[Proceedings of 22nd International Symposium on Fundamentals of Computation Theory (2019), 18-28]

Some Observations on Dynamic Random Walks and Network Renormalization*

Bernard Chazelle[1 [0000−0001−8542−0247]]

Department of Computer Science, Princeton University, Princeton NJ 08540, USA
chazelle@cs.princeton.edu

Abstract. We recently developed a general bifurcation analysis framework for establishing the periodicity of certain time-varying random walks. In this work, we look at the special case of lazy uniform-inflow random walks and show how a much simpler version of the argument can be used to resolve their analysis. We also revisit a renormalization technique for network sequences that we introduced earlier and we propose a few simplifications. This work can be viewed as a gentle introduction to *Markov influence systems*.

Keywords: Markov influence systems · Dynamic random walks · Network renormalization.

1 Introduction

Markov chains have remarkably simple dynamics: they either mix toward a stationary distribution or oscillate periodically. The periodic regime can be easily ruled out by introducing self-loops; thus, from a dynamical-systems perspective, Markov chains are essentially trivial. Not so with time-varying Markov chains [1, 4–12, 14]. We recently introduced *Markov influence systems* (*MIS*) to model random walks in graphs whose transition probabilities and topologies change over time endogenously [3]. The presence of a feedback loop, through which the next graph is chosen as a function of the current distribution of the walk, plays a crucial role. Indeed, the dynamics ranges over the entire spectrum from fixed-point attraction to chaos. This stands in sharp contrast to not only classical Markov chains but also time-varying chains whose temporal changes are driven randomly [1].

We showed that, if the Markov chains used at each step are irreducible, then the *MIS* is almost surely asymptotically periodic [3]. We prove a similar result in the next section, but for a different family of random walks, called *uniform-inflow*. Though the proof borrows much of the architecture of our previous one, it is much simpler and can be viewed as a gentle introduction to Markov influence systems.

* 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.

The main weakness of our bifurcation analysis is to impose topological restrictions on the graphs. As a step toward overcoming this limitation, we have developed a *renormalization* technique for graph sequences [3]. The motivation was to extend the standard classification of Markov chain states to the time-varying case. We revisit this technique in §3 and propose a number of useful simplifications.

2 Time-Varying Random Walks

Recall the definition of a Markov influence system [3]. Let $\mathbb{S}^{n-1}$ be the probability 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 row-stochastic matrices. An *MIS* is a discrete-time dynamical system with phase space $\mathbb{S}^{n-1}$, which is defined by the map $f : \mathbf{x}^\top \mapsto f(\mathbf{x}) := \mathbf{x}^\top S(\mathbf{x})$, where $\mathbf{x} \in \mathbb{S}^{n-1}$ and S is a piecewise-constant function $\mathbb{S}^{n-1} \mapsto \mathcal{S}$ over the cells $\{C_k\}$ of a hyperplane arrangement $\mathcal{H}$ within $\mathbb{S}^{n-1}$ (fig.1); over the discontinuities $h \cap \mathbb{S}^{n-1}$ ($h \in \mathcal{H}$), we define f as the identity.[1]

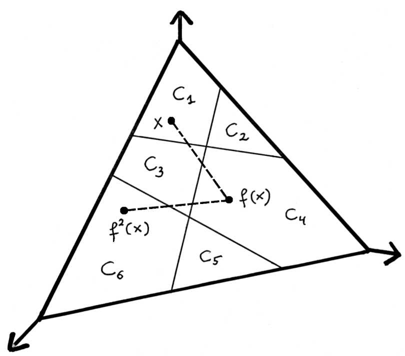

Fig. 1. The arrangement $\mathcal{H}$ consists of three hyperplanes. Each cell C_i in the simplex $\mathbb{S}^{n-1}$ is associated with a stochastic matrix defining the map f over it. The figure shows the first two iterates of $\mathbf{x}$ under f. The case $|\mathcal{H}| = 0$ corresponds to an ordinary random walk.

We focus our attention on *lazy uniform-inflow* random walks: each matrix $S(\mathbf{x})$ is associated with a probability distribution $(p_0(\mathbf{x}), \ldots, p_n(\mathbf{x})) \in \mathbb{S}^n$, such that $S(\mathbf{x})_{ij} = p_j(\mathbf{x}) + \delta_{ij} p_0(\mathbf{x})$ and δ_{ij} is the Kronecker delta. The cases $p_0(\mathbf{x}) = 0, 1$ are both trivial, so we may assume that $0 < p_0(\mathbf{x}) < 1$. Thus, any given $S(\mathbf{x})$ is the transition matrix of a lazy random walk. We state our main result:

Theorem 1. *Every orbit of a lazy uniform-inflow Markov influence system is almost surely asymptotically periodic.*

Note that lazy uniform-inflow random walks are not necessarily irreducible, so the theorem does not follow from [3].[2] In the remainder of this section, we discuss the

[1] The discontinuities can also be chosen to be real-algebraic varieties.

