horizon analysis of general systems is still open.

Some intuition

Is there a quick, intuitive explanation why the analysis of *MIS* should require all of that renormalization machinery? Of course, perhaps it does not and future work will show how to bypass it. But the specific challenges raised by the model are easy to state. The first hurdle is that Markov influence systems are not globally contractive. Worse, the eigenspaces over which they are not may be constantly changing over time. It is this spectral incoherence that renormalization attempts to "tame." To see why this has a strong graph-theoretic flavor, consider the fact that the stationary distributions may change at each time step and so can their number. The key insight is that these changes can be read off the topology of the graph: for example, the number of sinks in the condensation is precisely equal to the dimension of the principal eigenspace. Renormalization can thus be seen as an attempt to

---

[3] This requires making the polynomials over x_i homogeneous, which we can do by using the identity $\sum_i x_i = 1$.

[4] The key fact behind this result is that the first-order theory of the reals is decidable by quantifier elimination. This allows us to pick the next stochastic matrix at each time step on the basis of the truth value of a Boolean logic formula with arbitrarily many quantifiers. See [4] for details.

restore coherence to an ever-changing spectral landscape via a dynamic hierarchy of graphs, subgraphs, and homomorphs.

The bifurcation analysis at the heart of our analysis of Markov influence systems follows an approach commonly used in dynamics [2, 16, 21]: the idea is develop a notion of "general position" in order to bound the growth rate of the induced symbolic dynamics. The root of the problem is a clash between order and randomness. (This is the same conflict that arises between entropy and energy in statistical mechanics.) All Markov chains are attracted to a limit cycle (ie, order). Changing the chain at each step introduces pseudorandomness into the process (ie, disorder) and the question is to know which one of the two "forces" of order or disorder will prevail. The conflict is mediated by introducing a perturbation parameter and locating its critical values. We show that, in this case, the critical region forms a Cantor set of Hausdorff dimension strictly less than 1.

Previous work

There is a growing body of literature on dynamic graphs [3, 14, 17, 19] and their random walks [1, 6, 7, 8, 12, 9, 10, 15]. By contrast, as mentioned earlier, most of the research on nonlinear Markov chains has been done within the framework of stochastic differential calculus. The closest analog to the *MIS* model are the diffusive influence systems we introduced in [4]. The relation is interesting. Random walks and diffusion are dual processes that coincide only when the underlying operator is self-adjoint (which is not the case here). As a rule of thumb, diffusion is easier to analyze because even in a changing medium the constant function is always a principal eigenfunction. As a result, a diffusion model can converge to a fixed point while its dual Markov process does not. The reason the dynamics is so different is that, as has long been known [20], multiplying stochastic matrices from the right is harder than from the left.[5] Our renormalization scheme is new, but the idea of parsing graph sequences is not. We introduced it in [4] as a way of tracking the flow of information across changing graphs. The parsing method we discuss here is entirely different, however: being *topological* rather than *informational*, it is vastly more general and, we believe, likely to be useful in other applications of dynamic networks.

2 How to Parse a Graph Sequence

Throughout this work, a *digraph* refers to a directed graph with vertices in $[n]$ and a self-loop at each vertex.[6] We denote digraphs by lower-case letters (g, h, etc) and use boldface symbols for sequences. A *digraph sequence* $\mathbf{g} = (g_k)_{k>0}$ is an ordered, finite or infinite, list of digraphs over the vertex set $[n]$. The digraph $g_i \times g_j$ consists of all the edges (x, y) such that there exist at least an edge (x, z) in g_i and another one (z, y) in g_j. The operation $\times$ is associative.[7] We define the *cumulant* $\prod_{\leq k} \mathbf{g} = g_1 \times \cdots \times g_k$ and write $\prod \mathbf{g} = g_1 \times g_2 \times \cdots$ for finite $\mathbf{g}$. The cumulant indicates all the pairs of vertices that can be joined by a temporal walk of a given length. We need additional terminology:

[5] For example, consider the product AB of two stochastic matrices, where $\text{rank}(B) = 1$. We have $AB = B$ whereas BA can be any old stochastic matrix of rank 1.
[6] The graphs and digraphs (words we use interchangeably) have no multiple edges and $[n] := \{1, \ldots, n\}$.
[7] The sign is meant to highlight the connection with the multiplication of incidence matrices.

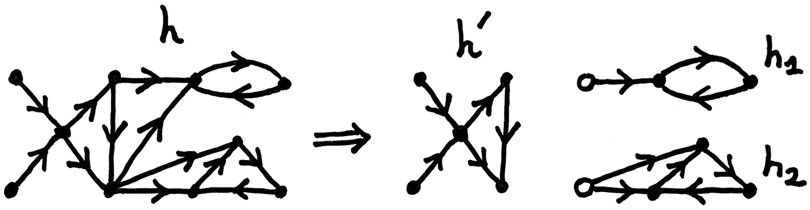

Figure 1 The decomposition of h into its stem h' and its petals h_1, h_2.

- *Transitive front of g*: An edge (x, y) of a digraph g is *leading* if there is u such that (u, x) is an edge of g but (u, y) is not.[8] The non-leading edges form a subgraph $tf(g)$, called the transitive front of g. We omit the (easy) proof that the transitive front is indeed transitive, ie, if (x, y) and (y, z) are edges of $tf(g)$ then so is (x, z). Given two graphs g, h over the same vertex set, we write $g \preceq h$ if all the edges of g are in h (with strict inclusion denoted by the symbol $\prec$). Because of the self-loops, $g, h \preceq g \times h$. We easily check that the transitive front of g is the (unique) densest graph h such that $g \times h = g$.
- *Subgraphs and contractions*: Given two digraphs g, h with vertex sets $V_g \supseteq V_h$, we denote by $g_{|\,h}$ the subgraph of g induced by V_h. Pick $U \subseteq V_h$ and contract all these vertices into a single one. By abuse of notation, we still designate by $g_{|\,h}$ the graph derived from g by first taking the subgraph induced by V_h and then contracting the vertices of U; note that the notation $g_{|\,(V_h,U)}$ would be more accurate but it will not be needed. Given a sequence $\mathbf{g} = (g_k)_{k>0}$, we use the shorthand $\mathbf{g}_{|\,h}$ for $(g_{k\,|\,h})_{k>0}$. Finally, $\mathbb{K}$ denotes the set of all complete digraphs (of any size) with self-loops, while $\mathbb{K}\otimes 1$ consists of the complete digraphs with an extra vertex pointing to all the others unidirectionally.[9]
- *Stem decomposition of h*: The strongly connected components of a graph h form, by contraction, an acyclic digraph called its *condensation*. Let $V_1, \ldots, V_\ell$ be the vertex sets from $[n]$ corresponding to the ℓ sinks of the condensation.[10] The remaining vertices of h induce a subgraph h' called the *stem* of h. For each $i \in [\ell]$, the *petal* h_i is the subgraph induced by V_i if no vertex outside V_i links to it; else h_i is the subgraph induced by V_i and h', with all the vertices of h' subsequently contracted into a single vertex and the multiple edges removed (fig.1).

The parser

The parse tree of a (finite or infinite) graph sequence $\mathbf{g}$ is a rooted tree whose leaves are associated with $g_1, g_2, \ldots$ from left to right; each internal node assigns a syntactical label to the subsequence $g_i, \ldots, g_j$ formed by the leaves of its subtree. The purpose of the parse tree is to monitor the formation of new temporal walks as time progresses. How to do that begins with the observation that the cumulant $\prod_{\leq k} \mathbf{g}$ is monotonically nondecreasing.[11] If the increase was strict at each step then the parse tree would be trivial: each graph of $\mathbf{g}$ would appear as a separate leaf with the root as its parent. Of course, the increase cannot

[8] For example, $tf(x \to y \to z)$ is the graph over x, y, z with the single edge $x \to y$ (and the three self-loops.) If g is transitive, then $tf(g) = g$. The transitive front of a directed cycle has no edges besides the self-loops.
[9] For example, ignoring self-loops, $\{(x, y), (y, x)\} \in \mathbb{K}$ and $\{(x, y), (y, x), (z, x), (z, y)\} \in \mathbb{K}\otimes 1$.
[10] These are the vertices with no outgoing edges: there is at least one of them; hence $\ell > 0$.
[11] All references to graph ordering are relative to $\preceq$.

go on forever. How to deal with time intervals within which the cumulant is "stuck" is the whole point of parsing: Short answer: proceed recursively. The grammar consists of only two pairs of productions, (1a,1b) and (2a,2b).

