Randomized reference models for temporal networks

Laetitia Gauvin, Mathieu Génois, Marton Karsai, Mikko Kivelä, Taro Takaguchi, Eugenio Valdano, Christian L. Vestergaard

Randomized reference models for temporal networks

L. Gauvin,[1] M. Génois,[2] M. Karsai,[3,4,5] M. Kivelä,[6] T. Takaguchi,[7] E. Valdano,[8] and C. L. Vestergaard[2,9,*]

[1] *Data Science Lab, ISI Foundation, Turin, Italy.*
[2] *Aix Marseille Univ., Univ. Toulon, CNRS, CPT, Marseille, France.*
[3] *Department of Network and Data Science, Central European University, 1100 Vienna, Austria.*
[4] *Université de Lyon, ENS de Lyon, INRIA, CNRS, LIP-UMR 5668, IXXI, 69364, Cedex 07 Lyon, France.*
[5] *Alfréd Rényi Insititute of Mathematics, H-1053 Budapest, Hungary*
[6] *Department of Computer Science, Aalto University School of Science, P.O. Box 12200, FI-00076 Aalto, Finland.*
[7] *National Institute of Information and Communications Technology, 4-2-1 Nukui-Kitamachi, Koganei, Tokyo 184-8795, Japan.*
[8] *Center for Biomedical Modeling, The Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, 760 Westwood Plaza, University of California Los Angeles, Los Angeles, CA 90024 USA.*
[9] *Decision and Bayesian Computation, Department of Computational Biology, Department of Neuroscience, CNRS USR 3756, CNRS UMR 3571, Institut Pasteur, 25 rue du Docteur Roux, Paris, 75015, France.*

(Dated: November 24, 2021)

Many dynamical systems can be successfully analyzed by representing them as networks. Empirically measured networks and dynamic processes that take place in these situations show heterogeneous, non-Markovian, and intrinsically correlated topologies and dynamics. This makes their analysis particularly challenging. Randomized reference models (RRMs) have emerged as a general and versatile toolbox for studying such systems. Defined as random networks with given features constrained to match those of an input (empirical) network, they may for example be used to identify important features of empirical networks and their effects on dynamical processes unfolding in the network. RRMs are typically implemented as procedures that reshuffle an empirical network, making them very generally applicable. However, the effects of most shuffling procedures on network features remain poorly understood, rendering their use non-trivial and susceptible to misinterpretation. Here we propose a unified framework for classifying and understanding microcanonical RRMs (MRRMs) that sample networks with uniform probability. Focusing on temporal networks, we survey applications of MRRMs found in literature, and we use this framework to build a taxonomy of MRRMs that proposes a canonical naming convention, classifies them, and deduces their effects on a range of important network features. We furthermore show that certain classes of MRRMs may be applied in sequential composition to generate new MRRMs from the existing ones surveyed in this article. We finally provide a tutorial showing how to apply a series of MRRMs to analyze how different network features affect a dynamic process in an empirical temporal network. Our taxonomy provides a reference for the use of MRRMs, and the theoretical foundations laid here may further serve as a base for the development of a principled and automatized way to generate and apply randomized reference models for the study of networked systems.

* cvestergaard@gmail.com

CONTENTS

I. INTRODUCTION

Random network models are responsible of major parts of our theoretical understanding of networked systems and practical knowledge extracted from networked data. Well-known examples of such models include the Erdős-Rényi model [1] – which generates random networks with fixed numbers of nodes and links – and the configuration model [2, 3] – which fixes the degree sequence. A main application of these models is as null models for hypothesis testing, though their use goes beyond this. They may notably be used more generally to investigate the relationship between different network features [4–8] and their roles in dynamic phenomena [9–35]. To underline their general scope, we here call these type of models *randomized reference models* (RRMs). Much of what we know of the behavior of dynamic processes in networks is based on them [27, 28], and they stand behind many prominent results in network science such as the absence of epidemic threshold [29], the vulnerability to attacks [30–33], and the robustness to failures [33–35] in certain types of networks. These models are also integral parts of many methods for network data analysis, such as popular network clustering methods [36, 37], network motif analysis [38, 39], and the analysis of structural correlations [40, 41]. The above is just a small selection of applications, but the examples are legion.

As network science has matured there has been an increasing need to go beyond the simple graph representation for networks, and at the same time repeat the success of RRMs for these new types of networks. An important extension to simple graphs is temporal networks, which allow the networks' topology to evolve in time. RRMs[1] have also emerged as a powerful toolbox for the study of the dynamics on and of temporal networks [45, 47, 55] and have been applied to complex systems in a broad range of fields, including sociology, epidemiology, infrastructure, economics, and biology. They have been used to study how given temporal network features affect other node- or interaction-level features [4–8], and how the features affect dynamical processes unfolding in the network [9–26] as well as the network's controllability [49, 50, 56]. Systems studied using temporal network RRMs include: human face-to-face interactions and physical proximity [12, 14–16, 19–22, 25, 26, 51, 53, 54, 57]; prostitution networks [10, 15, 22, 26]; functional connections in the brain [57–59]; human mobility [60, 61]; livestock transport [42]; mobile phone calls and text messages [7, 9, 11, 13, 19]; email correspondences [9, 15, 16, 19, 22, 50, 52, 54]; online communities [15, 22, 52, 57, 62, 63]; editing of Wikipedia pages [46]; and world trade [43, 44, 48].

The popularity of RRMs for the study of complex networks may be explained by the fact that they can often be defined simply as numerical procedures that generate random networks by shuffling the original data, thus avoiding the need to specify a complete generative model. The resulting set of randomized networks typically serves as a null reference that is compared to the original temporal network, or it may be compared to a second set of networks generated by another RRM. For example, by comparing how a given dynamic process evolves on the original network with how it evolves on different sets of randomized networks, we may identify how various features of the network affect the dynamic process.

The algorithmic definition of RRMs as shuffling methods makes them simple to apply in very general settings and with little domain-specific tweaking needed. However, an important downside to the algorithmic representation is that the effects of RRMs on network features are rarely investigated systematically and remain poorly understood. This lack of systematic understanding of the methods is not only a theoretical problem but it has led to severe practical problems in the literature. First, there are no unified naming conventions for the RRMs. This makes it difficult to compare the methods used in different studies and has lead to a situation where the algorithms producing equivalent RRMs are given a multitude of different names, and possibly worse, where multiple algorithms producing different RRMs are given the same names [64] (see also Section IV B). Second, researchers are confronted with the problems of how to choose and develop randomization techniques, in which order they should be applied, and how to interpret the results. These are crucial choices in order to be able to identify important features for each given dynamical phenomenon and for each temporal network under study (problems that are non-trivial even for the study of simple graphs [8]).

We review temporal network RRMs used in the literature and find that most of them fall into a class of methods that gives a uniform probability of sampling all networks with a given set of features constrained to the same value as that of the original data [13]. These RRMs are described by the concept of microcanonical ensembles from statistical physics [65, 66]. We will consequently call them *microcanonical* randomized reference models (MRRMs) and represent them in a formal framework where they are fully defined by the set of features they constrain. This principled approach has several advantages over the algorithmic representation: As MRRMs are completely defined by the constraints that they impose, we propose an unambiguous naming convention for MRRMs of temporal networks based on these constraints. Furthermore, the theoretical framework enables us to build a taxonomy of existing MRRMs, which lists their effects on important temporal network features and orders them by the amount of features they constrain.

[1] In the temporal network literature RRMs are also known as *null models* [4, 7, 9, 14, 15, 20, 21, 25, 42–49]; *reshuffling methods* [24], *randomization techniques* [47, 50], *randomization procedures* [14, 16, 42, 47, 50], *randomization strategies* [14], *randomization schemes* [47], *randomization methods* [51], or simply *randomizations* [16, 45, 47, 50–54]).

This hierarchy allows researchers to apply MRRMs so that the fixed features of the original data are systematically reduced. We also show how and when new MRRMs can be devised by applying previously implemented algorithms one after another. We finally illustrate how series of MRRMs may be applied to analyze temporal network data with a walk-through example.

Reference models, which keep parts of the features of original data and shuffle the rest, are clearly widely applicable outside of temporal networks. For example, MRRMs are closely related to exact (permutation) tests of classical statistics [67] and to conditionally uniform graph tests (CUGTs) found in the sociology literature [52, 68, 69]. Furthermore, even though we are here mostly concerned with temporal networks, our framework of MRRMs is directly applicable to a far more general class of systems, which can be considered as a realization of a state in a predefined discrete state space. In particular, it may be applied e.g. to correlation matrices and to more general types of relational data such as multilayer networks or hypergraphs.

It is our aim that, in addition to categorizing previous RRMs and surveying the literature, the unified framework and taxonomy we present would serve as a starting point for the development of a general and principled randomization-based approach for the characterization and analysis of networked dynamical systems. To this end, we provide (at `https://github.com/mgenois/RandTempNet`) a pure python library implementing the MRRMs presented in this article; furthermore, for applications to larger networks we provide (at `https://github.com/bolozna/Events`) a fast Python library with core functions written in C++ implementing most of the MRRMs.

We begin in Section II by providing theoretical definitions needed to describe temporal networks (Section II A) and MRRMs (Section II B). We use these to define classes of shuffling methods that are used to generate randomized temporal networks in practice (Section II C). In Section III we survey applications of MRRMs found in the literature, using the theoretical foundations posed in the previous section to provide consistent names for the MRRMs used in each study. In Section IV we develop the theoretical framework needed to compare and build hierarchies of network features and of MRRMs (Section IV A), and we use this to build a taxonomy of MRRMs found in the literature that characterizes and orders them hierarchically (Section IV B). In Section V we develop a theory for composing MRRMs by applying one algorithm after another (Section IV A). We exploit this to show which classes of shuffling methods can be combined to form new MRRMs from existing ones (Section V B), and we classify several MRRMs found in the literature that are compositions of two other MRRMs (Section V C). In Section VI we provide a brief overview of temporal network RRMs that do not fall into the class of microcanonical methods. In Section VII we finally provide a walk-through example showcasing how to apply nested series of MRRMs to analyze an empirical temporal network.

An appendix provides formal definitions that are left out in the main text and gives proofs for the propositions and theorems developed in Sections IV A, V A, and V B. We furthermore provide a supplemental material, which includes supplementary tables detailing features of temporal networks as well as four supplementary notes: Supplementary Note 1 discusses how we can randomize the durations of the events in a temporal network using shuffling methods for networks with instantaneous events; Supplementary Note 2 provides detailed definitions for a large selection of important features of temporal networks and orders them hierarchically; Supplementary Note 3 provides a second walk-through example describing an analysis of features of a face-to-face interaction network using series of MRRMs; Supplementary Note 4 finally lists the names of the algorithms in the python library used for generating the MRRMs reviewed here.

II. FUNDAMENTAL DEFINITIONS

In this section we define fundamental concepts for temporal networks (Section II A) and for microcanonical randomized reference models (MRRMs) (Section II B). Based on this, we propose a canonical naming convention for MRRMs that unambiguously and completely defines a MRRM (Section II B). We finally describe several important classes of algorithms for sampling randomized networks and we classify them according to what overall structures of a temporal network they randomize (Section II C).

We focus on microcanonical RRMs as these represent the class of maximum entropy models that can be obtained directly by randomly sampling constrained permutations of an empirically observed network (i.e. by shuffling the network). As shuffling is by far the most dominant method for generating randomized temporal networks, we shall in the following refer to any algorithm for generating random networks according to a given MRRM as a *shuffling method* or simply a *shuffling*.

A. Temporal network

We consider a system consisting of N individual nodes engaging in intermittent dyadic interactions observed over a period of time from $t = t_{\min}$ to $t = t_{\max}$; in a social network, for example, the nodes are persons, in an ecological network nodes are species, while in a transport network they are locations. A temporal network is our representation of such an observation.

Definition II.1. *Temporal network*[2]. A temporal network $G = (\mathcal{V}, \mathcal{C})$ is defined by the set of nodes $\mathcal{V} = \{1, 2, ..., N\}$ and a set of events $\mathcal{C} = \{c_1, c_2, \ldots, c_C\}$, where each event $c_q = (i, j, t, \tau)$ represents an interaction between nodes i and j during the time-interval $[t, t + \tau)$.

Definition II.1 encompasses both temporal networks with directed interactions (e.g. for phone-call or instant messaging networks) and undirected interactions (e.g. for face-to-face or proximity networks), but does not consider possible weights of the events. For directed networks, we may adopt the convention that the direction of an event (i, j, t, τ) is from i to j. For undirected networks, the presence of the event (i, j, t, τ) implies the symmetric interaction (j, i, t, τ), and in practice we may impose $i < j$ to avoid double counting.

For simplicity, we consider only undirected temporal networks in the main text. However, all models and methods may be applied directly to temporal networks with directed and/or weighted events by defining the appropriate features (see Supplementary Table S2 for features that explicitly account for directed events).

Example II.1. Figure II.1(a) shows a schematic temporal network consisting of four nodes.

For some systems, e.g., email communications or instant messaging [9, 15, 16, 19, 22, 50, 52, 54], events are instantaneous; in other cases, event durations are so short compared to the time-intervals between them, the *inter-event durations*, that they may be treated as instantaneous [9, 13]. Both cases are included in the above framework by setting $\tau = 0$. We may then reduce our representation of the sequence of events to a sequence of reduced events, leading us to the following definition.

Definition II.2. *Instant-event temporal network.* An instant-event temporal network $G = (\mathcal{V}, \mathcal{E})$ is defined by the set of nodes $\mathcal{V} = \{1, 2, ..., N\}$ and a set of instantaneous events $\mathcal{E} = \{e_1, e_2, \ldots, e_E\}$, where each event $e_q = (i, j, t)$ describes an interaction between nodes i and j at time t, but where the duration is implicit.

Some systems with continuous dyadic activity, notably face-to-face interaction and proximity networks [71, 72], are recorded with a coarse time resolution at evenly spaced points in time, $t = t_{\min}, t_{\min} + \Delta t, t_{\min} + 2\Delta t, \ldots, t_{\max} - \Delta t$. In this case we may also represent the system as an *instant-event network*, where the events mark a beginning of an activity at each measurement time and the time-resolution $\tau = \Delta t$ is implicit.

Example II.2. Figure II.1(b) shows a schematic discrete-time instant-event temporal network corresponding to the temporal network of Example II.1 (shown in Fig. II.1(a)).

[2] Our definition of a temporal network is equivalent to what is called a link stream in [70].

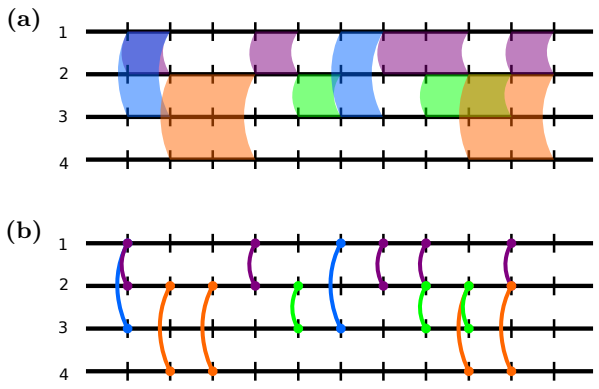


FIG. II.1: Graphical (*node-timeline*) representations of temporal network. Each node is represented as a timeline and events as links between corresponding timelines at corresponding points in time. (a) Temporal network with event durations recorded at discrete-time resolution. (b) Instant-event temporal network representation of the temporal network in (a).

1. *Temporal network feature*

We are in general interested in comparing the values of given features of an observed temporal network (either recorded empirically or generated by a model) to their values in randomized networks generated by a MRRM that constrains certain other features. To formalize both the notion of a temporal network feature and that of a MRRM, we first need to define a *state space* comprising all temporal networks.