[2] For example, the lazy random walk specified by the matrix $\left(\begin{smallmatrix} 1 & 0 \\ 0.5 & 0.5 \end{smallmatrix}\right)$ is not irreducible. Also, unlike in [3], we do not require the matrices to be rational.

meaning of the result and then we prove it. The *orbit* of $\mathbf{x} \in \mathbb{S}^{n-1}$ is the infinite sequence $(f^t(\mathbf{x}))_{t\geq 0}$ and its *itinerary* is the corresponding sequence of cells C_i's 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. The discontinuities in the map f occur at the intersections of the simplex $\mathbb{S}^{n-1}$ with the hyperplanes $\{\mathbf{x} \in \mathbb{R}^n \,|\, \mathbf{a}_i^\top \mathbf{x} = 1\}$ of $\mathcal{H}$. The hyperplanes are perturbed into the form $\mathbf{a}_i^\top \mathbf{x} = 1 + \delta$, for $\delta \in \Omega = [-\omega, \omega]$ and $\omega > 0$. Assuming that $\mathcal{H}$ is in general position, ω can always be chosen small enough so that the perturbed arrangement remains topologically invariant over all $\delta \in \Omega$. Theorem 1 follows from the existence of a set of Lebesgue measure zero (coverable by a Cantor set of Hausdorff dimension less than one) such that, for any $\delta \in \Omega$ outside of it, there is a finite set of stable periodic orbits (ie, discrete limit cycles) such that every orbit is asymptotically attracted to one of them.

It is useful to begin with a few observations about the stochastic matrices involved in lazy uniform-inflow random walks:

1. The matrix $S(\mathbf{x})$ can be written as
$$p_0(\mathbf{x})I + \mathbf{1}(p_1(\mathbf{x}), \ldots, p_n(\mathbf{x})) \tag{1}$$
and it has the unique stationary distribution $\pi(\mathbf{x}) = \frac{1}{1-p_0(\mathbf{x})}(p_1(\mathbf{x}), \ldots, p_n(\mathbf{x}))$. The family of such matrices is closed under composition. Indeed,
$$S(\mathbf{x})S(\mathbf{y}) = q_0 I + \mathbf{1}(q_1, \ldots, q_n),$$
where $q_0 = p_0(\mathbf{x})p_0(\mathbf{y})$ and, for $i > 0$, $q_i = p_i(\mathbf{x})p_0(\mathbf{y}) + p_i(\mathbf{y})$.
2. Let $\mathcal{M}$ be the (finite) set of all the matrices $S(\mathbf{x})$ that arise in the definition of f. We just saw that $p_0(\mathbf{x})$ is multiplicative. In this case, it is equal to the coefficient of ergodicity of the matrix [13], which is defined as half the maximum ℓ_1-distance between any two of its rows. By our assumption, $\tau := \sup_{\mathbf{x}} p_0(\mathbf{x}) < 1$. Given $M_1, \ldots, M_k \in \mathcal{M}$, if $\boldsymbol{\pi}$ denotes the stationary distribution of $M_1 \cdots M_k$, then
$$\begin{cases} M_1 \cdots M_k = qI + (1-q)\mathbf{1}\boldsymbol{\pi}^\top \\ \mathrm{diam}_{\ell_\infty}(\mathbb{S}^{n-1} M_1 \cdots M_k) = q \leq \tau^k. \end{cases} \tag{2}$$

Let D be the union of the discontinuities (defined by the intersection of the perturbed hyperplanes with the simplex), for some fixed $\delta \in \Omega$. Put $Z_t = \bigcup_{0\leq k\leq t} f^{-k}(D)$ and $Z = \bigcup_{t\geq 0} Z_t$ and note that $Z_\nu = Z_{\nu-1}$ implies that $Z = Z_\nu$. Indeed, suppose that $Z_{t+1} \supset Z_t$ for $t \geq \nu$; then, $f^{t+1}(\mathbf{y}) \in D$ but $f^t(\mathbf{y}) \notin D$ for some $\mathbf{y} \in \mathbb{S}^{n-1}$; in other words, $f^\nu(\mathbf{x}) \in D$ but $f^{\nu-1}(\mathbf{x}) \notin D$ for $\mathbf{x} = f^{t-\nu+1}(\mathbf{y})$, which contradicts the equality $Z_\nu = Z_{\nu-1}$. The key to periodicity is to prove that ν is finite.[3]

Lemma 1. *There is a constant $c > 0$ such that, for any $\varepsilon > 0$, there exist an integer $\nu \leq c\log(1/\varepsilon)$ and a finite union K of intervals of total length at most ε such that $Z_\nu = Z_{\nu-1}$, for any $\delta \in \Omega \setminus K$.*

[3] The constants in this paper may depend on any of the input parameters, such as the dimension n, the number of hyperplanes, the hyperplane coefficients, and the matrix elements.