1. TIME RENORMALIZATION Let m be the smallest index k at which $\prod_{\leq k} \mathbf{g}$ achieves its maximal value; write $\mathbf{g}_l = (g_k)_{k<m}$, $\mathbf{g}_r = (g_k)_{k>m}$, and $h = tf(\prod_{\leq m} \mathbf{g})$. The two productions below cluster time into the relevant intervals.
   - **a.** *Transitivization.* Using a parenthesis system to express the parse tree, the first production supplies the root with at most three children:

$$\mathbf{g} \longrightarrow \big(\, \mathbf{g}_l \,\big)\, g_m \,\big(\, \mathbf{g}_r \bigtriangleup h \,\big)\,, \tag{1a}$$

   where h is transitive, and $\mathbf{g}_l$ or $\mathbf{g}_r$ (or both) may be the empty sequence $\varnothing$. If $\mathbf{g}_l \neq \varnothing$, then $\prod \mathbf{g}_l \prec \prod_{\leq m} \mathbf{g} = \prod \mathbf{g}$. The right sibling of $(\mathbf{g}_l)$ is the terminal symbol g_m (a leaf of the parse tree) followed by $\mathbf{g}_r \bigtriangleup h$. The annotation $\bigtriangleup h$ indicates that $\prod \mathbf{g}_r \preceq h$ and that h will "guide" the parsing of $\mathbf{g}_r$.[12] Observe that h is available when needed but not earlier. This ensures that the parsing is of the *LR type*, meaning that it can be carried out bottom-up in a single left-to-right scan.[13]
   - **b.** *Cumulant completion.* We parse $\mathbf{g} \bigtriangleup h$ in the special case where h is in $\mathbb{K}$ or $\mathbb{K}\otimes 1$. Recall that the notation $\bigtriangleup$ implies that $\prod \mathbf{g} \preceq h$. Partition the sequence $\mathbf{g}$ into minimal subsequences $\mathbf{g}_k\, g_{m_k}$ such that $\prod \mathbf{g}_k \prec h = \big(\prod \mathbf{g}_k\big) \times g_{m_k}$:

$$\mathbf{g} \bigtriangleup h \longrightarrow \big(\mathbf{g}_1\big)\, g_{m_1} \,\big(\mathbf{g}_2\big)\, g_{m_2} \cdots \tag{1b}$$

   The list on the right-hand side could be finite or infinite; if finite, it could be missing the final g_{m_k}. This production is the one doing the heavy lifting in that it establishes a bridge between renormalization and Lyapunov exponents.
2. NETWORK RENORMALIZATION Two productions parse the rightmost term in (1a) by recursively breaking down the graph into clusters. This is done either by carving out subgraphs or taking homomorphs. In both cases, it is assumed that $\prod \mathbf{g} \preceq h$ and that h is transitive but *not* in $\mathbb{K}$ or $\mathbb{K}\otimes 1$.
   - **a.** *Decoupling.* If the number of connected components $h_1, \ldots, h_k$ of h exceeds one, then[14]

$$\mathbf{g} \bigtriangleup h \longrightarrow \big(\, \mathbf{g}_{\,|\, h_1} \bigtriangleup h_1 \,\big) \,\|\, \cdots \,\|\, \big(\, \mathbf{g}_{\,|\, h_k} \bigtriangleup h_k \,\big)\,. \tag{2a}$$

   In terms of the parse tree, the node has k children that model processes operating in parallel. Intuitively, the production decouples the system into the subsystems formed by the components. As we show below, this does not always imply the independence of the respective dynamics, however.
   - **b.** *One-way coupling.* If the undirected version of h has a single connected component, we use its stem decomposition $h', h_1, \ldots, h_\ell$ to cluster the digraphs of $\mathbf{g}$:

$$\mathbf{g} \bigtriangleup h \longrightarrow \big(\, \mathbf{g}_{\,|\, h'} \bigtriangleup h' \,\big) \;\Vdash\; \Big\{\, \big(\, \mathbf{g}_{\,|\, h_1} \bigtriangleup h_1 \,\big) \,\|\, \cdots \,\|\, \big(\, \mathbf{g}_{\,|\, h_\ell} \bigtriangleup h_\ell \,\big) \,\Big\}\,. \tag{2b}$$

---

[12] By definition of g_m, no temporal walk from $\mathbf{g}_r$ can extend one from $(g_k)_{k\leq m}$. This shows that $\big(\prod_{\leq m} \mathbf{g}\big) \prod \mathbf{g}_r = \prod_{\leq m} \mathbf{g}$; hence $\prod \mathbf{g}_r \preceq h$.

[13] This is usually a requirement for the bifurcation analysis. If not for that, we could use the simpler production $\mathbf{g} \to \mathbf{g} \bigtriangleup h$, instead.

[14] This refers to the subgraphs of h induced by each one of the vertex subsets of the connected components of the undirected version of h.

Since h is neither in $\mathbb{K}$ nor in $\mathbb{K}\otimes 1$, its stem and petals both exist (with $\ell > 0$). The assumed transitivity of h implies that each $h_i \in \mathbb{K}\otimes 1$. We iterate the production if h' is neither in $\mathbb{K}$ nor in $\mathbb{K}\otimes 1$. System-wise, the symbol ╠ indicates the direction of the information flow. None flows into $\mathbf{g}_{\,|\,h'} \triangle\, h'$, so its dynamics is decoupled from the rest. Such decoupling does not hold for the petals, so it is one-way. This allows us to renormalize the stem into a single vertex for the purposes of the petals: the common 1 in all the instances of $\mathbb{K}\otimes 1$. In terms of the parse tree, the nodes has $\ell + 1$ children that operate in parallel, with the last ℓ of them collecting information from the first one.

Network renormalization exploits the fact that the information flowing across the system might get stuck in portions of the graph for some period of time: we cluster the graph when and where this happens. Sometimes only time renormalization is possible. Consider the infinite sequence $(g_k)_{k>0}$, where $g_k = h_{k \,(\mathrm{mod}\, n)}$ and, for $k = 1, \ldots, n$, h_k consists of the graph at the vertices and edges from k to all $n-1$ other vertices: the cumulant never ceases to grow until it reaches $\mathbb{K}$, at which point the process repeats itself; the parsing involves n applications of (1a) with $\mathbf{g}_r = \varnothing$, followed by infinitely many calls to (1b). There is no network renormalization. Quite the opposite, the case of an infinite single-graph sequence features abundant network renormalization (fig.2).

The depth of the parse tree

It is easily verified that cumulants $\prod \mathbf{g}$ lose at least one edge from parent to child, which puts an obvious bound of n^2 on the maximal height of the parse tree.[15] This quadratic bound is tight. Indeed, consider the sequence (g_k), where $g_{k+1} = h_{k \,(\mathrm{mod}\, n-1)}$ for $0 \le k \le (n-2)(n-1)$, and (besides self-loops) h_k consists of the single edge $(k+2, k+1)$ for $k = 0, \ldots, n-2$. The j-th copy of h_k adds to the cumulant the new edge $(k+j+1, k+1)$, which creates, in total, a quadratic number of increments. The bounded depth implies that the parse tree for an infinite sequence includes exactly one node with an infinite number of children. That node is expressed by a production of type (1b).

Undirected graphs

Note that the cumulant of a sequence of undirected graphs might itself be directed.[16] Suppose that an undirected edge e of the digraph g does not extend any edge of g into a new temporal 2-edge walk.[17] Obviously, e must already be in the digraph; furthermore, any vertex linking to one of its endpoints must link to the other one as well. This implies that the *undirected transitive front* of a directed graph (say, a cumulant) consists of disjoint cliques of undirected edges. This simplifies the parsing since the condensation is trivial and the parsing tree has no nodes of type (2b). The complexity of the parse tree can still be as high as quadratic, however. To see why, consider the following recursive construction: given a clique C_k over k vertices at time t, attach to it, at time $t+1$, a two-edge path $(x, y), (y, z)$, say, at x. The cumulant gains the edge (y, z) as well as all k edges joining y to the clique. At time $t+2, \ldots, t+k+1$, visit each one of these k edges by using single-edge graphs g_i. Each such step will see the addition of a new edge to the cumulant, until it becomes a clique C_{k+2}.