Definition II.3. *State space $\mathcal{G}$.* The *state space* $\mathcal{G}$ is a predefined finite[3] set of temporal networks.

With the definition of the state space $\mathcal{G}$, representing the world of possible networks, we can now define any feature of a network as a function on $\mathcal{G}$.

Definition II.4. *Temporal network feature.* A *feature* $\mathbf{x}$ is any function that takes as an input any temporal network $G \in \mathcal{G}$. Formally, given a state space $\mathcal{G}$ (Def. II.3), $\mathbf{x}$ is a function that has $\mathcal{G}$ as domain.

Often a feature is a vector-valued function. However, the definition allows for more general functions, e.g. a function that returns a graph. Furthermore, the feature

[3] Note that from a practical point of view it is enough for our purposes to consider only finite state spaces. Considering all possible temporal networks would make the state space at least countably infinite, but as all of the reference models encountered in the literature keep the system size (measured in the number of nodes and the length of the observation period) fixed, we can always fix the state space to contain only networks of fixed size. Furthermore, the time can always be considered finite as the measurement resolution is finite.

does not need to be a structural feature of a network. It may for example quantify the outcome of a dynamic process on the network, such as the expected time needed for a contagion to infect a given proportion of the nodes.

An important temporal network feature is the *static graph*, which summarizes the time-aggregated topology of a temporal network.

Definition II.5. *Static graph.* The static graph, G^{stat}, is a function which returns a simple (i.e. static, unweighted, and undirected) graph $G^{\text{stat}}(G) = (\mathcal{V}, \mathcal{L})$ with the same set of nodes $\mathcal{V}$ as the original temporal network G and the set of *links* $\mathcal{L} = \{(i,j) : (i,j,t,\tau) \in \mathcal{C}\}$, which includes all pairs of nodes (i,j) that interact at least once in G.

Note that by Def. II.5 the static graph is a feature of a temporal network. Conversely, we may see the temporal network as a direct generalization of the static graph to include information about the time-evolution of the system's topology [45]. Note also that here we have defined the static graph as an unweighted graph, but one could also use the number of events between each pair of nodes or their cumulative duration to define link weights.

2. *Two-level temporal network representations*

Sometimes it is useful to separate the static structure and the temporal aspect in the definition of the temporal network as opposed to having them mixed together like in definitions II.1 and II.2. This can be done by separating these two aspects into two levels, either by first defining the static network structure and then the activation times of each link or by first defining the sequence of activation times and then the network structure at each of those times [47]. We call the first of these options a *link-timeline network* and the second a *snapshot-graph sequence*. These two-level temporal network representations will be practical for designing and implementing temporal network MRRMs (Sec. II C).

Definition II.6. *Link-timeline network.* A link-timeline network represents a temporal network using an edge-valued graph $G_{\mathcal{L}} = (\mathcal{V}, \mathcal{L}, \boldsymbol{\Theta})$. It uses the static graph $G^{\text{stat}} = (\mathcal{V}, \mathcal{L})$ to indicate the pairs of nodes that interact at least once during the observation period (Def. II.5). To each link $(i,j) \in \mathcal{L}$ it associates a *timeline* $\Theta_{(i,j)} \in \boldsymbol{\Theta}$, which indicates when the corresponding nodes interact. Each timeline is given by a sequence,

$$\Theta_{(i,j)} = \left(\left(t^1_{(i,j)}, \tau^1_{(i,j)}\right), \left(t^2_{(i,j)}, \tau^2_{(i,j)}\right), \ldots, \left(t^{n_{(i,j)}}_{(i,j)}, \tau^{n_{(i,j)}}_{(i,j)}\right)\right) , \tag{1}$$

where $t^m_{(i,j)}$ is the start of the mth event on link (i,j), with $\tau^m_{(i,j)}$ its duration, and $n_{(i,j)}$ is the total number of events taking place over the link [4].

[4] A link-timeline network is equivalent to the *interval graph* defined in [45]. If we set $\tau_m = 0$, i.e. in the instantaneous event limit, we obtain what is termed *contact sequences* in [45].

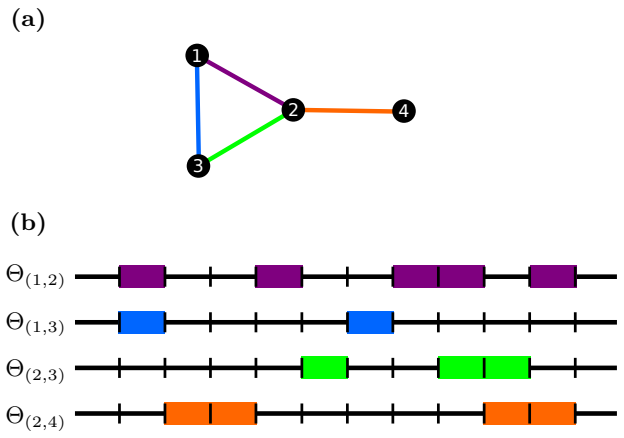


FIG. II.2: Link-timeline network. Graphical representation of the link-timeline representation $G_{\mathcal{L}}$ of the discrete-time temporal network shown in Fig. II.1. (a) Static graph $G^{\text{stat}} = (\mathcal{V}, \mathcal{L})$ showing links between nodes. Links are drawn between pairs (i,j) of nodes that interact at least once. (b) Timelines $\boldsymbol{\Theta} = \left(\Theta_{(1,2)}, \Theta_{(1,3)}, \Theta_{(2,3)}, \Theta_{(2,4)}\right)$ of the links $\mathcal{L}$ in the graph, showing when each link is active.

Example II.3. Figure II.2 shows the link-timeline representation of the temporal network of Example II.1.

Alternative to the link timeline representation we may think of a temporal network as a time-varying sequence of instantaneous graph snapshots. This leads to the following definition:

Definition II.7. *Snapshot-graph sequence.* A *snapshot-graph sequence*, $G_{\mathcal{T}} = (\mathcal{T}, \boldsymbol{\Gamma})$, represents a temporal network using a sequence of times, $\mathcal{T} = (t_1, t_2, \ldots, t_T)$, and a sequence of snapshot graphs,

$$\boldsymbol{\Gamma} = (\Gamma^1, \Gamma^2, \ldots, \Gamma^T) \ , \tag{2}$$

where for each $m = 1, 2, \ldots, T$, $\Gamma^m \in \boldsymbol{\Gamma}$ is associated to $t_m \in \mathcal{T}$. The *snapshot graphs* are defined as graphs $\Gamma^m = (\mathcal{V}, \mathcal{E}^{t_m})$, where $\mathcal{V}$ is the set of nodes and $\mathcal{E}^t$ is the set of edges for which there is an event taking place at time t,

$$\mathcal{E}^t = \{(i,j) : (i,j,t) \in \mathcal{E}\} \ . \tag{3}$$

Instantaneous-event networks can be represented as snapshot-graph sequences by constructing the sequence of times $\mathcal{T}$ as the times at which at least one event takes place. This is a natural representation especially for networks which are recorded with a fixed time resolution Δt, as the sequence of times becomes $\mathcal{T} = (\Delta t, 2\Delta t, \ldots, T)$, and if the time resolution is coarse enough so that the individual snapshot graphs do not become too sparse. We

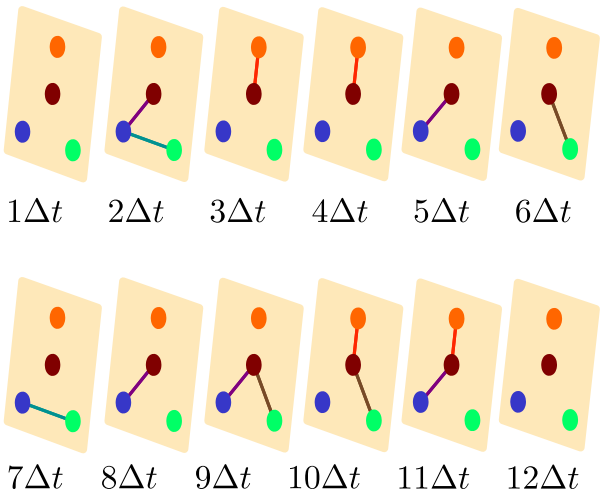


FIG. II.3: **Snapshot-graph sequence.** Sequence of snapshot graphs of the temporal network shown in Figs. II.1 and II.2.

use the shorthand Γ^t to refer to the snapshot graph associated with the time $t \in \mathcal{T}$ (i.e. Γ^m for m such that $t_m = t$).

Example II.4. Figure II.3 shows the snapshot-graph sequence for the temporal network of Example II.2.

Note that the two-level temporal networks do not add anything new to the temporal network structure. They are simply alternative ways of representing them because any temporal network can be uniquely represented as a link-timeline network and any instant-event temporal network can be uniquely represented as a snapshot-graph sequence. Despite this, the representations are often used for specific types of systems and they come with their own perspective on temporal networks.

The link-timeline networks are often used for data that is sparse in time such that only a small fraction of links are active at each time instant. Furthermore, because the static network is made explicit in its definition it is easy to think of the temporal network as having a latent static network which manifests as activation events of the links. For example, for an email communication data represented as link-timeline network one might consider the static graph as an acquaintance or friendship network where each social tie is activated during the communication events. The structure also guides the generation of randomized reference models, because it is easy to either randomize the static graph and keep the link timelines, or randomize the timelines while keeping the static graph.

The snapshot-graph sequences are more likely used for data that are dense in time such that each snapshot graph contains a reasonable number of links. Furthermore, this representation is natural for networks that change in time while the network nature of the system is still important at each separate time instance. For example, for networks in which the structure changes on the same or longer timescales as dynamics on that network, it is important to be able to look at the topology of the network at each time instance. Here, again, the structure guides the construction of randomized reference models: It is convenient to define shuffling methods where the order of the snapshot graphs change or where each snapshot is independently randomized.

The two-level temporal network representations thus provide convenient ways to define and generate MRRMs that constrain certain overall properties. We will explore this in detail in Subsection II C below and see that many RRMs found in the literature are implemented this way in Section IV B.

B. Microcanonical randomized reference model (MRRM)

Here we give a rigorous definition of a microcanonical randomized reference model (MRRM). We use this to propose an unambiguous *canonical* naming convention for MRRMs. We finally discuss several equivalent ways to represent a MRRM, each of which is useful for different purposes.

We consider a predetermined state space $\mathcal{G}$ (Def. II.3), where states are here temporal networks, and a single observation $G^* \in \mathcal{G}$, termed the *input network*.

Many procedures leading to a sampling from a conditional probability distribution $P(G|G^*)$ defined on $\mathcal{G}$ could be considered to be *randomized reference models* (RRMs). In order for such models to be useful for testing hypothesis and finding effect sizes they need to retain some of the properties of the original network G^* and randomize others in a controlled way. In the context of simple graphs, the most popular choices of RRMs include Erdős-Rényi (ER) models [1, 2], configuration models [2, 3], and exponential random graph models [2].

Here we will focus on models that exactly preserve certain features but are otherwise maximally random. The principle of maximum entropy formalizes this notion, and it provides theoretical justification for using such models as they are the least possible biased w.r.t. all other degrees of freedom [65, 66]. Maximum entropy models that impose exact (i.e. *hard*) constraints are called microcanonical models, and we will thus refer to RRMs that exactly constrain given feature values as *microcanonical randomized reference models* (MRRMs).

Definition II.8. *Microcanonical randomized reference model (MRRM).* Consider any function $\mathbf{x}$ that has the set $\mathcal{G}$ as domain (i.e. a temporal network feature, Def. II.4). A MRRM is then a model which given $G^* \in \mathcal{G}$ returns $G \in \mathcal{G}$ with probability:

$$P_{\mathbf{x}}(G|G^*) = \frac{\delta_{\mathbf{x}(G),\mathbf{x}(G^*)}}{\Omega_{\mathbf{x}}(G^*)} \ , \tag{4}$$

where δ is the Kronecker delta function, and $\Omega_{\mathbf{x}}(G^*) = \sum_{G\in\mathcal{G}} \delta_{\mathbf{x}(G),\mathbf{x}(G^*)}$ is a normalization constant.

We will sometimes use the shorthand notation $\mathbf{x}^* = \mathbf{x}(G^*)$, and $\Omega_{\mathbf{x}^*} = \Omega_{\mathbf{x}}(G^*)$. Furthermore, because the conditional probability depends only on the value of $\mathbf{x}$ in G^* we can define the notation $P_{\mathbf{x}}(G|\mathbf{x}^*) = P_{\mathbf{x}}(G|G^*)$.

In the above definition the feature function $\mathbf{x}$ defines the features of G^* that are retained in the randomized reference model. In statistical physics terms $\Omega_{\mathbf{x}^*}$ is the *microcanonical partition function*.

Note that restricting ourselves to a single feature entails no loss in generality since any number of distinct features may be combined into one tuple-valued feature, e.g., for two distinct features $\mathbf{x}$ and $\mathbf{y}$, we may simply define a third tuple-valued feature $\mathbf{z} = (\mathbf{x}, \mathbf{y})$. This defines what we shall call the *intersection* of two MRRMs:

Definition II.9. *Intersection of randomized reference models.* The *intersection* of two features $\mathbf{x}$ and $\mathbf{y}$ is the tuple ($\mathbf{x}$,$\mathbf{y}$), and for the associated MRRM we write $\mathrm{P}[\mathbf{x}, \mathbf{y}] = \mathrm{P}[(\mathbf{x}, \mathbf{y})]$

1. Naming convention

A MRRM is completely defined by the feature(s) it constrains (Def. II.8). This lets us propose a rigorous naming convention that specifies a MRRM by listing the corresponding features.

Definition II.10. *Naming convention for MRRMs.* A MRRM that constrains the individual temporal network features $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_Q$ is named $\mathrm{P}[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_Q]$.

Note that our naming convention is not unique as the list of features is not required to be non-overlapping, so we may always devise different ways to name the same MRRM (for a practical example see the description of the MRRM $\mathrm{P}[\mathbf{w}, \mathbf{t}]$ in Section IV B 8). It is however unambiguous as a set of features uniquely defines a single MRRM (Def. II.8). This means that a name always uniquely defines a MRRM.

Example II.5. In the context of simple graphs, the variant of the ER model [1] that returns uniformly at random a graph with N nodes and L edges, and the variant of the configuration model that returns uniformly a randomly selected simple graph with degree sequence $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$, also known as the *Maslov-Sneppen* model [38], are MRRMs. In the space of simple graphs with N nodes, the ER model is defined as $\mathrm{P}[L]$. It maps an input graph G^* to a microcanonical ensemble of graphs that all have the same number of links L as the input graph G^*, but are otherwise uniformly random. The Maslov-Sneppen model is defined as $\mathrm{P}[\mathbf{k}]$, and it maps G^* to a microcanonical ensemble of graphs that all have the same sequence $\mathbf{k}$ of node degrees as G^*.

Note that MRRMs are always defined relative to a state space $\mathcal{G}$ (Def. II.3), which should also be specified. In the context of reference models for temporal networks $\mathcal{G}$ contains all networks with the same set of nodes $\mathcal{V}$ and the same temporal duration $t_{\max} - t_{\min}$ as the original (input) network. We do thus not need to include these features in the names of temporal network MRRMs as they are always constrained.

Example II.6. A popular temporal network MRRM is the model which randomizes the time stamps of the instantaneous events completely inside each timeline without changing the aggregated topology of the network, leading the events to follow a Poisson process on each timeline. In the space of instant event temporal networks with a fixed set of nodes and observation interval it is named $\mathrm{P}[\mathbf{w}]$. Here $\mathbf{w} = (w_{(i,j)})_{(i,j)\in\mathcal{L}}$ is the sequence of link weights, which retains the number of instantaneous events on each link and the links' placement in the static graph. Several different names has been used in the literature to designate this MRRM: *random time(s)* [45, 47, 50, 52], *uniformly random times* [13], *temporal mixed edges* [42], *Poissonized inter-event intervals* [51], and *SRan* [12].