Proof of Theorem 1. The theorem can be shown to follow from Lemma 1 by using an argument from [3]. We reproduce the proof here for the sake of completeness. The polyhedral cells defined by the connected components of the complement of $Z = Z_\nu$ form the continuity pieces of $f^{\nu+1}$: by continuity, each one of them maps, under f, not simply to within a single cell of D but actually to within a single cell of Z itself.[4] This in turn implies the eventual periodicity of the symbolic dynamics. 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 the stochastic matrix of a lazy, uniform-inflow random walk, which implies convergence to a periodic point. In fact, better than that, we know from (2) that the matrix corresponds to a random walk that mixes to a unique stationary distribution, so the attracting periodic orbits are stable and there are only a finite number of them.

To complete the proof, we 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 $K^\infty = \bigcap_{l>0} K^l$; then $\mathrm{Leb}(K^l) \leq 2^{1-l}$ and hence $\mathrm{Leb}(K^\infty) = 0$. Theorem 1 follows from the fact that any $\delta \in \Omega$ outside of K^∞ lies outside of K^l for some $l > 0$. $\square$

As in [3], we begin the proof of Lemma 1 with a discussion of the symbolic dynamics of the system. Given $\Delta \subseteq \Omega$, let L^t_Δ denote the set of t-long prefixes of any itinerary for any starting position $\mathbf{x} \in \mathbb{S}^{n-1}$ and any $\delta \in \Delta$. Fix $\rho > 0$ and define $\mathcal{D}_\rho = \{\, [k\rho, (k+1)\rho] \cap \Omega \,|\, k \in \mathbb{Z} \,\}$.

Lemma 2. *There is a constant $b > 0$ such that, for any real $\rho > 0$ and any integer $T > 0$, there exist $t_\rho \leq b \log(1/\rho)$ and $V \subseteq \mathcal{D}_\rho$ of size at most b^T such that, for any $\Delta \in \mathcal{D}_\rho \setminus V$, any integer $t \geq t_\rho$, and any $\sigma \in L^t_\Delta$, we have $\big|\{\, \sigma' \,|\, \sigma \cdot \sigma' \in L^{t+T}_\Delta \,\}\big| \leq b$.*

Proof. Given $M_1, \ldots, M_k \in \mathcal{M}$, let $\varphi^k(\mathbf{x}) = \mathbf{x}^\top M_1 \cdots M_k$ for $\mathbf{x} \in \mathbb{R}^n$ and $k \leq T$; and let $h_\delta \colon \mathbf{a}^\top \mathbf{x} = 1 + \delta$ be some hyperplane in $\mathbb{R}^n$. Let $h_\Delta := \bigcup_{\delta \in \Delta} h_\delta$ and $X = \mathbf{x} + \rho[-1, 1]^n$, for $\mathbf{x} \in \mathbb{S}^{n-1}$. We define an exclusion zone U outside of which the T iterated images of X can meet h_Δ at most once. This is a general position claim much stronger than the one we used in [3] and closer in spirit to a dimensionality argument for planar contractions from [2] that inspired our approach.

Claim A. *For some constant $d > 0$ (independent of ρ), there exists $U \subseteq \mathcal{D}_\rho$ of size at most dT^2 such that, for any $\Delta \in \mathcal{D}_\rho \setminus U$ and $\mathbf{x} \in \mathbb{S}^{n-1}$, there are at most one integer $k \leq T$ such that $\varphi^k(X) \cap h_\Delta \neq \emptyset$.*

Proof. The crux of the claim is that it holds for *any* probability distribution $\mathbf{x}$. We assume the existence of two integers $j < k \leq T$ such that $\varphi^i(X) \cap h_\Delta \neq \emptyset$, for $i = j, k$ and $\Delta \in \mathcal{D}_\rho$. We draw the consequences and then negate them in order to rule out the assumption: this, in turn, specifies the set U. The assumption implies the existence of δ and $\mathbf{x} \in \mathbb{S}^{n-1}$ such that $|(\mathbf{x} + \mathbf{u})^T M_1 \cdots M_j \mathbf{a} - (1 + \delta)| \leq d_o \rho$, with $\|\mathbf{u}\|_\infty \leq \rho$ and constant

[4] Indeed, if that were not the case, then some $\mathbf{x}$ in a cell of Z, thus outside of Z, would be such that $f(\mathbf{x}) \in Z = Z_{\nu-1}$. It would follow that $f^k(\mathbf{x}) \in D$, for $k \leq \nu$; hence $\mathbf{x} \in Z_\nu = Z$, a contradiction.