[15] This differs from the linear-depth trees derived from the flow tracker [4].
[16] The product $(x \leftrightarrow y \leftrightarrow z) \times (x \quad y \leftrightarrow z)$ has a directed edge from x to z but not from z to x.
[17] An edge is undirected if both of its directed versions are present in the graph.

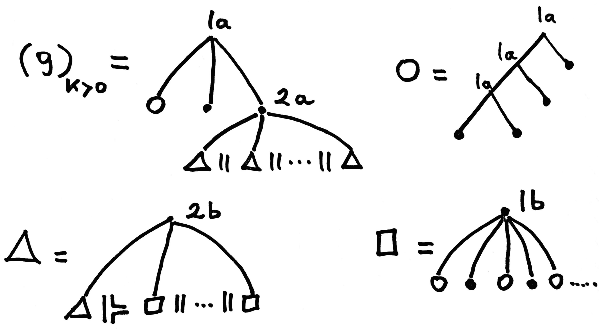

Figure 2 The parse tree of an infinite sequence consisting of the same graph g.

(Note that visiting (y, z) would ruin the whole construction.) The quadratic lower bound on the tree depth follows immediately.

Backward parsing

The sequence of graphs leads to products where each new graph is multiplied to the right, as would happen in a time-varying Markov chain. Algebraically, the matrices are multiplied from left to right. In diffusive systems (eg, multiagent agreement systems, Hegselmann-Krause models, Deffuant systems, voter models), however, matrices are multiplied from right to left. Although the dynamics can be quite different, the same parsing algorithm can be used. Given a sequence $\mathbf{g} = (g_k)_{k>0}$, its *backward parsing* is formed by applying the parser to the sequence $\overleftarrow{\mathbf{g}} = (h_k)_{k>0}$, where h_k is derived from g_k by reversing the direction of every edge, ie, (x, y) becomes (y, x). Once the parse tree for $\overleftarrow{\mathbf{g}}$ has been built, we simply restore each edge to its proper direction to produce the *backward parse tree* of $\mathbf{g}$.

3 The Markov Influence Model

Let $\mathbb{S}^{n-1}$ (or $\mathbb{S}$ when the dimension is understood) be the standard simplex $\{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} \geq \mathbf{0}\,, \|\mathbf{x}\|_1 = 1 \,\}$ and let $\mathcal{S}$ denote set of all n-by-n rational stochastic matrices. A *Markov influence system* (*MIS*) is a discrete-time dynamical system with phase space $\mathbb{S}$, which is defined by the map $f : \mathbf{x} \mapsto f(\mathbf{x}) := \mathbf{x}^\top S(\mathbf{x})$, where S is a function $\mathbb{R}^n \mapsto \mathcal{S}^n$ that is constant over the pieces of a finite polyhedral partition[18] $\mathcal{P} = \{P_k\}$ of $\mathbb{R}^n$. We define the digraph $g(\mathbf{x})$ (and its corresponding Markov chain) formed by the positive entries of $S(\mathbf{x})$. To avoid irrelevant technicalities, we assume the presence of self-loops in $g(\mathbf{x})$, ie, $S(\mathbf{x})_{ii} > 0$. In this way, any orbit of an *MIS* corresponds to a lazy, time-varying random walk with transitions defined endogenously.[19] We recall some basic terminology. The *orbit* of $\mathbf{x} \in \mathbb{S}$ is the infinite sequence $(f^t(\mathbf{x}))_{t\geq 0}$ and its *itinerary* is the corresponding sequence of cells P_k visited in the process. The orbit is *periodic* if $f^t(\mathbf{x}) = f^s(\mathbf{x})$ for any $s = t$ modulo a fixed integer. It is asymptotically periodic if it gets arbitrarily close to a periodic orbit over time.

For convenience, we assume a representation of the discontinuities induced by $\mathcal{P}$ as hyperplanes H_i of the form $\mathbf{a}_i^\top \mathbf{x} = 1 + \delta$, where $\delta \in \frac{1}{2}[-1, 1]$ (for concreteness). Note that the polyhedral partition is invariant up to scaling for all values of the bifurcation parameter,

---

[18] How f is defined on the discontinuities of the partition is immaterial.

[19] As discussed in the introduction, to access the full power of first-order logic in the stepwise choice of digraphs requires nonlinear partitions, which can be handled by a suitable tensor lift.

so the *MIS* remains well-defined as we vary δ. The parameter δ is necessary for the analysis: indeed, as we explain below in Section 5, chaos cannot be avoided without it. The *coefficient of ergodicity* $\tau(M)$ of a matrix M is defined as half the maximum ℓ_1-distance between any two of its rows [20]. It is submultiplicative for stochastic matrices, a direct consequence of the identity

$$\tau(M) = \max\Big\{ \|\mathbf{x}^\top M\|_1 : \mathbf{x}^\top \mathbf{1} = 0 \text{ and } \|\mathbf{x}\|_1 = 1 \Big\}.$$

Given $\Omega \subset \mathbb{R}$, let L^t_Ω denote the set of t-long prefixes of any itinerary for any starting position $\mathbf{x} \in \mathbb{S}$ and any $\delta \in \Omega$. We define the *ergodic renormalizer* $\eta = \eta(\Omega)$ as the smallest integer such that, for any $t \geq \eta$ and any matrix sequence $S_1, \ldots, S_t$ matching an element of L^t_Ω, the product $S_1 \cdots S_t$ is primitive (ie, some high enough power is a positive matrix) and its coefficient of ergodicity is less than $1/2$. We assume in this section that $\eta = \eta(\mathbb{R}) < \infty$ and discuss in §4 how to remove this assumption via renormalization. Let D be the union of the hyperplanes H_i from $\mathcal{P}$ in $\mathbb{R}^n$ (where δ is understood). We define $Z_t = \bigcup_{0 \leq k \leq t} f^{-k}(D)$ and $Z = \bigcup_{t \geq 0} Z_t$. Remarkably, for almost all δ, Z_t becomes strictly equal to Z in a finite number of steps.

▶ **Lemma 1.** *Given any $\varepsilon > 0$, there exists an integer $\nu \leq 2^{\eta^{O(1)}} |\log \varepsilon|$ and a finite union K of intervals of total length less than ε such that $Z = Z_\nu$ for any $\delta \notin K$.*

▶ **Corollary 2.** *For δ almost everywhere,*[20] *every orbit is asymptotically periodic.*

Proof. The equality $Z = Z_\nu$ implies the eventual periodicity of the symbolic dynamics. The period cannot exceed the number of connected components in the complement of Z. Once an itinerary becomes periodic at time t_o with period σ, the map f^t can be expressed locally by matrix powers. Indeed, divide $t - t_o$ by σ and let q be the quotient and r the remainder; then, locally, $f^t = g^q \circ f^{t_o + r}$, where g is specified by a stochastic matrix with a positive diagonal, which implies convergence to a periodic point at an exponential rate. Apply Lemma 1 repeatedly, with $\varepsilon = 2^{-l}$ for $l = 1, 2, \ldots$ and denote by K_l be the corresponding union of "forbidden" intervals. Define $K^l = \bigcup_{j \geq l} K_j$ and $K^\infty = \bigcap_{l > 0} K^l$: $\text{Leb}(K^l) < 2^{1-l}$; hence $\text{Leb}(K^\infty) = 0$. The lemma follows from the fact that any δ outside of K^∞ lies outside of K^l for some $l > 0$. ◀

The corollary states that the set of "nonperiodic" values of δ has measure zero in parameter space. Our result is actually stronger than that. We prove that the nonperiodic set can be covered by a Cantor set of Hausdorff dimension strictly less than 1. The remainder of this section is devoted to a proof of Lemma 1.