2. MRRM representations

While the definition of MRRMs is written as a conditional probability it is often useful to use alternative representations of MRRMs. Namely, as a *shuffling method* that uniformly samples randomized networks, as a partition of the state space, and as a transition matrix between states. All of these representations are equivalent in a sense that they completely and uniquely specify a MRRM. The power of the equivalence between the different representations is that any result proven for one representation automatically carries over to the others. We will in the following switch between the representations to use the one that is most convenient in each context: the definition as a conditional probability notably provided a consistent naming convention that fully characterizes any MRRM (Def. II.10), shuffling methods are how MRRMs are implemented in practice (we will explore this in Subsection II C below), while the partition and matrix pictures will provide theoretical underpinnings for building hierarchies of MRRMs (Section IV) and for generating new MRRMs from existing ones (Section V).

Definition II.11. *MRRM representations:*

1. *Shuffling method.* An algorithm that transforms G^* into G according to Def. II.8. These algorithms often shuffle some elements of G^*. Note that multiple algorithms or shuffling procedures might correspond to the same MRRM and in this case these are considered here to be the same shuffling method.

2. *Partition of the state space.* The feature function $\mathbf{x}$ (Def. II.4) defines an equivalence relation and thus partitions the state space $\mathcal{G}$ (Def. II.3) [73]: Given $\mathbf{x}$, one can construct a partition of the state space $\{\mathcal{G}_i\}$ (i.e., a set of subsets of $\mathcal{G}$ where each element

of $\mathcal{G}$ is in exactly one subset) such that $G, G' \in \mathcal{G}_i$ if $\mathbf{x}(G) = \mathbf{x}(G')$. The set which $G^* \in \mathcal{G}$ belongs to in this partition is the $\mathbf{x}$-*equivalence class* of G^* and is denoted by $\mathcal{G}_{\mathbf{x}^*} = \{G \in \mathcal{G} : \mathbf{x}(G) = \mathbf{x}^*\}$. Note that the partition function which normalizes the conditional probability (Def. II.8) is the cardinality of this set, $\Omega_{\mathbf{x}^*} = |\mathcal{G}_{\mathbf{x}^*}|$.

3. *Transition matrix.* A MRRM is a symmetric linear stochastic operator mapping the state space $\mathcal{G}$ to itself. For a given indexing of the state space $\mathcal{G}$, we can represent a MRRM by a transition matrix $\mathbf{P}^{\mathbf{x}}$ with elements

$$\mathbf{P}^{\mathbf{x}}_{ij} = P_{\mathbf{x}}(G_j|G_i) \ . \tag{5}$$

$\mathbf{P}^{\mathbf{x}}$ is always a block diagonal matrix where inside each block the elements have the same value.

Example II.7. To illustrate the different MRRM representations, we consider the state space $\mathcal{G}$ of all static graphs with 3 nodes and the MRRM P$[L]$, defined by the feature L which returns the number of edges in the network, corresponding to the Erdős-Rényi random graph model ER$(3, L)$. We number the 8 graphs in $\mathcal{G}$ such that $L(G_1) = 0$, $L(G_i) = 1$ for $i = 2, 3, 4$, $L(G_i) = 2$ for $i = 5, 6, 7$, and $L(G_8) = 3$ [Fig. II.4(a)], and we take as input state the graph $G^* = G_5$. Figure II.4 illustrates the four different representations of P$[L]$ for the state space $\mathcal{G}$ and the input graph $G^* = G_5$. Note that in a real application the number of states is typically much much larger than in this example. This in particular means that almost all states are sampled at most once in practice.

C. Shuffling methods and classes of temporal network MRRMs

We describe here several important classes of shuffling methods that are used to formulate and generate MRRMs in practice. The classes are formulated depending on which parts of a temporal network they randomize.

All the MRRMs we will encounter are implemented by methods that shuffle the positions of the events in a temporal network, or for instant-event networks, the instantaneous events. We shall call the former type of shuffling method an *event shuffling* and the latter an *instant-event shuffling*. The shufflings are generally implemented by randomizing any or all of the indices i, j, and t in the events $(i, j, t, \tau) \in \mathcal{C}$ or in the instantaneous events $(i, j, t) \in \mathcal{E}$.

These methods are practical for generating reference models as they all conserve the nodes $\mathcal{V}$, the temporal duration $t_{\max} - t_{\min}$ and the number of events, C (or instantaneous events, E). In addition to these features, event shufflings also conserve the events' durations, i.e.

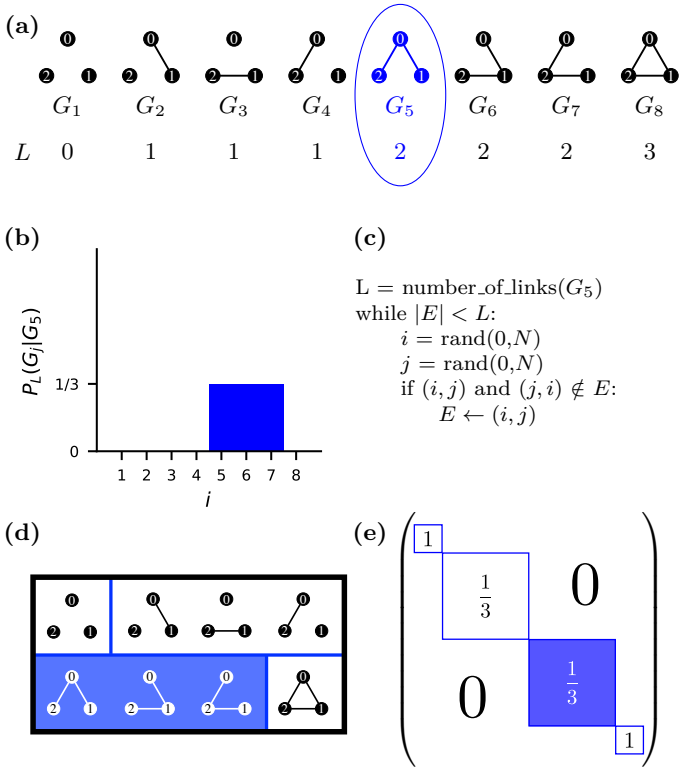


FIG. II.4: **Equivalent MRRM representations.** We consider the state space of all simple graphs with three nodes and the MRRM P$[L]$ which constrains the number of edges L in a graph. The state space $\mathcal{G}$ contains eight states (graphs) and is ordered as shown in (a). The input network is taken to be $G^* = G_5$ (highlighted in blue) for which $L^* = L(G_5) = 2$. Panels (b)-(e) illustrate graphically the four equivalent representations of P$[L]$ when applied to G_5: (b) The conditional probability $P_L(G_j|G_5)$ gives the probability to generate each state G_j using P$[L]$ with G_5 as the input state. (c) A shuffling method corresponding to P$[L]$ samples graphs G_j from the subset of networks with $L(G_5) = 2$ links, $\mathcal{G}_{L(G_5)} \subset \mathcal{G}$ according to $P_L(G_j|G_5)$. (d) The partition of $\mathcal{G}$ induced by P$[L]$ consists of the four distinct sets: $\mathcal{G}_0 = \{G_1\}$, $\mathcal{G}_1 = \{G_2, G_3, G_4\}$, $\mathcal{G}_2 = \{G_5, G_6, G_7\}$ (marked in blue), and $\mathcal{G}_3 = \{G_8\}$. (e) The block-diagonal stochastic matrix $\mathbf{P}^L$ gives the probability to generate any output state G_j from any given input state G_i (the block corresponding to the input state $G^* = G_5$ is marked in blue).

the multiset $p(\boldsymbol{\tau}) = [\tau]_{(i,j,t,\tau)\in\mathcal{C}}$, which contains the durations of all events in $\mathcal{C}$ including duplicate values. Different shuffling methods additionally constrain other network features, but they all conserve at least the above features.

Example II.8. The most random event shuffling possible, P$[p(\boldsymbol{\tau})]$, is the one that conserves only the events' durations and otherwise redistributes them completely

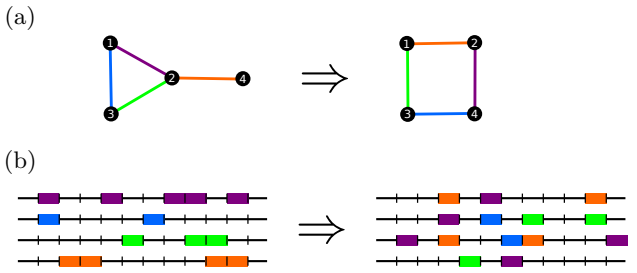


FIG. II.5: Illustration of the most random link and timeline shufflings. (a) The most random link shuffling, $\mathrm{P}[p_{\mathcal{L}}(\mathbf{\Theta})]$, completely randomizes the static graph but conserves the content of the individual timelines. (b) The most random timeline shuffling, $\mathrm{P}[\mathcal{L}, E]$, redistributes the instantaneous events between all timelines at random while conserving the static topology.

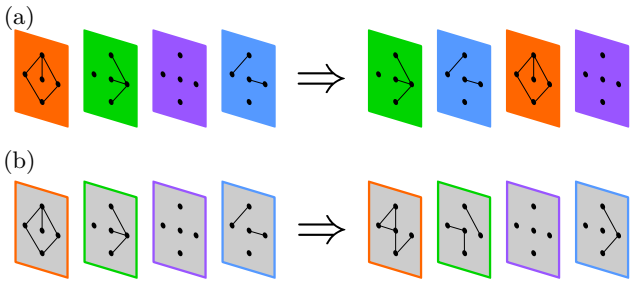


FIG. II.6: Illustration of the most random sequence and snapshot shufflings. (a) The most random sequence shuffling, $\mathrm{P}[p_{\mathcal{T}}(\mathbf{\Gamma})]$, conserves all the individual snapshot graphs but randomizes their temporal order. (b) The most random sequence shuffling, $\mathrm{P}[\mathbf{t}]$, completely randomizes each snapshot graph while conserving the temporal ordering of the snapshots.

at random, i.e. it draws the triplets (i, j, t) at random without replacement.

Example II.9. The most random instant-event shuffling is $\mathrm{P}[E]$. It draws the triplets (i, j, t) at random without replacement and conserves only the number of instantaneous events E.

We furthermore define several more restricted classes of shuffling methods that randomize specific temporal or topological aspects of a network using the two level representations introduced in Section II A 2 above.

1. *Link and timeline shufflings*

Based on the link-timeline representation (Def. II.6), we define the following two classes of shuffling methods.

Link shufflings conserve the content of the timelines, i.e. the multiset $p_{\mathcal{L}}(\mathbf{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$, but randomizes their placement. In practice, they are implemented by randomizing the links $\mathcal{L}$ in the static graph, using any shuffling method for static graphs, and redistributing the timelines $\Theta_{(i,j)} \in \mathbf{\Theta}$ on the new links without replacement. Note that link shufflings do not necessarily randomize the static topology of the network completely since the static graph shuffling may constrain any feature of G^{stat}, e.g. the nodes' degrees $\mathbf{k}$.

Example II.10. Using the Erdős-Rényi model for randomizing the static graph G^{stat} leads to the most random link shuffling possible, $\mathrm{P}[p_{\mathcal{L}}(\mathbf{\Theta})]$ [Fig. II.5(a)], while randomizing the G^{stat} using the configuration model leads to the more constrained link shuffling $\mathrm{P}[\mathbf{k}, p_{\mathcal{L}}(\mathbf{\Theta})]$, which constrains the nodes' degrees in G^{stat}.

Timeline shufflings, on the other hand, constrain the network's static topology, $G^{\mathrm{stat}} = (\mathcal{V}, \mathcal{L})$ and randomizes the content of the timelines $\Theta_{(i,j)} \in \mathbf{\Theta}$. In practice they are implemented by redistributing the (instantaneous) events in or between the timelines. Similarly to link shufflings, the timelines are not necessarily completely randomized as timeline shufflings may additionally constrain any feature of $\mathbf{\Theta}$.

Example II.11. The most random timeline shuffling, $\mathrm{P}[\mathcal{L}, E]$, is obtained by redistributing the instantaneous events in an instant-event network at random between the timelines [Fig. II.5(b)].

2. *Sequence and snapshot shufflings*

Based on the snapshot-sequence representation (Def. II.7), we define the following two classes of shuffling methods.

Sequence shufflings constrain the content of instantaneous snapshot graphs, i.e. the multiset $p_{\mathcal{T}}(\mathbf{\Gamma}) = [\Gamma^t]_{t\in\mathcal{T}}$, and randomize the order of the snapshots. They are implemented by shuffling the order of the snapshots.

Example II.12. Shuffling the temporal order of the individual snapshots completely at random leads to the most random sequence shuffling, $\mathrm{P}[p_{\mathcal{T}}(\mathbf{\Gamma})]$ [Fig. II.6(a)].

Snapshot shufflings constrain the time of each event, i.e. $\mathbf{t} = (t)_{(i,j,t)\in\mathcal{E}}$ and randomzes the individual snapshot graphs $\Gamma^t \in \mathbf{\Gamma}$. They are typically implemented by randomizing the snapshot graphs individually and independently using any shuffling method for static graphs.

Example II.13. Using the ER model to randomize each individual snapshot graph leads to the snapshot shuffling $\mathrm{P}[\mathbf{t}]$, which is the most random snapshot shuffling [Fig. II.6(b)].

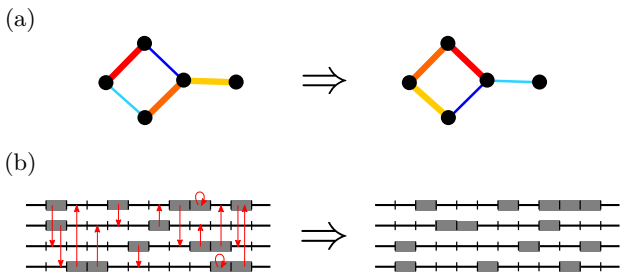


FIG. II.7: Illustration of intersections between shuffling methods. (a) The most random link-timeline intersection, $\mathrm{P}[\mathcal{L}, p_{\mathcal{T}}(\mathbf{\Gamma})]$, constrains the static topology redistributes the individual timelines on the links at random. (b) The most random timeline-snapshot intersection, $\mathrm{P}[\mathcal{L}, \mathbf{t}]$, conserves the timestamp of each instantaneous event and redistributes them at random between the existing links.

3. *Intersections of shuffling methods*

As we shall see in the following, several MRRMs exist which constrain both the content of individual timelines, i.e. $p_{\mathcal{L}}(\mathbf{\Theta})$, and the static topology, i.e. $G^{\mathrm{stat}} = (\mathcal{V}, \mathcal{L})$. This makes them intersections (Def. II.9) of link and timeline shufflings. They are typically implemented in a manner similar to link shufflings by redistributing the timelines between the links in G^{stat}, but without changing G^{stat}.

Example II.14. The intersection between the most random link shuffling, $\mathrm{P}[p_{\mathcal{L}}(\mathbf{\Theta})]$ and the most random timeline shuffling, $\mathrm{P}[\mathcal{L}, E]$, defines the most random link-timeline intersection: $\mathrm{P}[\mathcal{L}, p_{\mathcal{L}}(\mathbf{\Theta})]$ [Fig. II.7(a)]. This model constrains both the static topology and all temporal correlations on individual links, but destroys correlations between network topology and dynamics.