$d_o > 0$. Likewise, we have $|(\mathbf{x}+\mathbf{u}')^T M_1 \cdots M_k \mathbf{a} - (1+\delta)| \le d_o \rho$, with $\|\mathbf{u}'\|_\infty \le \rho$. Writing $\mathbf{v} = \mathbf{x}^T M_1 \cdots M_j$, we have $|\mathbf{v}^T \mathbf{a} - (1+\delta)| \le d'_o \rho$ and $|\mathbf{v}^T M_{j+1} \cdots M_k \mathbf{a} - (1+\delta)| \le d'_o \rho$, for constant d'_o dependent on $\mathbf{a}$. By (1), $M_{j+1} \cdots M_k = q_0 I + \mathbf{1}(q_1, \ldots, q_n)$, for some $(q_0, \ldots, q_n) \in \mathbb{S}^n$. Since $\mathbf{v} \in \mathbb{S}^{n-1}$, it follows that

$$\Big| q_0 \mathbf{v}^T \mathbf{a} + (q_1, \ldots, q_n)\mathbf{a} - (1+\delta) \Big| \le d'_o \rho;$$

hence $|\delta + 1 - (q_1, \ldots, q_n)\mathbf{a}/(1-q_0))| \le 2d'_o \rho/(1-\tau)$. To rule out the previous condition, we must keep δ outside of $O(1/(1-\tau))$ intervals of $\mathcal{D}_\rho$. The claim follows from the fact that the number of products $M_{j+1} \cdots M_k$ is quadratic in T. □

To complete the proof of Lemma 2, we define V as the union of the sets U formed by applying Claim A to each one of the hyperplanes h_δ of $\mathcal{H}$ and every possible sequence of T matrices in $\mathcal{M}$; hence $|V| \le b^T$ for constant $b > 0$. We fix $\Delta \in \mathcal{D}_\rho \setminus V$ and consider the (lifted) phase space $\mathbb{S} \times \Delta$ for the dynamical system induced by the map $f_\uparrow : (\mathbf{x}^\top, \delta) \mapsto (\,\mathbf{x}^\top S(\mathbf{x}), \delta\,)$. A continuity piece Υ_t for $f_\uparrow^t$ is a maximal polyhedron within $\mathbb{S}^{n-1} \times \Delta$ over which the t-th iterate of $f_\uparrow$ is linear.

Given any sequence $M_1, \ldots, M_k$ in $\mathcal{M}$, recall from (2) that $\mathrm{diam}_{\ell_\infty}(\mathbb{S}^{n-1} M_1 \cdots M_k) \le \tau^k$. This implies the existence of an integer $t_\rho \le b \log(1/\rho)$ (raising the previous constant b if necessary) such that, for any $t \ge t_\rho$, $f_\uparrow^t(\Upsilon_t) \subseteq (\mathbf{x} + \rho \mathbb{I}^n) \times \Delta$, for some $\mathbf{x} = \mathbf{x}(t, \Upsilon_t) \in \mathbb{S}^{n-1}$. Consider a nested sequence $\Upsilon_1 \supseteq \Upsilon_2 \supseteq \cdots$. Note that Υ_1 is a polyhedral cell within $\mathbb{S}^{n-1} \times \Delta$ and $f_\uparrow^k(\Upsilon_{k+1}) \subseteq f_\uparrow^k(\Upsilon_k)$. There is a *split* at k if $\Upsilon_{k+1} \subset \Upsilon_k$. Observe that, by Claim A, given any $t \ge t_\rho$, there are at most a constant number b_1 of splits between t and $t+T$ (at most one per hyperplane of $\mathcal{H}$). It follows that the number of nested sequences is bounded by the number of leaves in a tree of height T with at most b_1 nodes of degree greater than 1 along any path. Lemma 2 follows from the fact that no node has more than a constant number of children. □

Proof of Lemma 1. We use the notation of Lemma 2 and set T to a large enough constant. Fix $\varepsilon > 0$ and set $\rho = \frac{1}{2}\varepsilon/|V|$, and $\nu = t_\rho + kT$, where $k = T\log(1/\varepsilon)$. Since $t_\rho \le b\log(1/\rho)$, note that $\nu = O(\log 1/\varepsilon)$. Let $P = M_1 \cdots M_\nu$, where $M_1, \ldots, M_\nu$ is the matrix sequence matching an element of L_Δ^ν, for $\Delta \in \mathcal{D}_\rho \setminus V$. By (2), $\mathrm{diam}_{\ell_\infty}(\mathbb{S}^{n-1} P) \le \tau^\nu$, so there is a point $\mathbf{x}_P$ such that, given any point $\mathbf{y} \in \mathbb{S}^{n-1}$ whose ν-th iterate $f^\nu(\mathbf{y}) = \mathbf{z}^T$ is specified by $\mathbf{z}^T = \mathbf{y}^T P$, we have $\|\mathbf{x}_P - \mathbf{z}\|_\infty \le \tau^\nu$. Given a discontinuity $h_\delta : \mathbf{a}_i^\top \mathbf{x} = 1 + \delta$ of the system, the point $\mathbf{z}$ lies on one side of h_δ if and only if $\mathbf{x}_P$ lies on the (relevant) side of some $h_{\delta'}$, for $|\delta' - \delta| \le c_1 \tau^\nu$, for constant $c_1 > 0$. Thus, adding an interval of length $c_2 \tau^\nu$ to V, for constant c_2 large enough (independent of T), it is the case that, for any $h \in \mathcal{H}$, it holds that, for all $\mathbf{y} \in \mathbb{S}^{n-1}$, the ν-th iterates $f^\nu(\mathbf{y})$ specified by P all lie strictly on the same side of h_δ, for any $\delta \in \Delta$. We repeat this operation for every string L_Δ^ν and each one of the (at most) $1/\rho$ intervals $\Delta \in \mathcal{D}_\rho \setminus V$. This increases the length $\mathrm{Leb}(V)$ covered by V from its original $\rho|V| = \varepsilon/2$ to $\rho|V| + c_2 |L_\Delta^\nu| \tau^\nu/\rho \le \varepsilon$. This last inequality follows from:

$$
\begin{aligned}
(\rho\varepsilon)^{-1}|L^{\nu}_{\Delta}| &\le (\rho\varepsilon)^{-1}c_3^{t_\rho}b^k && [\text{ for constant } c_3 \text{ independent of } T\,] \\
&\le 2c_3^{t_\rho}b^{k+T}4^{k/T} && [\ 1/\rho\varepsilon \le 2b^T/\varepsilon^2 \le 2b^T 4^{k/T}\ ] \\
&\le 2c_3^{b^2k}b^{k+T}4^{k/T} && [\ t_\rho \le b\log(1/\rho) \le b^2 T\log(1/\varepsilon)\ ] \\
&\le T^k \le \tau^{-\nu}/(2c_2). && [\text{ for } T \text{ large enough }]
\end{aligned}
$$

Thus, for any $\delta \in \Omega$ outside a set of intervals covering a length at most ε, no $f^{\nu}(\mathbf{x})$ lies on a discontinuity. It follows that, for any such δ, we have $Z_\nu = Z_{\nu-1}$. □

This completes the proof of Theorem 1.

3 Revisiting Network Sequence Renormalization

In [3], we proposed a mechanism for expressing an infinite sequence of networks as a hierarchy of graph clusters. The intention was to generalize to the time-varying case the standard classification of the states of a Markov chain. We review the main parts of this "renormalization" technique and propose a number of simplifications along the way. Our variant maintains the basic division of the renormalization process into temporal and topological parts, but it simplifies the overall procedure. For example, the new grammar includes only three productions, as opposed to four.

Throughout this discussion, a *digraph* is a directed graph with vertices in $[n] := \{1,\ldots,n\}$ and a self-loop attached to each vertex. Graphs and digraphs (words we use interchangeably) have no multiple edges. A *digraph sequence* $\mathbf{g} = (g_k)_{k>0}$ is an ordered (possibly infinite) list of digraphs over the same vertex set $[n]$. We define the product $g_i \times g_j$ as the digraph consisting of all the edges (x,y) with an edge (x,z) in g_i and another one (z,y) in g_j for at least one vertex z. The operation $\times$ is associative but not commutative; it corresponds roughly to matrix multiplication. The digraph $\prod_{\le k}\mathbf{g} = g_1 \times \cdots \times g_k$ is called a *cumulant* and, for finite $\mathbf{g}$, we write $\prod \mathbf{g} = g_1 \times g_2 \times \cdots$

The cumulant links all the pairs of vertices that can be joined by a temporal walk of a given length. The mixing time of a random walk on a (fixed) graph depends on the speed at which information diffuses and, in particular, how quickly the cumulant becomes transitive. In the time-varying case, mixing is a more complicated proposition, but the emergence of transitive cumulants is still what guides the parsing process.

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. The non-leading edges form a subgraph of g, which is denoted by $tf(g)$ and called the transitive front of g. 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); on the other hand, the transitive front of a directed cycle over three or more vertices has no edges besides the self-loops. We denote by $cl(g)$ the transitive closure of g: it is the graph that includes an edge (x,y) for any two vertices x,y with a path from x to y. Note that $tf(g) \le g \le cl(g)$.

- An equivalent definition of the transitive front is that the edges of $tf(g)$ are precisely the pairs (i,j) such that $C_i \subseteq C_j$, where C_k denotes the set of vertices l such that (l,k) is an edge of g. Because each vertex has a self-loop, the inclusion $C_i \subseteq C_j$

implies that (i, j) is an edge of g. If g is transitive, then $tf(g) = g$. The set-inclusion definition of the transitive front shows that it is indeed transitive: ie, if (x, y) and (y, z) are edges, 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, $g, h \preceq g \times h$.