3.1 Shift spaces and growth rates

The *growth exponent* of a language is defined as $\lim_{n \to \infty} \frac{1}{n} \max_{k \leq n} \log N(k)$, where $N(k)$ is the number of words of length k; for example, the growth exponent of $\{0,1\}^*$ is 1. The language consisting of all the itineraries of a Markov influence system forms a *shift space* and its growth exponent is the *topological entropy* of its symbolic dynamics [21].[21] It can be strictly positive, which is a sign of chaos. We show that, for a typical system, it is zero, the key fact driving periodicity. Let $M_1, \ldots, M_T$ be n-by-n matrices from a fixed set $\mathcal{M}$ of

[20] Meaning everywhere in $\frac{1}{2}[-1, 1]$ outside a set of Lebesgue measure zero.
[21] Which should not be confused with the topological entropy of the *MIS* itself.

primitive stochastic rational matrices with positive diagonals, and assume that $\tau(M) < 1/2$ for $M \in \mathcal{M}$; hence $\tau(M_1 \cdots M_k) < 2^{-k}$. Because each product $M_1 \cdots M_k$ is a primitive matrix, it can be expressed as $\mathbf{1}\boldsymbol{\pi}_k^\top + Q_k$ (by Perron-Frobenius), where $\boldsymbol{\pi}_k$ is its (unique) stationary distribution.[22] If $\boldsymbol{\pi}$ is a stationary distribution for a stochastic matrix S, then its j-th row $\boldsymbol{s}_j$ satisfies $\boldsymbol{s}_j - \boldsymbol{\pi}^\top = \boldsymbol{s}_j - \boldsymbol{\pi}^\top S = \sum_i \boldsymbol{\pi}_i(\boldsymbol{s}_j - \boldsymbol{s}_i)$; hence, by the triangular inequality, $\|\boldsymbol{s}_j - \boldsymbol{\pi}\|_1 \leq \sum_i \pi_i \|\boldsymbol{s}_j - \boldsymbol{s}_i\|_1 \leq 2\tau(S)$. This implies that

$$\begin{cases} M_1 \cdots M_k = \mathbf{1}\boldsymbol{\pi}_k^\top + Q_k \\ \|Q_k\|_\infty \leq 2\tau(M_1 \cdots M_k) < 2^{1-k}. \end{cases} \tag{1}$$

Property U

Fix a vector $\mathbf{a} \in \mathbb{Q}^n$, and denote by $M^{(\theta)}$ the n-by-m matrix with the m column vectors $M_1 \cdots M_{k_i}\,\mathbf{a}$, where $\theta = (k_1, \dots, k_m)$ is an increasing integer sequence of nonnegative integers in $[T]$. We say that property $\mathbf{U}$ holds if there exists a vector $\mathbf{u}$ such that $\mathbf{1}^\top\mathbf{u} = 1$ and $\mathbf{x}^\top M^{(\theta)}\mathbf{u}$ does not depend on the variable $\mathbf{x} \in \mathbb{S}$.[23] Intuitively, property $\mathbf{U}$ is a quantifier elimination device for expressing "general position" for *MIS*. To see the connection, consider a simple statement such as "the three points (x, x^2), $(x+1, (x+1)^2)$, and $(x+2, (x+2)^2)$ cannot be collinear for any value of x." This can be expressed by saying that a certain determinant polynomial in x is constant. Likewise, the vector $\mathbf{u}$ manufactures a quantity, $\mathbf{x}^\top M^{(\theta)}\mathbf{u}$, that "eliminates" the variable $\mathbf{x}$. Note that some condition on $\mathbf{u}$ is obviously needed since we could pick $\mathbf{u} = \mathbf{0}$. We explain below why $\mathbf{1}^\top\mathbf{u} = 1$ is the right condition.

▶ **Lemma 3.** *There exists a constant $b > 0$ (linear in n) such that, given any integer $T > 0$ and any increasing sequence θ in $[T]$ of length at least $T^{1-\alpha}/\alpha$, property $\mathbf{U}$ holds, where $\alpha := \mu^{-b}$ and μ is the number of bits needed to encode any entry of M_k for any $k \in [T]$.*

Proof. By choosing b large enough, we can automatically ensure that T is as big as we want.[24] The proof is a mixture of algebraic and combinatorial arguments. We begin with a Ramsey-like statement about stochastic matrices.

▶ **Lemma 4.** *There is a constant $d > 0$ such that, if the sequence θ contains $j_0, \dots, j_n$ with $j_i \geq d\mu j_{i-1}$ for each $i \in [n]$, then property $\mathbf{U}$ holds.*

Proof. By (1), $\|Q_k\mathbf{a}\|_\infty < c_0 2^{-k}$ for constant $c_0 > 0$. Note that Q_k has rational entries over $O(\mu k)$ bits (with the constant factor depending on n). We write $M^{(\theta)} = \mathbf{1}\mathbf{a}^\top\Pi^{(\theta)} + Q^{(\theta)}$, where $\Pi^{(\theta)}$ and $Q^{(\theta)}$ are the n-by-m matrices formed by the m column vectors $\boldsymbol{\pi}_{k_i}$ and $Q_{k_i}\mathbf{a}$, respectively, for $i \in [m]$; recall that $\theta = (k_1, \dots, k_m)$. The key fact is that the dependency on $\mathbf{x} \in \mathbb{S}$ is confined to the term $Q^{(\theta)}$: indeed,

$$\mathbf{x}^\top M^{(\theta)}\mathbf{u} = \mathbf{a}^\top\Pi^{(\theta)}\mathbf{u} + \mathbf{x}^\top Q^{(\theta)}\mathbf{u}. \tag{2}$$

This shows that, in order to satisfy property $\mathbf{U}$, it is enough to ensure that $Q^{(\theta)}\mathbf{u} = 0$ has a solution such that $\mathbf{1}^\top\mathbf{u} = 1$. Let $\sigma = (j_0, \dots, j_{n-1})$. If $Q^{(\sigma)}$ is nonsingular then, because

---

[22] Positive diagonals play a key role here because primitiveness is not closed under multiplication: for example, $\left(\begin{smallmatrix}1&1\\1&0\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}0&1\\1&1\end{smallmatrix}\right)$ are both primitive but their product is not.

[23] Because $\mathbf{x}$ is a probability distribution, property $\mathbf{U}$ does not imply $\mathbf{x}^\top M^{(\theta)}\mathbf{u} = 0$; for example, we have $\mathbf{x}^\top\left(\mathbf{1}\mathbf{1}^\top\right)\mathbf{u} = 1$, for $\mathbf{u} = \frac{1}{n}\mathbf{1}$.

[24] All the constants in this work may depend on the input parameters such as n, $\mathcal{P}$, etc. Dependency on other parameters is indicated by a subscript.

each one of its entries is a rational over $O(\mu j_{n-1})$ bits, we have $|\det Q^{(\sigma)}| \geq c_1^{\mu j_{n-1}}$, for constant $c_1 > 0$. Let R be the $(n+1)$-by-$(n+1)$ matrix derived from $Q^{(\sigma)}$ by adding the column $Q^{j_n}\mathbf{a}$ to its right and then adding a row of ones at the bottom. If R is nonsingular, then $R\,\mathbf{u} = (0, \ldots, 0, 1)^\top$ has a (unique) solution in $\mathbf{u}$ and property $\mathbf{U}$ holds (after padding $\mathbf{u}$ with zeroes). Otherwise, we expand the determinant of R along the last column. Suppose that $\det Q^{(\sigma)} \neq 0$. By Hadamard's inequality, all the cofactors are at most a constant $c_2 > 0$ in absolute value; hence, for d large enough,

$$0 = |\det R\,| \geq |\det Q^{(\sigma)}| - nc_2\|Q_{j_n}\mathbf{a}\,\|_\infty \geq c_1^{\mu j_{n-1}} - nc_2c_0 2^{-j_n} > 0.$$

This contradiction implies that $Q^{(\sigma)}$ is singular, so (at least) one of its rows can be expressed as a linear combination of the others. We form the n-by-n matrix R' by removing that row from R, together with the last column, and setting $u_{j_n} = 0$ to rewrite $Q^{(\theta)}\mathbf{u} = 0$ as $R'\mathbf{u}' = (0, \ldots, 0, 1)^\top$, where $\mathbf{u}'$ is the restriction of $\mathbf{u}'$ to the columns indexed by R'. Having reduced the dimension of the system by one variable, we can proceed inductively in the same way; either we terminate with the discovery of a solution or the induction runs its course until $n = 1$ and the corresponding 1-by-1 matrix is null, so that the solution 1 works. Note that $\mathbf{u}$ has rational coordinates over $O(\mu T)$ bits. ◀