Other MRRMs constraint both the static topology, i.e. $\mathcal{L}$, and the timestamps of the events, i.e. $\mathbf{t}$. These are thus intersections of timeline and snapshot shufflings. They are typically implemented by exchanging the timestamps of the events inside each timeline, or alternatively by redistributing events between existing links while keeping their timestamps unchanged.

Example II.15. The intersection between the most random timeline shuffling, $\mathrm{P}[\mathcal{L}, E]$, and the most random snapshot shuffling, $\mathrm{P}[\mathbf{t}]$, defines the most random timeline-snapshot intersection: $\mathrm{P}[\mathcal{L}, \mathbf{t}]$ [Fig. II.7(b)].

4. *Compositions of shuffling methods*

The final classes of shuffling methods that we will encounter are methods that generate randomized networks by applying a pair of different shuffling methods in *composition*, i.e. by applying the second shuffling to the randomized networks generated by the first.

Not all compositions generate a microcanonical RRM however. They are e.g. not guaranteed to sample the randomized networks uniformly. But as we will show in Section V, compositions between link shufflings and timeline shufflings and between sequence shufflings and snapshot shufflings always result in a MRRM. Several such compositions have been used in the literature to produce MRRMs that randomize both topological and temporal aspects of a network at the same time (we describe and characterize them in Section V C).

Example II.16. The composition of the link shuffling $\mathrm{P}[p_{\mathcal{L}}(\mathbf{\Theta})]$ with the timeline shuffling $\mathrm{P}[\mathcal{L}, E]$ results in the MRRM $\mathrm{P}[L, E]$ which randomizes both the static topology and the temporal order of events while conserving the number of links $L = |\mathcal{L}|$ in the static graph.

Example II.17. The composition of the sequence shuffling $\mathrm{P}[p_{\mathcal{T}}(\mathbf{\Gamma})]$ with the snapshot shuffling $\mathrm{P}[\mathbf{t}]$ results in the MRRM $\mathrm{P}[p(\mathbf{A})]$ which randomizes both the topology of snapshots and their temporal order while conserving the multiset of the number of events in each snapshot, $p(\mathbf{A}) = [|\mathcal{E}^t|]_{t\in\mathcal{T}}$.

III. SURVEY OF APPLICATIONS OF RANDOMIZED REFERENCE MODELS

The applications of MRRMs for temporal networks are manifold, but all follow two main directions: (i) studying how the network and ongoing dynamical processes are controlled by the effects of temporal and structural correlations that characterize empirical temporal networks, (ii) highlighting statistically significant features in temporal networks.

(i) Dynamical processes have been studied by using data-driven models, where temporal networks are obtained from real data, while the ongoing dynamical process is modeled by using any conventional process definition [45, 74] and typically simulated numerically on the empirical and randomized temporal networks [74, 75]. One common assumption in all these models is that information can flow between interacting entities only during their interactions. This way the direction, temporal, and structural position, duration, and the order of interactions become utmost important from the point of view of the dynamical process. MRRMs provide a way to systematically eliminate the effects of these features and to study their influence on the ongoing dynamical process. This methodology has recently shown to be successful in indicating the importance of temporality, bursty dynamics, community structure, weight-topology correlations, and higher-order temporal correlations on the evolution of dynamical processes, just to mention a few examples.

(ii) MRRMs have commonly been used as null models to find statistically significant features in temporal net-

works (often termed interaction motifs) or correlations between network dynamics and node attributes. This approach is conceptually the same as using the configuration model to detect overrepresented subgraphs (termed *motifs*) in static networks [39, 76, 77]. The difference here is that the studied networks vary in time, which induces further challenges, and in particular drastically increases the number of possible null models.

We here review the main research directions and a selection of main results obtained using MRRMs to study temporal networks. We use the naming convention developed in the preceding section to provide consistent names for the shuffling methods applied in the different studies, and we classify them according to which aspects of a temporal network they randomize.

As we shall see in this section, several studies apply a single MRRM as null model analogous to standard hypothesis testing. However, in many cases we may not know how to choose the right null model, in which case it is problematic to choose an arbitrary model since results may crucially depend on this choice [3, 8, 78]. In other cases we are not interested in performing null hypothesis testing at all, but rather in investigating how a range of different features of a network affect each other or how they affect a given dynamical phenomenon. Instead of basing our analysis on a single model, we want in these cases to apply and compare a series of related MRRMs to understand how the various features and their combinations change the results. Sections IV and V develop the theoretical machinery needed to compare and order network features and MRRMs, and they provide a taxonomy of the MRRMs found in the literature which fully describes them, orders them, and characterize their effects on temporal network features. We refer to this taxonomy for detailed descriptions of each MRRM encountered in this section.

In the first three subsections of this section (Subsections III A–III C), we review studies applying MRRMs to study various dynamical processes in empirical temporal networks. In the fourth subsection (Subsection III D) applications to inferring statistically significant motifs and correlations in network dynamics will be discussed. Finally, in the last subsection (Subsection III E) we discuss a pair of recent papers that have applied MRRMs to study temporal network controllability. We will in the following include reference models that are not MRRMs (such models are briefly discussed in Section VI).

A. Contagion processes

Contagion phenomena is the family of dynamical processes that has been studied the most using MRRMs. Since epidemics, information, or influence are all transmitted by person-to-person interactions to a large extent, the approximation provided by contact-data-driven simulations are indeed closer to reality than other conventional methods based solely on analytical models. MRRMs became important in this case to help understand which temporal or structural features of real temporal networks control the speed, size, or the critical threshold of the outbreak of any kind of contagion process. In the following we will address various types of contagion dynamics ranging from simple to complex spreading processes, focusing on findings that are due to MRRMs. For detailed definitions and discussion of the different contagion processes we refer readers to the recent review by Pastor-Satorras et al. [74].

1. SI process

The susceptible-infected (SI) process is the simplest possible contagion model. Here nodes can be in two mutually exclusive states: susceptible (S) or infectious (I). Susceptible nodes (initially everyone except an initial seed node) become infected with rate β when in contact with an infected node. The single parameter β controls the speed of saturation, thus by considering the limit $\beta \to \infty$ one can simulate the fastest possible contagion dynamics on a given network. In this case the infection times correspond to the temporal distances between the seed and the nodes that get infected. This can be seen as a "light-cone" defining the horizon of propagation in the temporal network [52].

Early motivation to use RRMs of temporal networks was to understand why models of information diffusion unfold extremely slowly in various communication networks even when modeled by the fastest possible spreading model, i.e. an SI process with $\beta \to \infty$ [9]. The study introduced four MRRMs in order to quantify the contributions of topology and various temporal features of communication data to the spreading speed: (1) a model termed *configuration model* (corresponding to the composition of the link shuffling $\mathrm{P}[\mathbf{k},\chi_\lambda,p_{\mathcal{L}}(\mathbf{\Theta})]$ and the timeline-snapshot intersection $\mathrm{P}[\mathbf{w},\mathbf{t}]$—Sec. V C), removing all structural and temporal correlations while keeping only the empirical heterogeneities in the node degrees, $\mathbf{k}$, in the distribution of link weights, $p(\mathbf{w})$, and in the cumulative activity over time, $\mathbf{A}$. (2) a model termed *time shuffled* (the timeline-snapshot intersection $\mathrm{P}[\mathbf{w},\mathbf{t}]$—Sec. IV B 8), (3) the *link-sequence shuffled* model $\mathrm{P}[\mathcal{L},p_{\mathcal{L}}(\mathbf{\Theta})]$—Sec. IV B 7), and (4) the *equal-weight link-sequence shuffled* model (the link-timeline intersection $\mathrm{P}[\mathbf{w},p_{\mathcal{L}}(\mathbf{\Theta})]$—Sec. IV B 7), which eliminates all causal correlations between events taking place on adjacent links but conserves the weighted network structure and temporal correlations in individual timelines. The conclusions was that shuffling more in general makes the spreading faster, and that he bursty interaction dynamics and the Granovetterian weight-topology correlations [79] are dominantly responsible for the slow spread of information in these systems.

Effects of circadian fluctuations were studied in Ref. [9] via two canonical RRMs (Sec. VI A), where interaction times were generated by either a homogeneous or an in-

homogeneous Poisson process. The first model thus conserved the average link weights, while the second additionally conserved tyhe average activity at each point in time. The effect of circadian fluctuations could also be studied with MRRMs, as was done in a follow-up study by Kivelä et al. [13], who in addition applied a model termed *uniformly random times* (the timeline shuffling P[**w**]—Sec. IV B 4) to randomize all temporal correlations, including the circadian patterns, while conserving the aggregated structure. In order to clarify the role of network topology, they also introduced new models termed *configuration model* and *random network* (the link shufflings P[$\mathbf{k}, \chi_\lambda, p_\mathcal{L}(\mathbf{\Theta})$] and P[$\chi_\lambda, p_\mathcal{L}(\mathbf{\Theta})$], respectively—Sec. IV B 3) to randomize the static network topology while conserving all temporal correlations in individual timelines.

Another study by Gauvin et al. [14] analyzed face-to-face interaction networks and employed MRRMs to identify the effective dynamical features, responsible for driving the diffusion of epidemics in local settings like schools, hospitals, or scientific conferences. To understand the dominant temporal factors driving the epidemics in these cases, they took both a bottom-up approach by using generative network models, and a top-down approach by employing two shuffling methods and a bootstrap method. They shuffled event and inter-event durations on individual links using the model termed *interval shuffling* (the timeline shuffling P[$\boldsymbol{\pi}_\mathcal{L}(\boldsymbol{\tau})$,$\boldsymbol{\pi}_\mathcal{L}(\mathbf{\Delta}\boldsymbol{\tau})$]—Sec. IV B 4), they shuffled the timelines between existing links using the *link shuffling* model (the link-timeline intersection shuffling P[$\mathcal{L}$,$p_\mathcal{L}(\mathbf{\Theta})$]—Sec. IV B 7), and they finally bootstrapped the global distribution of event durations $p(\boldsymbol{\tau})$ while keeping the number of events $\mathbf{n}$ on each link fixed (Sec. VI C).

Perotti et al. [80] studied the effect of temporal sparsity, an entropy-based measure quantifying temporal heterogeneities on the empirical scale of average inter-event durations $\langle \Delta\tau^m_{(i,j)} \rangle$. As a reference model the authors used the timeline shuffling P[**w**] (Sec. IV B 4). They showed via the numerical analysis of several temporal datasets and using analytical calculations that there is a linear correspondence between the temporal sparsity of a temporal network and the slowing down of a simulated SI process.

A unique temporal interaction dataset was studied by Rocha et al. [10], which recorded the interaction events of sex sellers and buyers in Brazil. The system is a temporal bipartite network where connections only exist between sellers and buyers. Using this dataset the authors studied, among other questions, the effects of temporal and structural correlations on simulated SI (and SIR) processes. They introduced three different MRRMs imposing a bipartite network structure. Their first model, *random topological* (the metadata link shuffling P[$\mathbf{k}$,$p_\mathcal{L}(\mathbf{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_\mathcal{L}$]—Sec. IV B 9), was used to destroy any structural correlations in the bipartite structure while keeping temporal heterogeneities in the individual timelines unchanged. Conversely, their second null model, termed *random dynamic* (the timeline-snapshot intersection P[**w**,**t**]—Sec. IV B 8), destroyed all the temporal structure except global activity patterns, but kept the weighted (bipartite) network structure unchanged. Their third model, *random dynamic topological* was generated as the composition of the two others (Sec. V C). Interestingly, they observed that bursty patterns accelerate the spreading dynamics, contrary to other studies [9, 11, 13, 14, 80]. At the same time they showed that structural correlations slow down the dynamics in the long run, and by applying the two reference models at the same time, that bursty temporal patterns and structural correlations together slows spreading initially and speeds it up for later times. The authors arrived at the same conclusion using SIR model dynamics. Note that the accelerating effect of burstiness in this case was explained later by the non-stationarity of the temporal network [17, 26]

2. SIR and SIS processes

The Susceptible-Infected-Recovered (SIR) and Susceptible-Infected-Susceptible (SIS) processes are two other dynamical processes that have been widely studied on temporal networks using MRRMs. In addition to the S→I transition of the SI process, in the SIR (SIS) process infected nodes transition spontaneously to a recovered, R (susceptible, S), state with rate γ (or after a fixed time θ), after which they cannot (can) be re-infected. These processes are characterized by the basic reproduction number $R_0 = \beta/\gamma$ and display a phase-transition between a non-endemic and an endemic phase. An analogy with information diffusion can easily be drawn, where the infection is associated to the exposure to a given information, while spontaneous recovery mimics that the agent later forgets the given information.

One of the first studies addressing SIR dynamics using MRRMs was published by Miritello et al. [11] and investigated mobile phone communication networks. They used two reference models. The first was the timeline-snapshot intersection P[**w**,**t**] (Sec. IV B 8), used to study the effects of bursty interaction dynamics on global information spreading. Their second null model applied a local shuffling scheme that cannot evidently be interpreted as a MRRM for networks since it considers only local information and not the whole network: Both reference models preserve the link weights **w**, the duration of interactions, and also the circadian rhythms of human communications. As their first conclusion, they realized that relay times depend on two competing properties of communication. While burstiness induces large transmission times, thus hindering any possible infection, causual interaction patterns translate into an abundance of short relay times, favoring the probability of propagation.

Génois et al. studied the effects of sampling of face-to-face interaction data on data-driven simulations of SIR

and SIS processes [24], and proposed an algorithm for compensating for the sampling effect by reconstructing surrogate versions of the missing contacts from the incomplete data, taking into account the network group structure and heterogeneous distributions of $n_{(i,j)}$, $\tau^m_{(i,j)}$, and $\Delta\tau^m_{(i,j)}$. Using the reconstructed data instead of the sampled data allowed to trade in a large underestimation of the epidemic risk by a small overestimation; here the epidemic risk was quantified by the fraction of recovered (susceptible) nodes in the stationary state and the probability that the epidemics reached at least 20% of the population. They used MRRMs to investigate and explain the reasons for the small overestimation of the epidemic risk when using the reconstructed networks. They applied following reference models: a method termed *link shuffling* (the link-timeline intersection P[$\mathcal{L}, p_{\mathcal{L}}(\mathbf{\Theta})$]—Sec. IV B 7); their *CM-shuffling* (the blockmodel link shuffling P[$p_{\mathcal{L}}(\mathbf{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$]—Sec. IV B 9); a bootstrap method, resampling $p(\mathbf{n})$, $p(\boldsymbol{\tau})$, $p(\mathbf{\Delta}\boldsymbol{\tau})$, and $p(\mathbf{t}^1)$ (Sec. VI C); and they finally applied P[$p_{\mathcal{L}}(\mathbf{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$] in composition with the bootstrap method. This allowed them to conclude that the overestimation was due to higher order temporal and structural correlations in the empirical temporal networks, which however are notoriously hard to quantify and to model.

The effect of the timing of the first and last activations of the links in a network on epidemic spreading was demonstrated by Holme and Liljeros [17] using twelve empirical temporal networks. They investigated an ongoing link picture where the lifetime of social ties is irrelevant as links are assumed to be created and end before and after the observation period; and a link turnover picture where social links are assigned with a lifetime being created and dissolved during the observation. To understand which case is more relevant for modeling epidemic spreading, they defined three deterministic *poor man's reference models* [47] (see Sec. VI B). Their first reference model conserved the timings of the first and last events on each link, $\mathbf{t}^1$, $\mathbf{t}^w$, respectively, as well as the links' weights $\mathbf{w}$, and equalized all inter-event durations in the timelines, eliminating the effects of heterogeneous inter-event durations. Their second and third models aimed to neutralize the effects of the beginning and ending times of active intervals, thus they shifted the active periods of each link either to the beginning or to the end of the observation period, i.e. they set $t^1_{(i,j)} = t_{\min}$ or $t^w_{(i,j)} = t_{\max}$ for all $(i,j) \in \mathcal{L}$, respectively, while keeping the original sequence $\mathbf{\Delta}\boldsymbol{\tau}$ of inter-event durations on the links. The authors presented an exhaustive analysis by simulating SIR and SIS processes on each dataset using the original event sequences, and each reference model.