- A third characterization of $tf(g)$ is as the unique largest graph h over $[n]$ such that $g \times h = g$: we call this the *maximally-dense property* of the transitive front, and it is the motivation behind our use of the concept. Indeed, the failure of subsequent graphs to grow the cumulant implies a structural constraint on them. This is the sort of structure that parsing attempts to tease out.

A graph sequence $\mathbf{g} = (g_k)_{k>0}$ can be parsed into a rooted tree whose leaves are associated with $g_1, g_2, \ldots$ from left to right. The purpose of the parse tree is to track the creation of new temporal walks over time. This is based on the observation that, because of the self-loops, the cumulant $\prod_{\leq k} \mathbf{g}$ is monotonically nondecreasing with k (with all references to graph ordering being relative to $\preceq$). If the increase were strict at each step, then the parse tree would look like a fishbone. The cumulant cannot grow forever, obviously, and parsing is what tells us what to do when it reaches its maximum size. The underlying grammar consists of three productions: (1a) and (1b) renormalize the graph sequence along the time axis, while (2) creates the hierarchical clustering of the graphs in the sequence $\mathbf{g}$.

1. Temporal renormalization We express the sequence $\mathbf{g}$ in terms of minimal subsequences with cumulants equal to $\prod \mathbf{g}$. There is a unique decomposition

$$\mathbf{g} = \mathbf{g}_1, g_{m_1}, \ldots, \mathbf{g}_k, g_{m_k}, \mathbf{g}_{k+1}$$

such that
(i) $\mathbf{g}_1 = g_1, \ldots, g_{m_1 - 1}$; $\mathbf{g}_i = g_{m_{i-1}+1}, \ldots, g_{m_i - 1}$ $(1 < i \leq k)$; and $\mathbf{g}_{k+1} = g_{m_k+1}, \ldots$.
(ii) $(\prod \mathbf{g}_i) \times g_{m_i} = \prod \mathbf{g}$, for any $i \leq k$; and $\prod \mathbf{g}_i \prec \prod \mathbf{g}$, for any $i \leq k+1$.

The two productions below create the *temporal parse tree*. Unless specified otherwise, the node corresponding to the sequence $\mathbf{g}$ is annotated by the transitive graph $cl(\prod \mathbf{g})$, called its *sketch*.

- *Transitivization.* Assume that $\prod \mathbf{g}$ is not transitive. We define $h = tf(\prod \mathbf{g})$ and note that $h \prec \prod \mathbf{g}$. It follows from the maximally-dense property of the transitive front that $k = 1$. Indeed, $k > 1$ would imply that $\prod \mathbf{g} = (\prod \mathbf{g}_2) \times g_{m_2} \preceq tf\{(\prod \mathbf{g}_1) \times g_{m_1}\} = tf(\prod \mathbf{g})$, which would contradict the non-transitivity of $\prod \mathbf{g}$. We have $\mathbf{g} = \mathbf{g}_1, g_{m_1}, \mathbf{g}_2$ and the production

$$\mathbf{g} \longrightarrow (\mathbf{g}_1)\, g_{m_1} \,((\mathbf{g}_2) \vartriangle h). \tag{1a}$$

In the parse tree, the node for $\mathbf{g}$ has three children: the first one serves as the root of the temporal parse subtree for $\mathbf{g}_1$; the second one is the leaf associated with the graph g_{m_1}; the third one is a *special* node annotated with the label $\vartriangle h$, which serves as the parent of the node rooting the parse subtree for $\mathbf{g}_2$. The purpose of annotating a special node with the label $\triangle h$ is to provide an intermediate approximation of $\prod \mathbf{g}_2$ that is strictly finer than the transitive closure. These coarse-grained approximations form the sketches. Note that special nodes have only one child.

- *Cumulant completion.* Assume that $\prod \mathbf{g}$ is transitive. We have the production

$$\mathbf{g} \;\longrightarrow\; (\,\mathbf{g}_1)\, g_{m_1}\,(\,\mathbf{g}_2)\, g_{m_2}\, \cdots (\,\mathbf{g}_k)\, g_{m_k}\,(\,\mathbf{g}_{k+1}). \tag{1b}$$

Note that the index k may be infinite and any of the subsequences $\mathbf{g}_i$ might be empty (for example, $\mathbf{g}_{k+1}$ if $k = \infty$).