Let $N(T)$ be the largest sequence θ in $[T]$ such that property $\mathbf{U}$ does not hold. Divide $[T]$ into bins $[(d\mu)^k, (d\mu)^{k+1} - 1]$ for $k \geq 0$. By Lemma 4, the sequence θ can intersect at most $2n$ of them, so, if $T > t_0$, for some large enough $t_0 = (d\mu)^{O(n)}$, there is at least one empty interval in T of length $T/(d\mu)^{2n+3}$. This gives us the recurrence $N(T) \leq T$ for $T \leq t_0$ and $N(T) \leq N(T_1) + N(T_2)$, where $T_1 + T_2 \leq \beta T$, for a positive constant $\beta = 1 - (d\mu)^{-2n-3}$. The recursion to the right of the empty interval, say, $N(T_2)$, warrants a brief discussion. The issue is that the proof of Lemma 4 relies crucially on the property that Q_k has rational entries over $O(\mu k)$ bits—this is needed to lower-bound $|\det Q^{(\sigma)}|$ when it is not 0. But this is not true any more, because, after the recursion, the columns of the matrix $M^{(\theta)}$ are of the form $M_1 \cdots M_k\,\mathbf{a}$, for $T_1 + L < k \leq T$, where L is the length of the empty interval and $T = T_1 + T_2 + L$. Left as such, the matrices use too many bits for the recursion to go through. To overcome this obstacle, we observe that the recursively transformed $M^{(\theta)}$ can be factored as AB, where $A = M_1 \cdots M_{T_1+L}$ and B consists of the column vectors $M_{T_1+L+1} \cdots M_k\,\mathbf{a}$. The key observation now is that, if $\mathbf{x}^\top B\,\mathbf{u}$ does not depend on $\mathbf{x}$, then neither does $\mathbf{x}^\top M^{(\theta)}\,\mathbf{u}$, since it can be written as $\mathbf{y}^\top B\,\mathbf{u}$ where $\mathbf{y} = A^\top\mathbf{x} \in \mathbb{S}$. In this way, we can enforce property $\mathbf{U}$ while having restored the proper encoding length for the entries of $M^{(\theta)}$.

Plugging in the ansatz $N(T) = t_0 T^\gamma$, for some unknown positive $\gamma < 1$, we find by Jensen's inequality that, for all $T > 0$, $N(T) \leq t_0(T_1^\gamma + T_2^\gamma) \leq t_0 2^{1-\gamma}\beta^\gamma T^\gamma$. For the ansatz to hold true, we need to ensure that $2^{1-\gamma}\beta^\gamma \leq 1$. Setting $\gamma = 1/(1 - \log\beta) < 1$ completes the proof of Lemma 3. ◀

Define $\phi^k(\mathbf{x}) = \mathbf{x}^\top M_1 \cdots M_k$ and let $h_\delta : \mathbf{a}^\top\mathbf{x} = 1 + \delta$ be some hyperplane in $\mathbb{R}^n$. We consider a set of canonical intervals of length ρ (or less): $\mathcal{D}_\rho = \{\, [k\rho, (k+1)\rho] \cap \varepsilon\,\mathbb{I} \mid k \in \mathbb{Z}\,\}$, where $\mathbb{I} := [-1, 1]$ and $\varepsilon > 0$ is the parameter of Lemma 1. Roughly, the "general position" lemma below says that, for most δ, the ϕ^k-images of any ρ-wide cube centered in the simplex $\mathbb{S}^{n-1}$ cannot near-collide with the hyperplane $\mathbf{a}^\top\mathbf{x} = 1 + \delta$ for most values of $k \leq T$. This may sound seriously counterintuitive. After all, if the stochastic matrices M_i are the identity, the images do not move, so if the initial cube collide then all of the images will! The point is that M_i is primitive so it cannot be the identity. The low coefficients of ergodicity will also play a key role. Notation: α refers to its use in Lemma 3.

▶ **Lemma 5.** *For any real $\rho > 0$ and any integer $T > 0$, there exists $U \subseteq \mathcal{D}_\rho$ of size $c_T = 2^{O(\mu T)}$, where c_T is independent of ρ, such that, for any $\Delta \in \mathcal{D}_\rho \backslash U$ and $\mathbf{x} \in \mathbb{S}$, there* *are at most* $T^{1-\alpha}/\alpha$ *integers* $k \le T$ *such that* $\phi^k(X) \cap h_\Delta \neq \emptyset$*, where* $X = \mathbf{x} + \rho\mathbb{I}^n$ *and* $h_\Delta := \bigcup_{\delta\in\Delta} h_\delta$.

Proof. In what follows, $b_0, b_1, \dots$ refer to suitably large positive constants. We assume the existence of more than $T^{1-\alpha}/\alpha$ integers $k \le T$ such that $\phi^k(X) \cap h_\Delta \neq \emptyset$ and draw the consequences: in particular, we infer certain linear constraints on δ; by negating them, we define the forbidden set U and ensure the conclusion of the lemma. Let $k_1 < \cdots < k_m$ be the integers in question, where $m > T^{1-\alpha}/\alpha$. For each $i \in [m]$, there exists $\mathbf{x}(i) \in X$ such that $|\,\mathbf{x}(i)^\top M_1 \cdots M_{k_i}\mathbf{a} - 1 - \delta\,| \le \rho$. By the stochasticity of the matrices, $|\,\big(\mathbf{x}(i) - \mathbf{x}\big)^\top M_1 \cdots M_{k_i}\mathbf{a}\,| \le b_0\,\rho$; hence $|\,\mathbf{x}^\top M_1 \cdots M_{k_i}\mathbf{a} - 1 - \delta\,| \le (b_0+1)\rho$. By Lemma 3, there is a rational vector $\mathbf{u}$ such that $\mathbf{1}^\top\mathbf{u} = 1$ and $\mathbf{x}^\top M^{(\theta)}\mathbf{u} = \psi(M^{(\theta)}, \mathbf{a})$ does not depend on the variable $\mathbf{x} \in \mathbb{S}$; on the other hand, $|\,\mathbf{x}^\top M^{(\theta)}\mathbf{u} - (1+\delta)\,| \le b_1\rho$. Two quick remarks: (i) the term $1+\delta$ is derived from $(1+\delta)\mathbf{1}^\top\mathbf{u} = 1+\delta$; (ii) $b_1 \le (b_0+1)\|\mathbf{u}\|_1$, where $\mathbf{u}$ is a rational over $O(\mu T)$ bits. We invalidate the condition on $k_1, \dots, k_m$ by keeping δ outside the interval $\psi(M^{(\theta)}, \mathbf{a}) - 1 + b_1\rho\mathbb{I}$, which rules out at most $2(b_1+1)$ intervals from $\mathcal{D}_\rho$. Repeating this for all sequences $(k_1, \dots, k_m)$ raises the number of forbidden intervals, ie, the size of U, to $c_T = 2^{O(\mu T)}$. ◀

Topological entropy

We identify the family $\mathcal{M}$ with the set of all matrices of the form $S_1 \cdots S_k$ for $\eta \le k \le 3\eta$. By definition of the ergodic renormalizer $\eta = \eta(\Omega)$ (for a set Ω that will be specified later), any $M \in \mathcal{M}$ is primitive and $\tau(M) < 1/2$; furthermore, both μ and $\log|\mathcal{M}\,|$ are in $O(\eta)$. Our next result implies a bound of $\lim_{T\to\infty} T^{-\eta^{-O(1)}} = 0$ on the topological entropy of the shift space of itineraries.

▶ **Lemma 6.** *For any real* $\rho > 0$ *and any integer* $T > 0$*, there exists* $t_\rho = O(\eta|\log\rho\,|)$ *and* $V \subseteq \mathcal{D}_\rho$ *of size* $d_T = 2^{O(T)}$ *such that, for any* $\Delta \in \mathcal{D}_\rho \backslash V$*, any integer* $t \ge t_\rho$*, and any* $\sigma \in L^t_\Delta$, $\log\big|\,\{\,\sigma' \,|\, \sigma\cdot\sigma' \in L^{t+T}_\Delta\,\}\,\big| \le \eta^{b+1}T^{1-\eta^{-b}}$*, for constant* $b > 0$.