Valdano et al. [25] proposed an infection propagator approach to compute the epidemic threshold of discrete time SIS (and SIRS) processes on temporal networks. Their aim was to account for more realistic effects, namely a varying force of infection per contact, the possibility of waning immunity, and limited time resolution of the temporal network. To better understand the effects of temporal aggregation and correlations on the estimation of the epidemic threshold in face-to-face interaction datasets recorded in school settings, they employed three MRRMs: *reshuffle* (the snapshot shuffling P[$p_{\mathcal{T}}(\mathbf{\Gamma})$]—Sec. IV B 5), *reconfigure* (the timeline-snapshot intersection P[$\mathbf{w}$,$\mathbf{t}$]—Sec. IV B 8), and *anonymize* (the snapshot shuffling P[$\mathbf{iso}(\mathbf{\Gamma})$]—Sec. IV B 6). They measured, for different recovery rates, how the epidemic threshold changed as a function of the aggregation time window relative to the case with the highest temporal resolution. They considered two different aggregation strategies: where the link weights (i) were or (ii) were not considered. Finally, they considered a fourth, heuristic, reference model, which shuffled the snapshot order, but only within a given number of slices, this way keeping control on the length of temporal correlations it destroyed (see description in Sec. VI B).

Finally, there has been a single study using MRRMs with rumor spreading dynamics [18]. It considered the Daley-Kendall model, which is very similar to the SIR model with the exception that nodes do not recover spontaneously but via interactions with other infected or recovered (stifler) nodes. The aim of this study was to understand the effects of memory processes, inducing repeated interactions between people, on the global mitigation of rumors in large social networks. Using a mobile phone communication dataset they utilized a specific directed temporal network snapshot shuffling, P[$\mathbf{d}_{\rightarrow}$], which constrained the instantaneous out-degree $d^m_{i\rightarrow}$ of each node in each snapshot (see Supplementary Table S2). In practice this amounted to randomizing the called person for each event in order to eliminate the effects of repeated interactions over the same link. This MRRM randomized the topological and temporal correlations in the network, destroyed link weights, and increased the static node degrees considerably. Results were confronted with corresponding model simulations, which verified that memory effects play the same role in data-driven models as was observed in the case of synthetic model processes, namely they keep rumors local due to repeated interactions over strong ties.

3. Threshold models

A third family of spreading processes are *complex contagion* processes, which are often used to model social contagion. These models capture the effects of social influence, which is considered via a non-linear mechanism for contaminating neighboring nodes (typically a threshold mechanism). In the conventional definition of threshold models [81] nodes can be either of two mutually exclusive states, non-adopter (i.e. susceptible) – initially all but one node – and adopter (i.e. infectious) – initially a randomly selected *seed* node – and each node i is assigned a threshold ϕ_i defining the number k^I_i or fraction k^I_i/k_i of adopter neighbors necessary to make the node (with total degree k_i) adopt. We refer to the first variant as

the Watts threshold model with *absolute* thresholds, and the second as the Watts threshold model with *relative* thresholds. The central question here is the condition needed to induce a large adoption cascade that spreads all around the network. These models are highly constrained by the network structure and dynamics as the distribution of individual thresholds determine the conditions for global cascades. This is fundamentally different from the SIR type of dynamics (called *simple contagion processes*) which are highly stochastic, driven by random infection and recovery. The conventional threshold model introduced by Watts [81], and other related dynamical processes have been thoroughly studied on static networks, however their behavior on temporal networks has been addressed only recently by studies using RRMs.

Karimi and Holme [15] studied two different threshold models on six empirical datasets of time-resolved human interactions. They employed two MRRMs: one called *time reshuffle* (the timeline-snapshot intersection P[**w**,**t**]—Sec. IV B 8) and anoter termed *Erdős-Rényi* (the link shuffling P[$p_{\mathcal{L}}(\mathbf{\Theta})$]—Sec. IV B 3). Application of P[**w**,**t**] allowed them to conclude that burstiness plays an important role on how large cascades can appear in complex contagions. Backlund et al. [19] also studied the effects of temporal correlations on cascades in slightly different threshold models on temporal networks. They applied two MRRMs to four different temporal interaction datasets. They used the P[**w**,**t**] (Sec. IV B 8) model to destroy all temporal correlations while keeping circadian fluctuations, and introduced another model, P[**per**($\mathbf{\Theta}$)] (Sec. IV B 4), that randomly shifts each individual timeline using periodic boundary conditions to keep all temporal correlations inside each timeline and destroy correlations between events on adjacent links as well as circadian fluctuations. They found that the removal of temporal correlations using P[**w**,**t**] facilitates spreading. This way they concluded that burstiness negatively affects the size of the emerging cascades. At the same time, they found that higher order temporal-structural correlations, removed by P[**per**($\mathbf{\Theta}$)], facilitate the emergence of large cascades.

A somewhat different picture was proposed by Takaguchi et al. [16], where the authors used a threshold model denoted *history dependent contagion*. This model is an extension of an SI process with a threshold mechanism. Here each node has an internal variable measuring the concentration of pathogen and is increased by unity after a stimuli arrived via temporal interactions with infected neighbors. However, this concentration decays exponentially as function of time in the absence of interaction with infected nodes. A node becomes infected if its actual concentration reaches a given threshold, after which it remains in the infected state. They simulated this model on two different temporal interaction networks and measured the fraction of adopters as function of time. In order to identify the effects of bursty interaction patterns they used a model called *randomly permuted times* (the timeline-snapshot intersection P[**w**,**t**]—Sec. IV B 8), which led to slower spreading dynamics. From this they argued that burstiness increases the speed of spreading in both datasets. Furthermore, they showed through the analysis of single link dynamics, that this acceleration was mostly due to the bursty patterns on separate links and not due to correlations between bursty events on adjacent links or to the overall structure of the network.

B. Random walks

Random walks are some of the simplest and most studied dynamical processes on networks. On a temporal network, a random walk is defined by a walker, which is located at a node at time t, and can be re-located to one of the node's current neighbors in each timestep. The walker chooses the neighbor to which it jumps either at random or with a probability proportional some link weight.

Starnini et al. [12] studied stationary properties of random walks on temporal networks, and used reference models to define ways to synthetically extend their temporal face-to-face interactions datasets with a limited observation length. They assumed periodic temporal boundary conditions on their empirical temporal network (their first model), with weak induced biases as discussed in an earlier paper [60]. Their second model, *SRan* (the timeline shuffling P[**w**]—Sec. IV B 4), kept all weighted features of the aggregated network, but destroyed all temporal correlations and induced Poissonian interaction dynamics. Finally, they introduced a third heuristic reference model in which they impose a delta function constraint on the number of events starting at each time step (Sec VI B), randomly drawing the pairs of nodes that interact in order to approximately conserve **n** and finally bootstrap the event durations from $p(\boldsymbol{\tau})$. This approximately conserves certain important statistical properties of the empirical event sequence, namely $p(\mathbf{n})$ and $p(\boldsymbol{\tau})$, but not $\mathbf{A}$ and $p(\mathbf{\Delta}\boldsymbol{\tau})$. They measured the *mean-first passage time* (MFPT), defined as the average time taken by the random walker to arrive for the first time at a given node starting from some initial position in the network, and the *coverage*, defined as the number of different vertices that have been visited by the walker up to time t, on both the original temporal network and synthetic sequences. They found that the results for empirical sequences deviated systematically from the mean field prediction and from the results for the reference models, inducing a slowdown in coverage and MFPT. They concluded that this slowdown is not due to the heterogeneity of the durations of conversations, but uniquely due to what they term *temporal correlations* (which, given the reference models they tested, encompasses the time-varying cumulative activity, the broad distribution of inter-event durations, and higher-order temporal correlations between different events).

Delvenne et al. [22] also addressed random walks on temporal networks. They used MRRMs in order to un-

derstand which factor is dominant in determining the relaxation time of linear dynamical processes to their stationary state. They introduced a general formalism for linear dynamics on temporal networks, and showed that the asymptotic dynamics is determined by the competition between three factors: a structural factor (i.e. community structure) associated with the spectral properties of the Laplacian of the static network, and two temporal factors associated to the shape of the waiting-time distribution, namely its burstiness coefficient (defined in [82]) and the decay rate of its tail. They demonstrated their methodology on six empirical temporal interaction networks and used two RRMs. A MRRM termed *randomized structure* (the link shuffling $\mathrm{P}[\mathbf{k}, p_{\mathcal{L}}(\mathbf{\Theta})]$—Sec. IV B 3) aimed to remove the effects of the structural correlations. The other null model, a generative reference model using a homogeneous Poisson process to generate events and constraining only G^{stat} and the mean number of events $\langle E \rangle$ (Sec. VI A), destroyed all temporal and weight correlations while conserving the static network structure, leading to the evident dominance of the network structure in regulating the convergence to stationarity.

A greedy random walk process and a non-backtracking random walk process were studied by Saramäki and Holme on eight different human interaction datasets in Ref. [23]. A greedy random walker always moves from the occupied node to one of its neighbors whenever possible. Thus its dynamics is more sensitive to local temporal correlations in the network. A non-backtracking greedy random walker is additionally forbidden to return to its previous position. Thus, it is forced to move to a new neighbor or wait until the next event which moves it to a new neighbor. The authors studied what types of temporal correlations are determinant during these dynamics by using the *time-stamp shuffling* (the timeline-snapshot shuffling $\mathrm{P}[\mathbf{w},\mathbf{t}]$—Sec. IV B 8) and measuring the coverage of the walker after a fixed number of moves. They found that after removing temporal correlations using $\mathrm{P}[\mathbf{w},\mathbf{t}]$, the walker reached considerably more nodes. They finally traced the entropy of the greedy walkers and concluded that, on average, the entropy production rates measured in the original event sequences were lower than for randomized data, indicating more predictable node sequences of visited nodes in the empirical case.

C. Evolutionary games

Evolutionary games [83] define another set of dynamical processes which have historically been studied on networks. They are analogous to several social dilemmas where the balance of local and global payoffs drive the decision of interacting agents. Any agent may choose between two strategies (Cooperation or Defection) and can receive four different payoffs (Reward, Punishment, Sucker, or Temptation). The relative values of Temptation and Sucker determines the game, where players update their strategy depending on the state of their neighbors with a given frequency and tend to find an optimal strategy to maximize their benefits.

Cardillo et al. [20] studied various evolutionary games on temporal networks and asked two questions: "Does the interplay between the time scale associated with graph evolution and that corresponding to strategy updates affect the classical results about the enhancement of cooperation driven by network reciprocity?" and "what is the role of the temporal correlations of network dynamics in the evolution of cooperation?". They analyzed two human interaction sequences, and for comparison they applied a shuffling termed *random ordered* (the sequence shuffling $\mathrm{P}[p_{\mathcal{T}}(\mathbf{\Gamma})]$—Sec. IV B 5), and the activity-driven model [84]. As a parameter to control the time-scale of the network, they varied the size of the integration time window defining a single snapshot of the temporal networks and measured the fraction of cooperators after the simulated dynamics reached equilibrium. They showed for all social dilemmas studied that cooperation is seriously hindered when agent strategies are updated too frequently relative to the typical time scale of interactions, and that temporal correlations between links are present and lead to relatively small giant components of the graphs obtained at small aggregation intervals. However, when one uses randomized or synthetic time-varying networks that preserve the original activity potentials but destroys temporal correlations, the structural patterns change dramatically. Effects of the temporal resolution on cooperation are smoothed out, and due to the lack of temporal and structural correlations, cooperation may persistently evolve even for moderately small time periods.

D. Temporal motifs and networks with attributes

Another direction of application of RRMs is to highlight significant temporal correlations or motifs in interaction signals or when the interaction sequences may correlate with additional node attributes.

For directed temporal networks, one simple application of MRRMs was introduced by Karsai et al. [5], who analyzed the correlated activity patterns of individuals, which induced bursty event trains. They found that the number of consecutive events arriving in clusters are distributed as a power-law. To identify the reason behind this observation they used a MRRM that shuffled the inter-event durations between consecutive event pairs, $\mathrm{P}[\mathbf{s}_{\rightarrow}, p(\mathbf{\Delta\tau})]$ (see Supplementary Table S2). They found that in the shuffled signal, bursty event trains were exponentially distributed, which evidently indicated that bursty trains were induced by intrinsic correlations in the original system and were not simply due to the broad distribution of inter-event durations.

In another study, Karsai et al. [6] also applied this framework to identify whether correlated bursty trains of individuals is a property of nodes or links. Using

a large mobile phone call interaction dataset, the observation was made that bursty train size distributions were almost the same for nodes and links. This suggests that such correlated event trains were mostly induced by conversations by single peers rather than by group conversations. To further verify this picture, the fraction of bursty trains of a given size emerging between a varying number of individuals were calculated in the empirical event sequence and in shuffled networks generated using a MRRM] where the receivers of calls were shuffled between calls of the actual caller (P[$\mathbf{w}$,$p(\boldsymbol{\Delta\tau})$]—Supplementary Table S2). This reference model leaves the timing of each event unchanged, thus leading to the observation of the same bursty trains, and it keeps the instantaneous and static out-degrees of individuals. However, since the receivers are shuffled, potential correlations that induce bursty trains on single links are eliminated. Results showed that the fraction of single link bursty trains drops from $\sim 80\%$ to $\sim 20\%$ after shuffling in call and SMS sequences. This supports the hypothesis that single link bursty trains are significantly more frequent than one would expect from the null hypothesis, which is then rejected.

Real temporal networks commonly reveal more complicated temporal motifs, whose detection was first addressed by Kovanen et al. [4]. They proposed a method to identify mesoscale causal temporal patterns in interaction sequences where events of nodes do not overlap in time. This framework can be used to identify overrepresented patterns, called temporal motifs which are not only similar topologically but also in the temporal order of the events. RRMs are crucial in this framework for quantifying the significance of different temporal motifs. They used *time-shuffling* (the timeline-snapshot intersection P[$\mathbf{w}$,$\mathbf{t}$]—Sec. IV B 8), and they introduced a non-maximum-entropy reference model which biases the sampling of the temporal networks defined by P[$\mathbf{w}$,$\mathbf{t}$] in order to keep some temporal correlations in the sequence (see Sec. VI B). To do so, they selected randomly for each event in a motif m other events from the sequence and chose the one which was the closest in time to the original event in focus. If $m = 1$ the model is identical to P[$\mathbf{w}$,$\mathbf{t}$], while the larger m is the more candidate events there are, thus the more likely it is to find one close to the original event. They furthermore suggested that to remove causal correlations from the sequence, one may simply reverse the interaction sequence and repeat the motif detection procedure (see Sec. VI B). They used these reference models in the same spirit as the configuration model is typically used to identify motifs in static networks [76, 77]. Here, applying P[$\mathbf{w}$,$\mathbf{t}$] and its biased version as null models to detect motifs consisting of three events, they found that motifs between two nodes, i.e. bursty link trains, are the most frequent, and motifs which consist of potentially casually correlated events are more common than non-causal ones.