2. Topological renormalization 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: when this happens, we cluster the graphs using topological renormalization. Each nonspecial node v of the temporal parse tree is annotated by the sketch $cl(\prod \mathbf{g})$, where $\mathbf{g}$ is the graph sequence formed by the leaves of the subtree rooted at v. In this way, every path from the root of the temporal parse tree comes with a nested sequence of sketches $h_1 \succeq \cdots \succeq h_l$ (for both special and nonspecial nodes). Pick two consecutive ones, h_i, h_{i+1}: these are two transitive graphs whose strongly connected components, therefore, are cliques. Let $V_1, \ldots, V_a$ and $W_1, \ldots, W_b$ be the vertex sets of the cliques corresponding to h_i and h_{i+1}, respectively. Since h_{i+1} is a subgraph of h_i, it follows that each V_i is a disjoint union of the form $W_{i_1} \cup \cdots \cup W_{i_{s_i}}$.
   - *Decoupling.* We decorate the temporal parse tree with additional trees connecting the sketches along its paths. These *topological parse trees* are formed by all the productions of the type:

$$V_i \;\longrightarrow\; W_{i_1} \cdots W_{i_{s_i}}. \tag{3}$$

A sketch at a node v of the temporal tree can be viewed as an acyclic digraph over cliques: its purpose is to place limits on the movement of the probability mass in any temporal random walk corresponding to the leaves of the subtree rooted at v. In particular, it indicates how decoupling might arise in the system over certain time intervals specified by the temporal parse tree.

The maximum depth of the temporal parse tree is $O(n^2)$ because each child's cumulant loses at least one edge from its parent's (or grandparent's) cumulant. To see why the quadratic bound is tight, consider a bipartite graph $V = L \cup R$, where $|L| = |R|$ and each pair from $L \times R$ is added one at a time as a bipartite graph with a single nonloop edge; the leftmost path of the parse tree is of quadratic length.

Left-to-right parsing. The temporal tree can be built on-line by scanning the graph sequence $\mathbf{g}$ with no need to back up. Let $\mathbf{g}'$ denote the sequence formed by appending the graph g to the end of the finite graph sequence $\mathbf{g}$. If $\mathbf{g}$ is empty, then the tree $\mathcal{T}(\mathbf{g}')$ consists of a root with one child labeled g. If $\mathbf{g}$ is not empty and $\prod \mathbf{g} \prec \prod \mathbf{g}'$, the root of $\mathcal{T}(\mathbf{g}')$ has one left child formed by the root of $\mathcal{T}(\mathbf{g})$ as well as a right child (a leaf) labeled g. Assume now that $\mathbf{g}$ is not empty and that $\prod \mathbf{g} = \prod \mathbf{g}'$. Let v be the lowest internal nonspecial node on the rightmost path of $\mathcal{T}(\mathbf{g})$ such that $c_v \times g = c_v$, where c_u denotes the product of the graphs associated with the leaves of the subtree rooted at node u of $\mathcal{T}(\mathbf{g})$. Let w be the rightmost child of v; note that v and w always exist (the latter because v is internal). We explain how to form $\mathcal{T}(\mathbf{g}')$ by editing $\mathcal{T}(\mathbf{g})$.

1. *If c_v is transitive and w is a leaf.* Referring to (1b), v and w correspond to $\mathbf{g}$ and g_{m_k}, respectively, and $(\mathbf{g}_{k+1})$ is empty. If $g = c_v$, then $(\mathbf{g}_{k+1})$ remains empty while $g_{m_{k+1}} = g$ is created: accordingly, we attach a leaf to v as its new rightmost child and we label it g. On the other hand, if $g \prec c_v$, then $\mathbf{g}_{k+1}$ becomes the sequence consisting of g, so we attach a new rightmost child z to v and then a single leaf to z, which we label g, so as to form $(\mathbf{g}_{k+1})$.
2. *If c_v is transitive and w is not a leaf.* Again, referring to (1b), v and w correspond to $\mathbf{g}$ and the root of $(\mathbf{g}_{k+1})$, respectively. If $c_w \times g = c_v$, then $g = g_{m_{k+1}}$, so we attach a leaf to v as its new rightmost child and we label it g. On the other hand, if $c_w \times g \prec c_v$, then g is appended to the sequence $\mathbf{g}_{k+1}$. Because $c_w \prec c_w \times g$, we create a node z with w as its left child and, as its right child, a leaf labeled g: we attach z as the new rightmost child of v.
3. *If c_v is not transitive and w is a leaf.* Referring to (1a), v and w correspond to $\mathbf{g}$ and g_{m_1}, respectively, and $(\mathbf{g}_2)$ is empty. We know that $g \leq \triangle tf(c_v) \prec c_v$. Accordingly, we give v a new rightmost child z, which we make into a special node and annotate with the label $\triangle tf(c_v)$. We attach a leaf to z and label it g.
4. *If c_v is not transitive and w is not a leaf.* It follows then that w is a special node; let w' be its unique child. Referring to (1a), v and w correspond to $\mathbf{g}$ and the root of $(\mathbf{g}_2)$, respectively. Because $c_w \prec c_w \times g \leq \triangle tf(c_v) \prec c_v$, we create a node z with w' as its left child and, as its right child, a leaf labeled g: we attach z as the new unique child of the special node w.