Proof. In the lemma, t_ρ (resp. d_T) is independent of T (resp. ρ). The main point is that the exponent of T is bounded away from 1. We define the set V as the union of the sets U formed by applying Lemma 5 to each one of the hyperplanes h_δ involved in $\mathcal{P}$ and every possible sequence of T matrices in $\mathcal{M}$. This increases c_T to $2^{O(\eta T)}$. Fix $\Delta \in \mathcal{D}_\rho \backslash V$ and consider the (lifted) phase space $\mathbb{S} \times \Delta$ for the dynamical system induced by the map $f_\uparrow : \big(\mathbf{x}, \delta\big) \mapsto \big(\,\mathbf{x}^\top S(\mathbf{x}), \delta\,\big)$. The system is piecewise-linear with respect to the polyhedral partition $\mathcal{P}_\uparrow$ of $\mathbb{R}^{n+1}$ formed by treating δ as a variable in h_δ. Let Υ_t be a continuity piece for $f^t_\uparrow$, ie, a maximal region of $\mathbb{S} \times \Delta$ over which the t-th iterate of $f_\uparrow$ is linear. By the argument leading to (1), therefore, any matrix sequence $S_1, \dots, S_t$ matching an element of L^t_Δ is such that $S_1 \cdots S_t = \mathbf{1}\boldsymbol{\pi}^\top + Q$, where $\|Q\|_\infty < 2^{2-t/\eta}$; hence there exists $t_\rho = O(\eta|\log\rho\,|)$ such that, for any $t \ge t_\rho$, $f^t_\uparrow(\Upsilon_t) \subseteq (\mathbf{x} + \rho\mathbb{I}^n) \times \Delta$, for some $\mathbf{x} = \mathbf{x}(t, \Upsilon_t) \in \mathbb{S}$.

Consider a nested sequence $\Upsilon_1 \supseteq \Upsilon_2 \supseteq \cdots$.[25] We say there is a *split* at k if $\Upsilon_{k+1} \subset \Upsilon_k$, and we show that, given any $t \ge t_\rho$, there are only $O(\eta T^{1-\alpha}/\alpha)$ splits between t and $t+\eta T$, where $\alpha = \eta^{-b}$, for constant b (see Lemma 3 for definitions). We may confine our attention to splits caused by the same hyperplane h_δ (since $\mathcal{P}$ features only a constant number of them). Arguing by contradiction, we assume the presence of at least $6\eta T^{1-\alpha}/\alpha$ splits, which implies

[25] Note that Υ_1 is a cell of $\mathcal{P}_\uparrow$, $f^k_\uparrow(\Upsilon_{k+1}) \subseteq f^k_\uparrow(\Upsilon_k)$, and S_l is the stochastic matrix used to map $f^{l-1}_\uparrow(\Upsilon_l)$ to $f^l_\uparrow(\Upsilon_l)$ (ignoring the dimension δ).

that at least $N := 2T^{1-\alpha}/\alpha$ of those splits occur for values of k at least 2η apart. This is best seen by binning $[t+1, t+\eta T]$ into T intervals of length η and observing that at least $3N$ intervals must feature splits. In fact, this proves the existence of N splits at positions separated by a least two consecutive bins. Next, we use the same binning to produce the matrices $M_1, \ldots, M_T$, where $M_j = S_{t+1+(j-1)\eta} \cdots S_{t+j\eta}$.

Suppose that all of the N splits occur for values k of the form $t + j\eta$. In this case, a straightforward application of Lemma 5 is possible: we set $X \times \Delta = f^t_{\uparrow}(\Upsilon_t)$ and note that the functions ϕ^k are all products of matrices from the family $\mathcal{M}$, which happen to be η-long products. The number of splits, $2T^{1-\alpha}/\alpha$, exceeds the number allowed by the lemma and we have a contradiction. If the splits do not fall neatly at endpoints of the bins, we use the fact that $\mathcal{M}$ includes matrix products of any length between η and 3η. This allows us to reconfigure the bins so as to form a sequence $M_1, \ldots, M_T$ with the splits occurring at the endpoints: for each split, merge its own bin with the one to its left and the one to its right (neither of which contains a split) and use the split's position to subdivide the resulting interval into two new bins; we leave all the other bins alone.[26] This leads to the same contradiction, which implies the existence of fewer than $O(\eta T^{1-\alpha}/\alpha)$ splits at $k \in [t, t+\eta T]$; hence the same bound on the number of strict inclusions in the nested sequence $\Upsilon_t \supseteq \cdots \supseteq \Upsilon_{t+\eta T}$. The set of all such sequences forms a tree of depth ηT, where each node has at most a constant number of children and any path from the root has $O(\eta T^{1-\alpha}/\alpha)$ nodes with more than one child. Rescaling T to ηT and raising b completes the proof. ◀

3.2 Proof of Lemma 1

We show that the nonperiodic δ intervals can be covered by a Cantor set of Hausdorff dimension less than one. All the parameters below refer to Lemma 6 and are set in this order: $T(\eta)$, $\rho(T, \varepsilon)$, and $\nu(T, \rho, \varepsilon)$. The details follow. Let δ, Δ such that $\delta \in \Delta \in \mathcal{D}_\rho \backslash V$. Given a continuity piece $C^t \subseteq \mathbb{S}$ for f^t, the $(t+T)$-th iterate of f induces a partition of C^t into a finite number of continuity pieces $C^t_1, \ldots, C^t_m$, so we can define $\lambda_{t,T}(C^t) = \sum_i \operatorname{diam}_{\ell_\infty} f^{t+T}(C^t_i)$. As was observed in the proof of Lemma 6, $\operatorname{diam}_{\ell_\infty} f^{t+T}(C^t_i) = O(2^{-T/\eta} \operatorname{diam}_{\ell_\infty} f^t(C^t))$. That same lemma shows that if we pick $T = 2^{\eta^{2b}}$, for b large enough then, for any $t \geq t_\rho$,

$$\lambda_{t,T}(C^t) = \sum_{i=1}^{m} \operatorname{diam}_{\ell_\infty} f^{t+T}(C^t_i) \leq b 2^{\eta^{b+1} T^{1-\eta^{-b}}} 2^{-T/\eta} \operatorname{diam}_{\ell_\infty} f^t(C^t) \leq \tfrac{1}{2} \operatorname{diam}_{\ell_\infty} f^t(C^t). \quad (3)$$

Next we set $\rho = \varepsilon/(2d_T)$ so that the intervals of V cover a length of at most $\varepsilon/2$. This gives us an extra length of $\varepsilon/2$ worth of forbidden intervals at our disposal. For any $t = t_\rho + kT$ large enough, $f^t(\mathbb{S})$ is the union of (possibly overlapping) convex bodies $K_1, \ldots, K_p$. A key observation is that we can prevent any K_i from splitting at time $t + kT$ by keeping δ outside an interval of length $\operatorname{diam}_{\ell_\infty} K_i$ for each discontinuity of f. By iterating (3), we find that $\lambda_{t_\rho, kT}(C^{t_\rho}) \leq 2^{-k}$. We expand V by adding these intervals, which expands the total length covered by $2^{O(t_\rho)-k}$. To keep this expansion, as stated earlier, below $\varepsilon/2$, we set $k = O(t_\rho) + |\log \varepsilon|$. It follows that $Z_\nu = Z_{\nu+1}$ for $\nu + 1 = t_\rho + kT$, and hence $Z_t = Z$ for any $t \geq \nu$.[27] In view of $T = 2^{\eta^{2b}}$, $d_T = 2^{O(T)}$, $\rho = \varepsilon/(2d_T)$, and $t_\rho = O(\eta |\log \rho|)$, we

[26] We note the possibility of an inconsequential decrease in T caused by the merges. Also, we can now see clearly why Lemma 5 is stated in terms of the slab h_Δ and not the hyperplane h_δ. This allows us to express splitting caused by the hyperplane $\mathbf{a}^\top \mathbf{x} = 1 + \delta$ in lifted space $\mathbb{R}^{n+1}$.

[27] No point $\mathbf{x}$ is such that (a) $f^{\nu+1}(\mathbf{x})$ is in D (the union of the discontinuities) but $f^\nu(\mathbf{x})$ is not. To see why this implies that $Z_{t+1} = Z_t$ for any $t > \nu$, and hence $Z = Z_\nu$, suppose that $Z_{t+1} \supset Z_t$, ie,

observe that $|\log \rho| = |\log \varepsilon| + O(T)$, and both t_ρ and k are in $O(\eta|\log \varepsilon| + \eta T)$; hence

$$\nu = t_\rho + kT = O\big(\eta|\log \varepsilon|T + \eta T^2\big) = 2^{\eta^{O(1)}}|\log \varepsilon|,$$

which proves Lemma 1.

4 Applications

We show how the two sets of techniques developed above, renormalization and bifurcation analysis, allows us to resolve a few important families of *MIS*.