In another study by the same authors [7], the same methodology was used to identify motifs in temporal networks where nodes (individuals) were assigned with metadata attributes like gender, age, and mobile subscription types. Beyond the P[$\mathbf{w}$,$\mathbf{t}$] model (Sec. IV B 8), the authors introduced the metadata MRRM termed *node type shuffled data* (the metadata shuffling P[$G, p(\boldsymbol{\sigma}), \Sigma_{\mathcal{L}}$]—Sec. IV B 9), which shuffles single attributes between nodes. In addition, they applied the biased version of P[$\mathbf{w}$,$\mathbf{t}$] introduced in [4] (see Sec. VI B), which accounts for the frequency of motif emergence in the corresponding static weighted network without considering node attributes. Using this non-maximum-entropy reference model and the two MRRMs they found gender-related differences in communication patterns and showed the existence of temporal homophily, i.e. the tendency of similar individuals to participate in communication patterns beyond what would be expected on the basis of their average interaction frequencies.

The dynamics of egocentric network evolution was studied by Kikas et al. [85], where they used a large evolving online social network to analyze bursty link creation patterns. First of all they realized that link creation dynamics evolve through correlated bursty trains. They verified this observation by comparing the distribution of inter-event durations (measured between consecutive link creation events) to those generated by the directed-network MRRM P[$\mathbf{s}_{\rightarrow}, p(\boldsymbol{\Delta\tau})$] (see Supplementary Table S2), where inter-event durations were randomly shuffled. In addition, they classified users based on their link creation activity signals (where activity was measured as the number of new links added within a given month). They showed that bursty periods of link creation are likely to appear shortly after the creation of a user account, or when a user actively use free or paid services provided by the online social service. In order to verify these correlations they used a reference model where they shuffled link creation activity values between the active months of a given user and found considerably weaker correlations between the randomized link creation activity signals and service usage activity signals of people.

Finally, in a different framework, a special kind of metadata reference model was also used by Karsai et al. [86] to demonstrate whether the effect of social influence or homophily is dominating during the adoption dynamics of online services on static networks. This reference model did not consider randomizing the temporal networks, but rather node attributes linked to the dynamics of the game (i.e. a purely metadata MRRM – Sec. IV B 9 – coupled with a dynamical process on the network); we include it in this survey to demonstrate the scope of maximum entropy shuffling methods beyond randomizing structural network features. The authors used a reference model where they shuffled all adoption times between adopted nodes and confronted the emerging adoption rates of innovator, vulnerable, and stable adopters (for definitions see [81, 86]) to the adoption rates observed in the empirical system. They found that after

shuffling the rate of innovators considerably increased, while the rate of influence driven (vulnerable and stable) adoptions dropped. This verified that adoption times matters during real adoption dynamics, thus the social spreading process was predominantly driven by social influence. Note that in this case the network was static and shuffling was applied on the observed dynamical process.

E. Network controllability

We finally mention two recent studies of the controllability of temporal networks that have leveraged MRRMs. Control of a dynamical system aims at guiding a system to a desired state by designing the inputs to the system [87]. Although control theory has a long history as a branch of engineering applied to diverse subjects, it was only recently that we saw a general theory of the controllability of the systems in which elements interact in a networked manner [88].

It is natural to think of extending the theory for static networks to temporal networks. Pósfai and Hövel made the first study in this direction, in which they considered a discrete-time linear dynamical system with time-varying interactions [50]. Their focal measure of controllability is the size of the structural controllable subspace. The structural controllable subspace is defined by the subset of nodes satisfying that any of their final states at time t is realizable from any initial state in at most a number τ of time-steps by appropriately tuning the non-zero elements of the adjacency and input matrices as well as the input signals. First, they proved a theorem stating that a node subset is a structural controllable subspace if and only if any node in the subset are connected to disjoint time-respecting paths from the nodes receiving the input signals. This theorem implies that, keeping the same average instantaneous degree, the temporal network with a heavy-tailed distribution of instantaneous node degrees, $\boldsymbol{\pi}_{\mathcal{T}}(\mathbf{d})$, is more difficult to control than a network with a homogeneous $\boldsymbol{\pi}_{\mathcal{T}}(\mathbf{d})$ because the presence of hubs in snapshots decreases the number of disjoint time-respecting paths. They examine this theoretical argument by comparing the structural controllable subspace for an empirical temporal network to the ones produced by five MRRMs: one termed *random time* (the timeline shuffling P[$\mathbf{w}$]—Sec. IV B 4), one termed *random network* (P[$\mathbf{t}$]—Sec. IV B 6), one termed *degree-preserved network* (P[$\mathbf{d}$]—Sec. IV B 6), and two MRRMs both termed *shuffled time* (the sequence shufflings P[$p_{\mathcal{T}}(\boldsymbol{\Gamma})$] and P[$p_{\mathcal{T}}(\boldsymbol{\Gamma})$,$\boldsymbol{\chi}_{\mathbb{N}^+}(\mathbf{A})$]—Sec. IV B 5). The sizes of the maximum structural controllable subspace for P[$\mathbf{w}$,$\mathbf{t}$] and P[$\mathbf{t}$] were generally larger than that for the original network. This result suggests that the homogenization of $\boldsymbol{\pi}_{\mathcal{T}}(\mathbf{d})$ and thus the elimination of hubs in snapshots increases the controllability of networks, which is consistent with the theoretical argument. For the other two MRRMs, the controllability of networks generated by P[$\mathbf{d}$], which conserves the instantanteous degrees, is almost the same as the original network, and networks generated by P[$p_{\mathcal{T}}(\boldsymbol{\Gamma})$,$\boldsymbol{\chi}_{\mathbb{N}^+}(\mathbf{A})$], which randomizes the temporal order of the snapshots, has a slightly lower controllability than the original network. These results imply that the higher-order structural correlations in snapshots have little effect on network controllability and that the temporal correlations over successive snapshot present in the original network contribute to enhance the controllability to some extent.

Recently, Li et al. [56] showed that temporal networks have a fundamental advantage in controllability compared to their static network counterparts. They compared the time and energy required to achieve full structural controllability of the network, when driving nodes in the sequence of snapshots or the single aggregated network. The numerical experiments on multiple empirical networks demonstrated that temporal networks can be fully controlled more efficiently in terms of both time and energy than their static counterparts. They argued that this advantage comes from temporality itself, but not from particular temporal features, by showing that a set of different reference models achieve more efficient controllability than their aggregated counterparts. The models employed were: *randomly permuted times* (P[$\mathbf{w},\mathbf{t}$]—Sec. IV B 8), *randomized edges* (P[$\mathbf{k}, p_{\mathcal{L}}(\boldsymbol{\Theta})$]—Sec. IV B 3), and *randomized edges with randomly permuted times* (the composition of the two—Sec. V C), as well as time reversal (Sec. VI B).

IV. CHARACTERIZING AND ORDERING MICROCANONICAL RANDOMIZED REFERENCE MODELS

Some MRRMs randomize more (i.e. conserve less structure) than others. We here formalize this notion which allows us to compare MRRMs, which will let us build hierarchies between them and between different network features (Section IV A). Such hierarchies turn out to be useful for classification of MRRMs and for understanding how they affect different network features. We will use this in Section IV B to build a taxonomy of MRRMs found in the literature which orders them and characterizes their effects on temporal network features.

A. Theory: Hierarchies of MRRMs

We here develop the theory needed to compare and order MRRMs. To keep the presentation as accessible as possible, we have relegated proofs to Appendix VIII.

The central concept for building hierarchies of MRRMs is that of *comparability*.

Definition IV.1. *Comparability.* We will write P[$\mathbf{x}$] $\leq$ P[$\mathbf{y}$] if the set of states (e.g. temporal networks) that can be reached by applying P[$\mathbf{x}$] to any state is always contained in the set reached by P[$\mathbf{y}$], that is if $\mathcal{G}_{\mathbf{x}(G)} \subseteq$

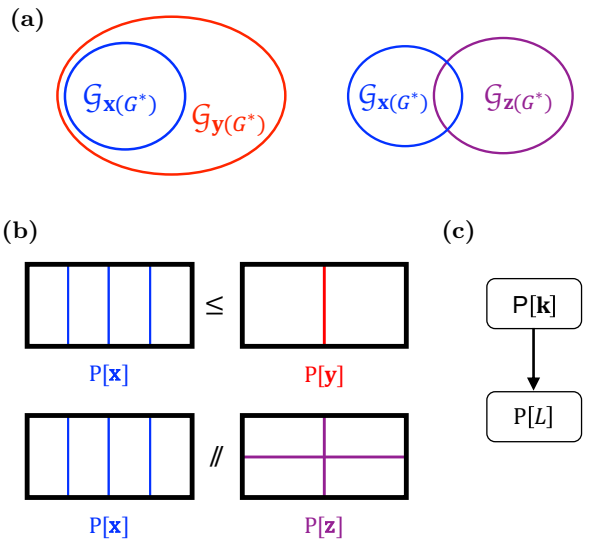


FIG. IV.1: **Comparability.** (a) Two MRRMs satisfy $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$ if the ensemble $\mathcal{G}_{\mathbf{x}(G^*)}$ generated by $\mathrm{P}[\mathbf{x}]$ is contained in the ensemble $\mathcal{G}_{\mathbf{y}(G^*)}$ generated by $\mathrm{P}[\mathbf{y}]$ for all $G^* \in \mathcal{G}$. Conversely, if the ensembles generated by two MRRMs $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{z}]$ only overlap partially, the MRRMs are not comparable, $\mathrm{P}[\mathbf{x}] \mathbin{/\!/} \mathrm{P}[\mathbf{z}]$. (b) In terms of partitions, $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$ means that $\mathrm{P}[\mathbf{x}]$ is finer than $\mathrm{P}[\mathbf{y}]$. Conversely, if $\mathrm{P}[\mathbf{x}] \mathbin{/\!/} \mathrm{P}[\mathbf{z}]$, neither partition is a refinement of the other. (c) Hasse diagram depicting the hierarchy of the two MRRMs $\mathrm{P}[\mathbf{k}]$ and $\mathrm{P}[L]$: the arrow indicates that $\mathrm{P}[\mathbf{k}]$ is finer than $\mathrm{P}[L]$.

$\mathcal{G}_{\mathbf{y}(G)}$ for all $G \in \mathcal{G}$. We say that $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ are *comparable* if $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$ or $\mathrm{P}[\mathbf{x}] \geq \mathrm{P}[\mathbf{y}]$.

The definition of comparability gives a precision notion that one MRRM randomizes more than another [Fig. IV.1(a)]. In practice it is often difficult to show directly that the ensembles of states generated by one model are always subsets of those generated by another. To compare MRRMs, we shall instead use that if the features that one MRRM constrains can be written as a function of the features constrained by another then the two MRRMs are comparable.

Proposition IV.1. *Equivalence between comparability of MRRMs and a functional relation between their features.* Two MRRMs $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ are comparable and $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$ if and only if there exists a function $\mathbf{f}$ for which $\mathbf{y}(G) = \mathbf{f}(\mathbf{x}(G))$ for all states $G \in \mathcal{G}$.

The following simple example illustrates how this proposition can be applied to order MRRMs.

Example IV.1. In the space of all static graphs with N nodes, one can define the MRRMs corresponding to the Erdős-Rényi random graph model [1], $\mathrm{P}[L]$, and the Maslov-Sneppen model [38], $\mathrm{P}[\mathbf{k}]$. We have $L = \sum_i k_i/2$, so L is a function of $\mathbf{k}$ and $\mathrm{P}[L] \geq \mathrm{P}[\mathbf{k}]$ by Proposition IV.1. Conversely, $\mathbf{k}$ is not a function of L as networks with different degree sequences can have the same number of links, so $\mathrm{P}[L] \not\leq \mathrm{P}[\mathbf{k}]$ which means that $\mathrm{P}[L]$ and $\mathrm{P}[\mathbf{k}]$ are not equivalent (i.e. $\mathrm{P}[L] \neq \mathrm{P}[\mathbf{k}]$).

Representing MRRMs as partitions (see Def. II.11) provides a useful and intuitive way of thinking about comparisons of MRRMs, because the MRRM comparison relation is exactly the natural comparison relation of the partitions.

Proposition IV.2. *Equivalence with partition refinements.* $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$ if and only if the partition $\mathcal{G}_{\mathbf{x}}$ is finer than $\mathcal{G}_{\mathbf{y}}$.

Borrowing the terminology from the theory of set partitions, we say for $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$ that $\mathrm{P}[\mathbf{x}]$ is *finer* than $\mathrm{P}[\mathbf{y}]$ and equivalently that $\mathrm{P}[\mathbf{y}]$ is *coarser* than $\mathrm{P}[\mathbf{x}]$. We will also refer to $\mathrm{P}[\mathbf{x}]$ as a *refinement* of $\mathrm{P}[\mathbf{y}]$ and to $\mathrm{P}[\mathbf{y}]$ as a *coarsening* of $\mathrm{P}[\mathbf{x}]$. Figure IV.1(b) illustrates the concept of comparability in terms of partitions.

The partition representation is especially useful here as the properties of refinements of set partitions are inherited to the comparison relation of the MRRMs. For example, we can now see that the use of the notation $\leq$ is appropriate as the relation it denotes is indeed a partial order. This follows immediately from the fact that partition refinement relations give partial orders [73]. As with any partially ordered set, one can draw Hasse diagrams to display the relationships between different MRRMs, and this turns out to be a convenient way of visually organizing the various MRRMs found in the literature [Fig. IV.1(c); see also Section IV B].

1. *The space of MRRMs*

The set partitions always have uniquely defined minimum and maximum partitions, and these are meaningful in the case of MRRMs. We call them the zero and unity elements.

Definition IV.2. *Zero and unity elements.* The *zero element*, $\mathrm{P}[0] = \mathrm{P}[G]$, is the MRRM which shuffles nothing, i.e. the one that always returns the input network and where the feature returns the entire temporal network. The *unity element*, $\mathrm{P}[1]$, is the MRRM that shuffles everything, i.e. the one that returns all networks in the state space with equal probability and where the feature is constant and does not depend on the input.

The zero element corresponds to the partition where each network is in its own set and the unity element to the partition where there is only a single set. The zero and unity elements are always in the top and bottom of any hierarchy of MRRMs, respectively: $\mathrm{P}[0] \leq \mathrm{P}[\mathbf{x}] \leq \mathrm{P}[1]$ for any $\mathbf{x}$.

Example IV.2. We continue from Example IV.1, limiting the state space to the set of simple graphs consisting of 3 nodes, $\mathcal{V} = \{1, 2, 3\}$, and 2 links. There are three such graphs: $\{G_1, G_2, G_3\}$ [Fig. IV.2(a)]. Since the

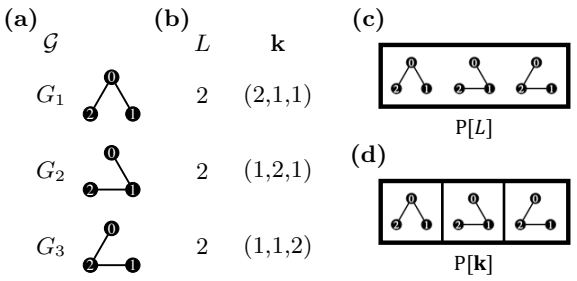


FIG. IV.2: **Example of zero and unity elements in a space of MRRMs.** (a) State space $\mathcal{G}$ consisting of all simple graphs with three nodes and two edges. (b) Values of the features constrained by the two MRRMs P[L] and P[$\mathbf{k}$] for each state $\mathcal{G}_i \in \mathcal{G}$. (c) Partition of $\mathcal{G}$ induced by P[L]. (d) Partition of $\mathcal{G}$ induced by P[$\mathbf{k}$].

number of links is the same in all graphs [Fig. IV.2(b)], the partition of the ER model contains only one set $\mathcal{G}_L = \{\{G_1, G_2, G_3\}\}$ [Fig. IV.2(c)]. However, the degree sequences of the networks differ [Fig. IV.2(b)], so the partition related to the Maslov-Sneppen model separates all networks $\mathcal{G}_\mathbf{k} = \{\{G_1\}, \{G_2\}, \{G_3\}\}$ [Fig. IV.2(d)]. For this state space the Maslov-Sneppen model is thus the zero element P[$\mathbf{k}$] = P[0] and the ER model is the unity element P[L] = P[1].