Undirected graphs. For our purposes, a graph is called undirected if any edge (x, y) with $x \neq y$ comes with its companion (y, x). Consider a sequence of undirected graphs over $[n]$. We begin with the observation that the cumulant of a sequence of undirected graphs might itself be directed; for example, the product $g_1 \times g_2 = (x \leftrightarrow y \quad z) \times (x \quad y \leftrightarrow z)$ has a directed edge from x to z but not from z to x. We can use undirectedness to strengthen the definition of the transitive front. Recall that $tf(g)$ is the unique maximal graph h such that $g \times h = g$. Its purpose is the following: if g is the current cumulant, the transitive front of g is intended to include any edge that might appear in subsequent graphs in the sequence without extending any path in g. Since, in the present case, the only edges considered for extension will be undirected, we might as well require that h itself (unlike g) should be undirected. In this way, we redefine the transitive front, now denoted by $utf(g)$, as the unique maximal *undirected* graph h such that $g \times h = g$. Its edge set includes all the pairs (i, j) such that $C_i = C_j$. Because of self-loops, the condition implies that (i, j) is an undirected edge of g. This forms an equivalence relation among the vertices, so that $utf(g)$ actually consists of disconnected, undirected cliques. To see the difference with the directed case, we take our previous example and note that

$tf(g_1 \times g_2)$ has the edges $(x, y), (x, z), (y, z), (z, y)$ in addition to the self-loops, whereas $utf(g_1 \times g_2)$ has the single undirected edge (y, z) plus self-loops.

As observed in [3], the depth of the parse tree can still be as high as quadratic in n. Here is a variant of the construction. Given a clique C_k over k vertices $x_1, \ldots, x_k$ at time t, attach to it, at time $t+1$, the undirected edge (x_1, y). The cumulant gains the undirected edge (x_1, y) and the directed edges (x_i, y) for $i = 2, \ldots, k$. At time $t + 2, \ldots, t + k$, visit each one of the $k - 1$ undirected edges (x_1, x_i) for $i > 1$, using single-edge undirected graphs with self-loops. Each such step will see the addition of a new directed edge (y, x_i) to the cumulant, until it becomes the undirected clique C_{k+1}. The quadratic lower bound on the tree depth follows immediately.

References

1. Avin, C., Koucký, M., Lotker, Z. *How to explore a fast-changing world (cover time of a simple random walk on evolving graphs)*, Proc. 35th ICALP (2008), 121–132.
2. Bruin, H., Deane, J.H.B. *Piecewise contractions are asymptotically periodic*, Proc. American Mathematical Society 137, 4 (2009), 1389–1395.
3. Chazelle, B. *Toward a theory of Markov influence systems and their renormalization*, Proc. 9th ITCS (2018), 58:1-58:18.
4. Condon, A., Hernek, D. *Random walks on colored graphs*, Random Structures and Algorithms 5 (1994), 285–303.
5. Denysyuk, O., Rodrigues, L. *Random walks on directed dynamic graphs*, Proc. 2nd International Workshop on Dynamic Networks: Algorithms and Security (DYNAS10), Bordeaux, France, July, 2010. Also arXiv:1101.5944 (2011).
6. Denysyuk, O., Rodrigues, L. *Random walks on evolving graphs with recurring topologies*, Proc. 28th International Symposium on Distributed Computing (DISC), Austin, Texas, USA, October 2014.
7. Holme, P., Saramäki, J. *Temporal networks*, Physics Reports 519 (2012), 97–125.
8. Iacobelli, G., Figueiredo, D.R. *Edge-attractor random walks on dynamic networks*, J. Complex Networks 5 (2017), 84–110.
9. Nguyen, G.H., Lee, J.B., Rossi, R.A., Ahmed, N., Koh, E., Kim, S. *Dynamic network embeddings: from random walks to temporal random walks*, 2018 IEEE International Conference on Big Data (2018), 1085–1092.
10. Perra, N., Baronchelli, A., Mocanu, D., Gonçalves, B., Pastor-Satorras, R., Vespignani, A. *Random walks and search in time-varying networks*, Phys. Rev. Lett. 109, 238701, 2012.
11. Petit, J., Gueuning, M., Carletti, T., Lauwens, B., Lambiotte, R. *Random walk on temporal networks with lasting edges*, Phys. Rev. E 98, 052307, 2018.
12. Ramiro, V., Lochin, E., Sénac, P., Rakotoarivelo, T. *Temporal random walk as a lightweight communication infrastructure for opportunistic networks*, 2014 IEEE 15th International Symposium (WoWMoM), June 2014.
13. Seneta, E. Non-Negative Matrices and Markov Chains, Springer, 2nd ed., 2006.
14. Starnini, M., Baronchelli, A., Barrat, A., Pastor-Satorras, R. *Random walks on temporal networks*, Phys. Rev. E 85, 5 (2012), 056115–26.