4.1 Irreducible systems

A Markov influence system is called *irreducible* if the Markov chain $g(\mathbf{x})$ is irreducible for all $\mathbf{x} \in \mathbb{S}$; with our self-loop assumption, this also means ergodic. All the digraphs $g(\mathbf{x})$ of an irreducible *MIS* are strongly connected; therefore, in the first instantiation of production (1a) $\mathbf{g} \to \big(\,\mathbf{g}_l\,\big)\, g_m \,\big(\,\mathbf{g}_r \bigtriangleup h\,\big)$, the right-hand side expands into

$$\mathbf{g} \longrightarrow ((((g_1)g_2)\cdots)g_{m-1})g_m\,\big(\,\mathbf{g}_r \bigtriangleup h\,\big), \tag{4}$$

with $h \in \mathbb{K}$ and $m < n$. In other words, every step sees growth in the cumulant until it is in $\mathbb{K}$ (the family of all complete digraphs). To see why, assume by contradiction that $\prod_{j<k} g_j = \prod_{j\le k} g_j$ for $k < m$. If so, then g_k is a subgraph of $tf(\prod_{j<k} g_j)$. Because the latter is transitive and it has in g_k a strongly connected subgraph that spans all the vertices, it must belong to $\mathbb{K}$; hence $k = m$, which contradicts our assumption. Since the last cumulant is in $\mathbb{K}$, the parsing of $\mathbf{g}_r \bigtriangleup h$ in (4) proceeds via (1b); hence

$$\begin{aligned}\mathbf{g} \longrightarrow\; & ((((g_1)g_2)\cdots)g_{m_1-1})\,g_{m_1} \\ & \Big(((((g_{m_1+1})g_{m_1+2})\cdots)g_{m_2-1})g_{m_2}\,((((g_{m_2+1})g_{m_2+2})\cdots)g_{m_3-1})g_{m_3}\,\cdots\Big).\end{aligned} \tag{5}$$

It follows that $\eta(\mathbb{R})$ is polynomial in n. By Lemma 1 and Corollary 2, this shows that irreducible Markov influence systems are typically asymptotically periodic. We strengthen this result in our next application.

4.2 Weakly irreducible systems

We now assume a fixed partition of the vertices such that each digraph $g(\mathbf{x})$ consists of disjoint strongly connected graphs defined over the subsets $V_1, \ldots, V_l$ of the partition. Irreducible systems correspond to the case $l = 1$. What makes weak irreducibility interesting is that the systems are not simply the union of independent irreducible systems. Indeed, note that communication flows among states in two ways: (i) directly, vertices collect information from neighbors to update their states; and (ii) indirectly, via the polyhedral partition $\mathcal{P}$, the sequence of graphs for V_i may be determined by the current states within V_j. In the extreme case, we can have the co-evolution of two systems V_1 and V_2, each one depending entirely on the other one yet with no links between the two of them. If the two subsystems were

that $f^{t+1}(\mathbf{y})$ is in D but $f^t(\mathbf{y})$ is not, for $\mathbf{y} \in \mathbb{S}$; then (a) holds for $\mathbf{x} = f^{t-\nu}(\mathbf{y})$, a contradiction. This shows that the continuity pieces for f^ν are the same as for any $f^{\nu+k}$, which implies that the f-image of any such piece must fall entirely inside a single one of them. The eventual periodicity of the itinerary follows.

independent, their joint dynamics could be expressed as a direct sum and resolved separately. This cannot be done, in general, and the bifurcation analysis requires some modifications.

As before, the right-hand side of production (1a) expands into (4). The difference is that h is now a collection of disjoint complete digraphs $h_1, \ldots, h_l$, one for each V_i. This gives us an opportunity to use network renormalization (2a) to derive

$$\mathbf{g} \triangle h \longrightarrow \big(\mathbf{g}_{\,|\, h_1} \triangle h_1 \big) \parallel \cdots \parallel \big(\mathbf{g}_{\,|\, h_l} \triangle h_l \big) \,.$$

Each $\mathbf{g}_{\,|\, h_i} \triangle h_i$ is parsed as in (1b) into $\big(\mathbf{g}^1_{\,|\, h_i} \big)\, g^{m_1}_{\,|\, h_i} \big(\mathbf{g}^2_{\,|\, h_i} \big)\, g^{m_2}_{\,|\, h_i} \cdots$ (with indices moved up for clarity). For the bifurcation analysis, Lemma 4 relies on the rank-l expansion

$$\mathbf{x}^\top M^{(\theta)} = \sum_{i=1}^{l} \Big(\sum_{j \in V_i} x_j \Big)\, \mathbf{a}^\top \Pi_i^{(\theta)} + \mathbf{x}^\top Q^{(\theta)} \,,$$

where (i) $M_1 \cdots M_k = \sum_{i=1}^{l} \mathbf{1}_{|V_i} \boldsymbol{\pi}_{i,k}^\top + Q_k$; (ii) all vectors $\mathbf{1}_{|V_i}$ and $\boldsymbol{\pi}_{i,k}$ have support in V_i; (iii) Q_k is block-diagonal and $\|Q_k\|_\infty < 2^{1-k}$; (iv) $\Pi_i^{(\theta)}$ and $Q^{(\theta)}$ are formed, respectively, by the column vectors $\boldsymbol{\pi}_{i,k_j}$ and $Q_{k_j}\, \mathbf{a}$ for $j \in [m]$, with $\theta = (k_1, \ldots, k_m)$. Property **U** no longer holds, however (see §3.1): indeed, if $l > 1$, it is no longer true that $\mathbf{x}^\top M^{(\theta)} \mathbf{u}$ is independent of the variable $\mathbf{x} \in \mathbb{S}$. The dependency is confined to the sums $s_i := \sum_{j \in V_i} x_j$ for $i \in [l]$. The key observation is that these sums are time-invariant. We fix them once and for all and redefine the phase space as the invariant manifold $\prod_{i=1}^{l} \big(s_i\, \mathbb{S}^{|V_i|-1} \big)$, which induces a foliation of the original simplex $\mathbb{S}^{n-1}$ via $\mathbb{S}^{l-1}$. The rest of the proof mimics the irreducible case, whose conclusion therefore still applies.

4.3 Condensation systems

We now assume that the condensations of the $g(\mathbf{x})$ all share the same transitive reduction. In other words, the condensations may change with time but all of them feature the same pairs of path-connected vertices.[28] Past the first n steps, a temporal walk will have been set up joining all pairs of path-connected vertices. We ignore the possibility of a call to (2a), which would be handled as in the weakly irreducible case. The parse tree features a node labeled (2b), where the cumulant h is the common transitive closure of any $g(\mathbf{x})$:

$$\mathbf{g} \triangle h \longrightarrow \big(\mathbf{g}_{\,|\, h'} \triangle h' \big) \;\Vdash\; \Big\{ \; \big(\mathbf{g}_{\,|\, h_1} \triangle h_1 \big) \parallel \cdots \parallel \big(\mathbf{g}_{\,|\, h_l} \triangle h_l \big) \; \Big\} .$$

It helps to think of the right-hand side of $\Vdash$ as the absorbing states of a time-varying Markov chain. Every n steps, another temporal walk is established to match any path in h. Let $\sum_{h'} x_i$ be the sum of the probabilities at the nodes of the stem h'. It is easy to see that, after every interval of n steps, $\sum_{h'} x_i$ is multiplied by less than $1 - O(1)^{-n} < 1/c$ for constant $c > 1$. We are now ready to reduce the problem to the case of weakly irreducible systems. Indeed, the sums s_i approach a fixed value with an additive error rate of $c^{-t/n}$, which is fast enough to keep the previous analysis valid. We omit the technical details, which simply recycle the reasoning from the previous sections.

The term "typically" below means almost everywhere, ie, for δ in a subset of $\frac{1}{2}[-1,1]$ of full Lebesgue measure 1. Recall that the result below applies, de facto, to irreducible and weakly irreducible systems.

---

[28] It is is implicit that the vertices of the condensations must match the same set of vertices. Note that the condensations have the same transitive closure and that any system with a time-invariant condensation forms a condensation *MIS*.

▶ **Theorem 7.** *Typically, every orbit of a condensation Markov influence system is asymptotically periodic.*

5 Hyper-Torpid Mixing and Chaos

Among the *MIS* that converge to a single stationary distribution, we show that the mixing time can be super-exponential. The creation of new timescales is what most distinguishes *MIS* from standard Markov chains. As we mentioned earlier, the system can be chaotic. We prove all of these claims below.