As we shall see, rich hierarchical structure can be found between the zero and unity elements. Again, the set partition representation gives us a glimpse of the theoretical understanding of this structure. The total number of possible MRRMs for a given state space $\mathcal{G}$ is the same as the number of possible partitions of the space. This is given by the Bell number B_Ω [89], which grows faster than exponentially with the state space size $\Omega = |\mathcal{G}|$ [90]. We also know that even though the number of MRRMs in the hierarchy can be large, it can only be relatively flat as compared to this number: The largest possible number of MRRMs all satisfying a total order (i.e. for which $\mathrm{P}[\mathbf{x}_1] \geq \mathrm{P}[\mathbf{x}_2] \geq \mathrm{P}[\mathbf{x}_3] \ldots$) is the maximum chain length in the set of partitions of the state space, which is equal to $\Omega + 1$. Thus, if we restrict ourselves to selecting a collection of MRRMs that is totally ordered [8], we can at most include an exponentially vanishing part of the possible MRRMs. In practice these theoretical limitations are not of much concern as the number of possible networks, Ω, typically is extremely large and it is not possible in practice to explore even a small fraction of all possible MRRMs for a given state space. Of more practical concern is the fact that we are often interested in studying a set of MRRMs that do not satisfy a linear order. In the context of temporal networks, this is notably the case for MRRMs that randomize temporal features and MRRMs that randomize topological features of a network [13].

Intersections of MRRMs (Def. II.9) play an important role in hierarchies of MRRMs: for two MRRMs, P[$\mathbf{x}$] and P[$\mathbf{y}$], their intersection P[$\mathbf{x}, \mathbf{y}$] defines the maximally random MRRM that shuffles less than both of them. In terms of set partitions, the partition of $\mathcal{G}$ induced by P[$\mathbf{x}, \mathbf{y}$] is trivially given by the set of pairwise intersections between the $\mathbf{x}$-equivalence classes and the $\mathbf{y}$-equivalence classes, i.e. $\mathcal{G}_{(\mathbf{x},\mathbf{y})}(G) = \mathcal{G}_\mathbf{x}(G) \cap \mathcal{G}_\mathbf{y}(G)$ for all $G \in \mathcal{G}$, and $\Omega_{(\mathbf{x}^*,\mathbf{y}^*)} = |\mathcal{G}_{\mathbf{x}^*} \cap \mathcal{G}_{\mathbf{y}^*}| = \sum_{G\in\mathcal{G}} \delta_{\mathbf{x}(G),\mathbf{x}^*}\delta_{\mathbf{y}(G),\mathbf{y}^*}$ (Fig. IV.3).

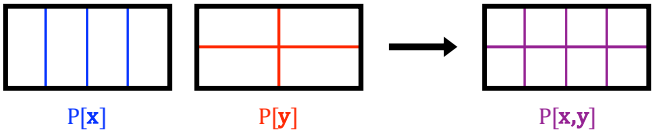


FIG. IV.3: **Intersection of MRRMs.** The intersection P[$\mathbf{x}, \mathbf{y}$] of two MRRMs, P[$\mathbf{x}$] and P[$\mathbf{y}$], shuffles less than either of the two. In terms of partitions, P[$\mathbf{x}, \mathbf{y}$] produces the greatest lower bound of P[$\mathbf{x}$] and P[$\mathbf{y}$].

The effects of intersection with the zero and unity elements are also easy to see. The unity is a neutral element that has no effect on the intersection P[$\mathbf{x}, 1$] = P[$\mathbf{x}$], because adding a constant to the feature function output does not affect the partitioning of the networks at all. The zero is an absorbing element P[$\mathbf{x}, 0$] = P[0], because adding extra information to the feature function that already contains the full network doesn't change anything. In fact, from set partitions we know that the intersection gives the greatest lower bound of the two partitions [73].

B. Taxonomy of temporal network MRRMs found in the literature

Using the theory developed above, we now describe and classify MRRMs found in the literature. We also introduce several new MRRMs which we will use in the walkthrough example in Section VII. We use our canonical naming convention (Def. II.10) to provide unambiguous names for the MRRMs, and we order them hierarchically and describe which temporal network features they conserve.

We separate the descriptions of event shufflings which shuffle the events in temporal networks while conserving their durations, and instant-event shufflings which shuffle instant-event temporal networks (Sec. II C). (Note though that it is possible to randomize the event durations of a temporal network by first discretizing its events and then shuffling this instant-event network using an instant-event shuffling—see Supplementary Note 1).

The following subsections list and describe the different MRRMs. For each MRRM we provide a canonical name as well as an informal name that may be easier to retain and use outside of the formal definition. We furthermore describe the features that the MRRM constrains, give references to the literature, and provide details on how the MRRM is implemented algorithmically. A comprehensive table (Table IV.1) provides detailed definitions

of temporal network features of interest, and a pair of tables at the end of this section (Tables IV.2 and IV.3) show how each MRRM affects them.

Subsection IV B 1 introduces temporal network features needed to describe and characterize the MRRMs described here. Subsection IV B 2 presents the coarsest, i.e. the maximally random, instant-event and event shufflings. Subsections IV B 3–IV B 6 present the four restricted classes of shufflings defined in Section II C, namely link shufflings (IV B 3), timeline shufflings (IV B 4), sequence shufflings (IV B 5), and snapshot shufflings (IV B 6). Subsection IV B 7 describes MRRMs that are intersections of link and timeline shufflings and Subsection IV B 8 describes intersections of timeline and snapshot shufflings. Subsection IV B 9 introduces and describes MRRMs that use additional metadata on nodes.

1. *Important temporal network features*

Many MRRMs constrain features that can be described as an ordered sequence of lower-dimensional features, e.g., the degree sequence of a static graph, $\mathbf{k}$, is given by the sequence of the individual node degrees k_i. Other features of interest are those defined as functions of such a sequence. The most important in practice are empirical distributions of feature values, e.g. the degree distribution $p(\mathbf{k})$, and moments, e.g. the mean degree $\mu(\mathbf{k})$.

We begin by introducing the ordered sequence of a collection of features. It retains both the values of the individual features and what they designate in the network. A MRRM that constrains such a sequence thus produces reference networks with exactly the same values and configuration of these features as in the input network. In order to make the notation simpler, and without loss of generality, we will assume that features returning values of multiple named entities, such as nodes or links, return them as sequences that have an arbitrary but fixed order.

Definition IV.3. *Sequence of features.* A *sequence* of features is a tuple $\mathbf{x} = (x_q)_{q\in\mathcal{Q}}$ of individual features ordered according to an arbitrary but fixed index $q \in \mathcal{Q}$.

The individual features x_q may be any functions, e.g. scalar functions, sequences of other features (i.e. vectors), or graphs. When the individual features are vectors we will generally use boldface to indicate this, i.e. $\mathbf{x}_q = (x_q^r)_{r\in\mathcal{R}_q}$. In this case we will refer to $\mathbf{x} = (\mathbf{x}_q)_{q\in\mathcal{Q}}$ as a *sequence of sequences*.

Typically, each x_q depends on a different part of the temporal network such as a node i, a link (i, j), or a time t. We shall in the following use a subscript to index individual topological features (e.g. x_i or $x_{(i,j)}$ for a feature of a single node or link, respectively) and superscript for temporal ones (e.g. x^t for a feature of a single snapshot). Such features are generally scalar and are assembled into a sequence that runs over all values of the index, i.e. over all nodes $i \in \mathcal{V}$, all links $(i, j) \in \mathcal{L}$, or all times $t \in \mathcal{T}$.

Example IV.3. The sequence of static degrees $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$ is a paradigmatic example of a sequence of scalar features.

Individual features that depend both on topology and time are given both a subscript and a superscript index (e.g. x_i^m or $x_{(i,j)}^m$, where m refers to a given temporal ordering). Such features are generally assembled into a sequence of sequence that runs over both indices.

Example IV.4. An important example of a sequence of features that are themselves sequences of scalar features is the sequence of instantaneous node degrees $\mathbf{d} = (\mathbf{d}^t)_{t\in\mathcal{T}}$, where $\mathbf{d}^t = (d_i^t)_{i\in\mathcal{V}}$ is the degree sequence of the snapshot graph at time t, and where each instantaneous degree d_i^t is the number of events that node i partakes in at time t. Note that since the ranges of i and t do not depend on each other, we can reverse their order, $\mathbf{d} = (\mathbf{d}_i)_{i\in\mathcal{V}}$ with $\mathbf{d}_i = (d_i^t)_{t\in\mathcal{T}}$, which is formally the same feature since it imposes the same constraints.

Table IV.1 lists and defines a selection of elementary temporal network features. It serves as reference when reading the description of MRRMs in the taxonomy below.

Instead of constraining an ordered sequence itself, many MRRMs constrain marginal distributions of a sequence. A distribution of feature values returns the number of times each possible value of individual features $x_q \in \mathbf{x}$ appears in a measured sequence. We formally define a distribution as a multiset, and we will in the following use the two terms interchangeably.

Definition IV.4. *Distribution of feature values.* Given a sequence of features, $\mathbf{x}$, we can define a distribution for it. A distribution is defined as the multiset $[x_q]_{q\in\mathcal{Q}}$ containing the values of all elements $x_q \in \mathbf{x}$ including duplicate values.

The individual features in a sequence may be scalar, sequences of scalars, or other more general functions. This means that multiple types of distributions may be defined from a sequence of features $\mathbf{x}$, depending on the type of sequence.

The distribution constructed from a sequence of scalar features, $\mathbf{x} = (x_q)_{q\in\mathcal{Q}}$, is simply the multiset containing all individual feature values, $p(\mathbf{x}) = [x_q]_{q\in\mathcal{Q}}$.

Example IV.5. The sequence of static degrees, $\mathbf{k}$, is a sequence of scalar features, and we can construct the simple distribution $p(\mathbf{k}) = [k_i]_{i\in\mathcal{V}}$ from it.

From a sequence of vector valued features, $\mathbf{x} = (\mathbf{x}_q)_{q\in\mathcal{Q}}$ with $\mathbf{x}_q = (x_q^r)_{\mathcal{R}_q\in\mathcal{Q}}$, we may construct several different types of distributions by marginalizing over the inner or outer indices, or both. We show the different types of distributions that can be obtained in the following example (see Supplementary Note 2 for formal definitions of each type of distribution).

TABLE IV.1: **Elementary features of a temporal network.** Below, "$(\cdot)$" denotes a sequence, "$\{\cdot\}$" denotes a set, "$|\cdot|$" denotes the cardinality of a set, and ":" means *for which* or *such that*.

Symbol	*Meaning of symbol*	*Definition*
$[t_{\min}, t_{\max}]$	Period of observation.	
G	(Instant-event) temporal network.	$G = (\mathcal{V}, \mathcal{C})$ (Def. II.1)[a] / $G = (\mathcal{V}, \mathcal{E})$ (Def. II.2)[b]
$\mathcal{V}$	Set of nodes in G.	$\mathcal{V} = \{1, 2, \ldots, N\}$
$\mathcal{C}$ / $\mathcal{E}$	Set of (instantaneous) events in G.	$\mathcal{C} = \{c_1, c_2, \ldots, c_C\}$ [a] / $\mathcal{E} = \{e_1, e_2, \ldots, e_E\}$ [b]
c_q / e_q	qth (instantaneous) event.	$c_q = (i_q, j_q, t_q, \tau_q)$ [a] / $e_q = (i_q, j_q, t_q)$ [b]
i_q, j_q	Indices for nodes partaking in the qth event.	
t_q	Start time of the qth event.	
τ_q	Duration of the qth event.[c]	
N	Number of nodes in G.	$N = \lvert\mathcal{V}\rvert$
C / E	Number of events in G.	$C = \lvert\mathcal{C}\rvert$ [a] / $E = \lvert\mathcal{E}\rvert$ [b]
Link-timeline representation		
$G_{\mathcal{L}}$	Link-timeline network.	$G_{\mathcal{L}} = \left(G^{\text{stat}}, \mathbf{\Theta}\right)$ (Def. II.6)
G^{stat}	Static graph.	$G^{\text{stat}} = (\mathcal{V}, \mathcal{L})$
$\mathcal{L}$	Links in G^{stat}.	$\mathcal{L} = \{(i,j) : (i,j,t,\tau) \in \mathcal{C}\}$ [a] / $\mathcal{L} = \{(i,j) : (i,j,t) \in \mathcal{E}\}$ [b]
L	Number of links in G^{stat}.	$L = \lvert\mathcal{L}\rvert$
$\mathcal{V}_i$	Neighborhood of node i.	$\{j : (i,j) \in \mathcal{L}\}$
$\mathbf{\Theta}$	Sequence of timelines.	$\mathbf{\Theta} = (\Theta_{(i,j)})_{(i,j)\in\mathcal{L}}$
$\Theta_{(i,j)}$	Link timeline.	$\Theta_{(i,j)} = \left(\left(t^1_{(i,j)}, \tau^1_{(i,j)}\right), \left(t^2_{(i,j)}, \tau^2_{(i,j)}\right), \ldots, \left(t^{n_{(i,j)}}_{(i,j)}, \tau^{n_{(i,j)}}_{(i,j)}\right)\right)$ [a] / $\Theta_{(i,j)} = \left(t^1_{(i,j)}, t^2_{(i,j)}, \ldots, t^{w_{(i,j)}}_{(i,j)}\right)$ [b]
Snapshot-sequence representation [d]		
$G_{\mathcal{T}}$	Snapshot-graph sequence	$G_{\mathcal{T}} = (\mathcal{T}, \mathbf{\Gamma})$ (Def. II.7)[d]
$\mathcal{T}$	Sequence of snapshot times.	$\mathcal{T} = (t_m)_{m=1}^{T}$ [d]
$\mathbf{\Gamma}$	Sequence of snapshot graphs.	$\mathbf{\Gamma} = (\Gamma^t)_{t\in\mathcal{T}}$ [d]
Γ^t	Snapshot graph at time t.	$\Gamma^t = (\mathcal{V}, \mathcal{E}^t)$ [d]
$\mathcal{E}^t$	Instantaneous events at time t.	$\mathcal{E}^t = \{(i,j) : (i,j,t) \in \mathcal{E}\}$ [d]
Topological-temporal (two-level) features		
$t^m_{(i,j)}$	Event start time.	Start time of mth event in timeline $\Theta_{(i,j)}$ (Def. II.6)
$\tau^m_{(i,j)}$	Event duration.	Duration of mth event in timeline $\Theta_{(i,j)}$ (Def. II.6) [c]
$\Delta\tau^m_{(i,j)}$	Inter-event duration.	$\Delta\tau^m_{(i,j)} = t^{m+1}_{(i,j)} - (t^m_{(i,j)} + \tau^m_{(i,j)})$ [a] / $\Delta\tau^m_{(i,j)} = t^{m+1}_{(i,j)} - t^m_{(i,j)}$ [b]
$t^w_{(i,j)}$	End time of last event on timeline.	$t^w_{(i,j)} = t^{n_{(i,j)}}_{(i,j)} + \tau^{n_{(i,j)}}_{(i,j)}$ [a] / $t^w_{(i,j)} = t^{w_{(i,j)}}_{(i,j)}$ [b]
d^t_i	Instantaneous degree at time t.	$d^t_i = \lvert\{j : (i,j,t',\tau) \in \mathcal{C} \text{ and } t' \le t < t' + \tau\}\rvert$ [a] / $d^t_i = \lvert\{j : (i,j) \in \mathcal{E}^t\}\rvert$ [b]
v^m_i	Activity start time.	Start time of mth interval of consecutive activity of node i.
α^m_i	Activity duration.	Duration of mth interval of consecutive activity of node i.[c]
$\Delta\alpha^m_i$	Inactivity duration.	$\Delta\alpha^m_i = v^{m+1}_i - (v^m_i + \alpha^m_i)$ [a] / $\Delta\alpha^m_i = v^{m+1}_i - v^m_i$ [b]
Aggregated (one-level) features		
$n_{(i,j)}$	Link event frequency.	$n_{(i,j)} = \lvert\Theta_{(i,j)}\rvert$ [c]
$w_{(i,j)}$	Link weight.	$w_{(i,j)} = \sum_{m=1}^{n_{(i,j)}} \tau^m_{(i,j)}$ [a] / $w_{(i,j)} = \lvert\Theta_{(i,j)}\rvert$ [b]
a_i	Node activity.	$a_i = \sum_{j\in\mathcal{V}_i} n_{(i,j)}$ [c]
s_i	Node strength.	$s_i = \sum_{j\in\mathcal{V}_i} w_{(i,j)}$
k_i	Node degree.	$k_i = \lvert\mathcal{V}_i\rvert$
A^t	Cumulative activity at time t.	$A^t = \lvert\mathcal{E}^t\rvert$
Special features		
χ_λ	Indicator of connectedness of G^{stat}.	$\chi_\lambda = 1$ if G^{stat} is connected, $\chi_\lambda = 0$ elsewise.
$\text{iso}(\Gamma^t)$	Isomorphism class of Γ^t.	Set of graphs obtained by all permutations of node indices in Γ^t.
$\chi_{\mathbb{N}^+}(A^t)$	Indicator of activity at t.	Indicator function for $A^t \in \mathbb{N}^+$, returning 0 if $A^t = 0$ and 1 if $A^t \ge 1$.
$\text{per}(\Theta_{(i,j)})$	All periodic shifts of timeline $\Theta_{(i,j)}$.	$[\Theta^{\Delta T}_{(i,j)}]_{\Delta T\in\mathcal{T}}$, where each $(t^m_{(i,j)})' = t^m_{(i,j)} + \Delta T \mod (t_{\max} - t_{\min})$.