5.1 A super-exponential mixer

How can reaching a fixed point distribution take so long? Before we answer this question formally, we provide a bit of intuition. Imagine having three unit-volume water reservoirs A, B, C alongside a clock that rings at noon and 1pm every day. Initially, the clock is at 2pm and A is full while B and C are empty. Reservoir A transfers half of its contents to B and repeats this each hour until the clock rings noon. At this point, reservoir A empties into C the little water that it has left. Next, the clock now rings 1pm and B empties its contents into A. At 2pm, we resume what we did the day before at the same hour, ie, A transfers half of its water contents to B, etc. This goes on until some day, at 1pm, reservoir C finds its more than half full. (Note that the water level of C rises by about 10^{-3} every day.) At this point, both B and C transfer all their water back to A, so that at 2pm on that day, we are back to square 1. The original 12-step clock has been extended into a new clock of period roughly 1,000. The proof below shows how to simulate this iterative process with an *MIS*.

▶ **Theorem 8.** *There exist Markov influence systems that mix to a stationary distribution in time equal to a tower-of-twos of height linear in the number of states.*

Proof. We construct an *MIS* with a periodic orbit of length equal to a power-of-twos. It is easy to turn it into an orbit with a fixed-point attractor that reaches a stationary distribution and we omit this part of the discussion. Assume, by induction, that we have a Markov influence system M cycling through states $1, \ldots, p$, for $p > 1$. We build another one with period c^p, for fixed $c > 1$, by adding a “gadget" to it consisting of a three-vertex graph $1, 2, 3$ with probability distribution $(x, y, z) \in \mathbb{S}$. We initialize the system by placing M in state 1 (ie, 2pm in our clock example) and setting $x = 1$. The dynamic graph is specified by these rules:

1. Suppose that $z > 1/2$. If M is in state p, then the graph has the edges $(2, 1)$ and $(3, 1)$; both are given probability 1 (so that nodes 2, 3 have no self-loops). If M is in any other state, the graph has only three self-loops, each one assigned probability 1.
2. Suppose that $z \leq 1/2$.
   a. If M is in state $1, \ldots, p-2$, then the graph has the edge $(1, 2)$, which is assigned probability $1/2$, as is the self-loop at 1.
   b. If M is in state $p-1$, then the graph has the edge $(1, 3)$, which is assigned probability 1 (hence no self-loop at 1).
   c. If M is in state p, then the graph has the edge $(2, 1)$, which is assigned probability 1 (hence no self-loop at 2).

Suppose that M is in state 1 and that $y = 0$ and $z \leq 1/2$. When M reaches state $p-1$, the probability vector is $\big((1-z)2^{2-p}, (1-z)(1-2^{2-p}), z \big)$. At the next step, M is in state p and the vector becomes $\big(0, (1-z)(1-2^{2-p}), z + (1-z)2^{2-p} \big)$. If the last coordinate is still at most $1/2$, then M moves to state 1 and the vector becomes $\big((1-z)(1-2^{2-p}), 0, z + (1-z)2^{2-p} \big)$. The key observation is that z increases by $(1-z)2^{2-p}$, which is between 2^{1-p} and 2^{2-p} as long as $z \le 1/2$. Since z begins at 0, it will cross the threshold 1/2 after on the order of $p2^p$ steps. Transfers of mass to vertex 3 only happens when M is in state $p-1$, so at the next step, M is in state p and $z > 1/2$. The system moves to state 1 and restores its initial vector $(1, 0, 0)$: the cycle is closed. The construction on top of M adds three new vertices so we can push this recursion roughly $n/3$ times to produce a Markov influence system that is periodic with a period of length equal to a tower-of-twos of height roughly $n/3$. We tie up the loose ends now:

- The construction needs to recognize two consecutive states of M: they are labeled $p-1$ and p in our description but, by symmetry, they could be any other consecutive pair. We give each one of these two states their own distinct polyhedral cell. The obvious choice is the unique state satisfying $z > 1/2$, which is followed by the only state such that $x \ge 1$.
- The basis case of our inductive construction consists of a two-vertex system of constant period k with initial distribution $(1, 0)$. If $x > 2^{1-k}$, the graph has an edge from 1 to 2 assigned probability 1/2; else an edge from 2 to 1 given probability 1 to reset the system.
- The construction assumes probabilities summing up to 1 within each of $\lfloor (n-2)/3 \rfloor + 1$ gadgets, which is clearly wrong. Being piecewise-linear, however, the system suggests an easy fix: we divide the probability weights equally among each gadget and adjust the linear discontinuities appropriately. ◀

5.2 Chaos

We give a simple 5-state construction with chaotic symbolic dynamics:

$$A = \frac{1}{3}\begin{pmatrix} 2 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} \text{ if } x_1 + x_2 > x_4 \quad \text{and} \quad B = \frac{1}{3}\begin{pmatrix} 1 & 0 & 2 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix} \text{ else,}$$

for $\mathbf{x} \in \mathbb{S}^4$. We focus our attention on $\Sigma = \{\, (x_1, x_2, x_4) \mid 0 < x_1 \le x_4/2 \text{ and } x_4/2 \le x_2 < x_4 \,\}$, and easily check that it is an invariant manifold. At time 0, we fix $x_4 = 1/4$ and $x_5 = 0$; at all times, of course, $x_3 = 1 - x_1 - x_2 - x_4 - x_5$. The variable $y := (2x_2 - x_4)/(2x_1 - x_4)$ is always nonpositive over Σ.[29] It evolves as follows:

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

Writing $z = (y+1)/(y-1)$, we note that $-1 \le z < 1$ and it evolves according to $z \mapsto 2z + 1$ if $z \le 0$, and $z \mapsto 2z - 1$ otherwise, a map that conjugates with the baker's map and is well known to be chaotic [11].

Acknowledgments. I wish to thank Maria Chudnovsky and Ramon van Handel for helpful comments.

References

1 C. Avin, M. Koucký, and Z. Lotker. How to explore a fast-changing world (cover time of a simple random walk on evolving graphs). *Proc. 35th ICALP*, pages 121–132, 2008.

[29] We used a different but similar construction in [4].

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

3 C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, 2009.

4 B. Chazelle. Diffusive influence systems. *SIAM J. Comput.*, 44:1403–1442, 2015.

5 B. Chazelle, Q. Jiu, Q. Li, and C. Wang. Well-posedness of the limiting equation of a noisy consensus model in opinion dynamics. *J. Differential Equations*, 263:365–397, 2017.

6 A. Condon and D. Hernek. Random walks on colored graphs. *Random Structures and Algorithms*, 5:285–303, 1994.

7 A. Condon and R.J. Lipton. On the complexity of space bounded interactive proofs. *Proc. 30th IEEE Symp. on Foundations of Computer Science*, pages 462–267, 1989.

8 D. Coppersmith and P. Diaconis. Random walk with reinforcement. *Unpublished manuscript*, 1986.

9 O. Denysyuk and L. Rodrigues. Random walks on directed dynamic graphs. *Proc. 2nd International Workshop on Dynamic Networks: Algorithms and Security (DYNAS'10), Bordeaux, France (Also arXiv:1101.5944 (2011))*, 2010.

10 O. Denysyuk and L. Rodrigues. Random walks on evolving graphs with recurring topologies. *Proc. 28th International Symposium on Distributed Computing (DISC)*, 2014.

11 R.L. Devaney. An Introduction to Chaotic Dynamical Systems (2nd ed.). *Westview Press*, 2003.

12 P. Diaconis. Recent progress on de Finetti's notions of exchangeability. *Bayesian Statistics*, 3, 1988.

13 T.D. Frank. Strongly nonlinear stochastic processes in physics and the life sciences. *ISRN Mathematical Physics*, Article ID 149169, 2013.

14 P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519:97–125, 2012.

15 G. Iacobelli and D.R. Figueiredo. Edge-attractor random walks on dynamic networks. *J. Complex Networks*, 5:84–110, 2017.

16 A. Katok and B. Hasselblatt. Introduction to the Modern Theory of Dynamical Systems. *Cambridge University Press*, 1996.

17 D. Kempe, J. Kleinberg, and A. Kumar. Connectivity and inference problems for temporal networks. *J. Computer and System Sciences*, 64:820–842, 2002.

18 V.N. Kolokoltsov. Nonlinear Markov Processes and Kinetic Equations. *Cambridge Tracks in Mathematics (182)*, Cambridge Univ. Press, 2010.

19 M. Othon. An introduction to temporal graphs: an algorithmic perspective. *Internet Mathematics*, 12, 2016.

20 E. Seneta. Non-Negative Matrices and Markov Chains. *Springer, 2nd ed.*, 2006.

21 S. Sternberg. Dynamical Systems. *Dover Books on Mathematics*, 2010.