[a] Definition for a temporal network with event durations.
[b] Definition for an instant-event temporal network.
[c] Only defined for a temporal network with event durations.
[d] Only defined for an instant-event temporal network.

Example IV.6. The sequence of instantaneous degrees, $\mathbf{d} = ((d_i^t)_{i\in\mathcal{V}})_{t\in\mathcal{T}}$, is a sequence of sequences of scalar features. We can thus marginalize over the two indices i and t in different ways to construct different types of distributions:

- By marginalizing over both i and t, we obtain the *global distribution* $p(\mathbf{d}) = \cup_{t\in\mathcal{T}}[d_i^t]_{i\in\mathcal{V}}p(\mathbf{x}_q) = [d_i^t]_{t\in\mathcal{T},i\in\mathcal{V}}$.
- By marginalizing over the outer index t, we get the distribution of the degree sequences of each snapshot graph, $p_{\mathcal{T}}(\mathbf{d}) = [\mathbf{d}^t]_{t\in\mathcal{T}}$, where $\mathbf{d}^t = (d_i^t)_{i\in\mathcal{V}}$.
- Marginalizing over the inner index i gives the temporal sequence of the nodes' instantaneous degree distributions, $\boldsymbol{\pi}_{\mathcal{T}}(\mathbf{d}) = (p(\mathbf{d}^t))_{t\in\mathcal{T}}$, where $p(\mathbf{d}^t) = [d_i^t]_{i\in\mathcal{V}}$ is the instantaneous degree distribution in the tth snapshot.

While inverting the order of the indices in the definition of the sequence $\mathbf{d}$ (i.e. letting $\mathbf{d} = ((d_i^t)_{t\in\mathcal{T}})_{i\in\mathcal{V}}$) leads to exactly the same feature, it will lead to different distributions when marginalizing over the indices:

- By marginalizing over t with t as the inner index in $\mathbf{d} = ((d_i^t)_{t\in\mathcal{T}})_{i\in\mathcal{V}}$ we get the sequence of the temporal distribution of each node's instantaneous degree, $\boldsymbol{\pi}_{\mathcal{V}}(\mathbf{d}) = (p(\mathbf{d}^i))_{i\in\mathcal{V}}$, with $p(\mathbf{d}^i) = [d_i^t]_{t\in\mathcal{T}}$.
- Finally, marginalizing over i with i as the outer index leads to the distribution of the temporal sequences of the individual node's degree, $p_{\mathcal{V}}(\mathbf{d}) = [\mathbf{d}_i]_{i\in\mathcal{V}}$, with $\mathbf{d}_i = (d_i^t)_{t\in\mathcal{T}}$.

Finally, many MRRMs conserve mean values of network features.

Definition IV.5. *Mean of a sequence of features.* The mean $\mu(\mathbf{x})$ of a sequence of features is defined as the average over all individual scalar elements in $\mathbf{x}$.

From a sequence of scalars, the mean is naturally given as their average value: $\mu(\mathbf{x}) = \sum_{q\in\mathcal{Q}} x_q/Q$, where Q is the number of elements in $\mathbf{x}$. For a sequence of sequences of scalars, we may construct either the *global* mean, providing an average over all scalar elements: $\mu(\mathbf{x}) = \sum_{q\in\mathcal{Q}}\sum_{r\in\mathcal{R}_q} x_q^r/(\sum_{q\in\mathcal{Q}} R_q)$, where R_q is the number of elements in $\mathbf{x}_q$; or the sequence of local means, providing the averages over each local sequence $\mathbf{x}_q \in \mathbf{x}$: $\boldsymbol{\mu}_{\mathcal{Q}}(\mathbf{x}) = (\mu(\mathbf{x}_q))_{q\in\mathcal{Q}}$ with $\mu(\mathbf{x}_q) = \sum_{r\in\mathcal{R}_q} x_q^r/R_q$.

The distributions and moments are all functions of a sequence of features, so they are coarser than the sequence itself (Proposition IV.1). This means that a MRRM that constrains a distribution or the mean of a collection of feature values randomizes more than a MRRM that constrains their ordered sequence.

A detailed list of the general form that different distributions and moments take for features of nodes, links, and snapshots is given in Supplementary Table S1. Supplementary Note 2 additionally provides a detailed description of how the all the distributions and moments listed in Table S1 are constructed and discusses how to order them using the finer/coarser relation of Def. IV.1.

We are now ready to build a taxonomy of MRRMs which rigorously characterizes and orders them. We provide for each MRRM informal definitions of the principal features that it constrains and refer to Table IV.1 for detailed definitions of the features. The effects of each shuffling on a wide selection of other temporal network features are shown in Table IV.2. We additionally provide details and a graphical illustration of the typical algorithmic implementation of each shuffling method.

2. The basic instant-event and event shufflings

We first present the coarsest (i.e. the most random) instant-event and event shufflings possible.

a. Instant-event shuffling

P[E]. **Common name:** *Instant-event shuffling.* **Features constrained:** The number of instantaneous events, E. **Reference:** [26].

P[E] draws i, j, and t at random without replacement for each instantaneous event $(i, j, t) \in \mathcal{E}$.

b. Event shuffling

P[$p(\boldsymbol{\tau})$]. **Common name:** *Event shuffling.* **Features constrained:** the distribution (Def. IV.4) of event durations, $p(\tau) = [\tau_q]_{q=1}^{C}$. **Reference:** Supplementary Note 3.

P[$p(\boldsymbol{\tau})$] draws i, j, and t at random without replacement for each event $(i, j, t, \tau) \in \mathcal{C}$. It conserves the events durations but not their order, which is equivalent to constraining their distribution $p(\boldsymbol{\tau})$.

3. Link shufflings

Link shufflings alter the aggregated network topology but conserve temporal structure locally on each link. Different time-aggregated features may be constrained or randomized depending on the model.

Link shufflings are defined and implemented exactly the same way for temporal networks with and without event durations. We thus do not need to distinguish between instant-event and event shuffling versions of these. They are ordered hierarchically in the Hasse diagram shown in Fig. IV.27(a).

P[$p_{\mathcal{L}}(\boldsymbol{\Theta})$]. **Common name:** *Link shuffling.* **Features constrained:** the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\boldsymbol{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$. **References:** [15] (*Erdős-Rényi model*); Supplementary Note 3.

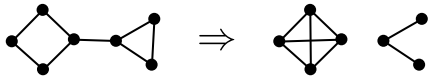


FIG. IV.4: P$[p_{\mathcal{L}}(\mathbf{\Theta})]$ shuffles the links and associated timelines $\Theta_{(i,j)} \in \mathbf{\Theta}$ between all node pairs (i,j) without any constraints on the static network structure (i.e. corresponding to the Erdős-Rényi (ER) [1] model). All nodes and links in the static network are equivalent and are shown in the same color.

P$[\chi_\lambda, p_{\mathcal{L}}(\mathbf{\Theta})]$. **Common name:** *Connected link shuffling.* **Features constrained:** the connectedness $\chi_\lambda(G^{\text{stat}})$ of the static graph $G^{\text{stat}} = (\mathcal{V}, \mathcal{L})$; the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\mathbf{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$. **References:** [51] (*rewiring*); [13] (*random network*).

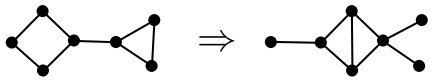


FIG. IV.5: P$[\chi_\lambda, p_{\mathcal{L}}(\mathbf{\Theta})]$ generates randomized networks in the same manner as P$[p_{\mathcal{L}}(\mathbf{\Theta})]$, with the additional constraint that the static graph G^{stat} of the sampled networks must be connected if it was in the input network. All nodes and links in the static network are equivalent and are shown in the same color.

P$[\mathbf{k}, p_{\mathcal{L}}(\mathbf{\Theta})]$. **Common name:** *Degree-constrained link shuffling.* **Features constrained:** the static degree sequence, $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$; the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\mathbf{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$. **References:** [45, 52, 56] (*randomized edges*); [47] (*random link shuffling*); [22] (*randomized structure*); Supplementary Note 3.

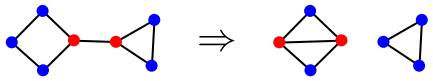


FIG. IV.6: P$[\mathbf{k}, p_{\mathcal{L}}(\mathbf{\Theta})]$ shuffles the links and associated timelines $\Theta_{(i,j)} \in \mathbf{\Theta}$ between all node pairs (i,j) while constraining the sequence of degrees of the nodes in the static network, $\mathbf{k}$ (typically implemented using the algorithm of Maslov and Sneppen [38] or by using a stub matching algorithm [2, 3]). Nodes are colored by their degree in the static graph G^{stat}, which is conserved by the shuffling.

P$[\mathbf{k}, \chi_\lambda, p_{\mathcal{L}}(\mathbf{\Theta})]$. **Common name:** *Connected degree-constrained link shuffling.* **Features constrained:** the connectedness $\chi_\lambda(G^{\text{stat}})$ of the static graph $G^{\text{stat}} = (\mathcal{V}, \mathcal{L})$; the static degree sequence, $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$; the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\mathbf{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$. **References:** [13] (*configuration model*).

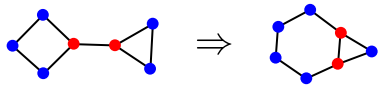


FIG. IV.7: P$[\mathbf{k}, \chi_\lambda, p_{\mathcal{L}}(\mathbf{\Theta})]$ generates randomized networks in the same manner as P$[\mathbf{k}, p_{\mathcal{L}}(\mathbf{\Theta})]$, with the additional constraint that the static network of the sampled networks must be connected if it was in the input network. Nodes are colored by their degree in the static graph G^{stat}, which is conserved by the shuffling.

4. *Timeline shufflings*

Timeline shufflings randomize the individual timelines $\Theta_{(i,j)}$ without changing the topology of the aggregated network. They typically randomize temporal features of both links and nodes to different extents as described below.

The timeline shufflings listed below are ordered hierarchically in the Hasse diagram shown in Fig. IV.27(b).

a. *Instant-event shufflings*

P$[\mathcal{L}, E]$. **Common name:** *Timeline shuffling.* **Features constrained:** the static graph $G^{\text{stat}} = (\mathcal{V}, \mathcal{L})$; the number of instantaneous events, E. **References:** [45, 52] (*random(ized) contacts*).

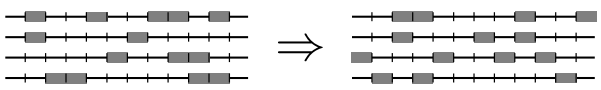


FIG. IV.8: P$[\mathcal{L}, E]$ redistributes the instantaneous events completely at random between the existing timelines. Since all events are equivalent, they are marked in the same color (grey).

P$[\mathbf{w}]$. **Common name:** *Weight-constrained timeline shuffling.* **Features constrained:** the sequence of link weights, $\mathbf{w} = (w_{(i,j)})_{(i,j)\in\mathcal{L}}$ (numbers of instantaneous events per link). **References:** [45, 47, 50, 52, 60] (*random time(s)*); [13] (*uniformly random times*); [42] (*temporal mixed edges*); [51] (*poissonized inter-event intervals*); [12] (*SRan*); [26, 80]; Section VII.

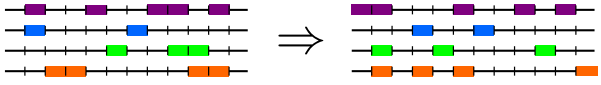


FIG. IV.9: P$[\mathbf{w}]$ randomizes the timestamps of the instantaneous events inside each individual timeline. Events of the same color stay on the same timeline after shuffling.

P$[\boldsymbol{\pi}_{\mathcal{L}}(\mathbf{\Delta\tau}), \mathbf{t}^1]$. **Common name:** *Inter-event shuffling.* **Features constrained:** the sequence of local distributions of inter-event durations on each link,

$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}) = ([\Delta\tau^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$; the sequence of times of the first event on each link, $\mathbf{t}^1 = (t^1_{(i,j)})_{(i,j)\in\mathcal{L}}$. **References:** [51, 53] (*shuffled inter-event intervals*); Section VII.

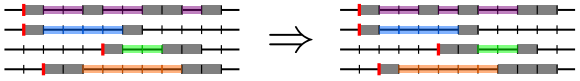


FIG. IV.10: P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$,$\mathbf{t}^1$] shuffles the inter-event durations between the instantaneous events on each link while keeping the times of the first event on each link fixed. Inter-event durations on the same link are marked in the same color and red vertical lines mark the start times of the first event on each link. Both are conserved by the shuffling.

b. Event shufflings

P[$\mathcal{L}$,$p(\boldsymbol{\tau})$] **Common name:** *Timeline shuffling.* **Features constrained:** the static graph $G^{\text{stat}} = (\mathcal{V},\mathcal{L})$; the distribution (Def. IV.4) of event durations, $p(\tau) = [\tau_q]^C_{q=1}$. **Reference:** Supplementary Note 3.

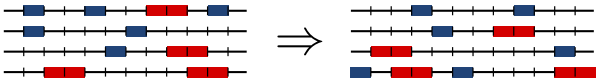


FIG. IV.11: P[$\mathcal{L}$,$p(\boldsymbol{\tau})$] constrains the static network structure G^{stat} and otherwise shuffles the events completely at random between all timelines. Colors mark the events' durations, which are conserved by the shuffling.

P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$] **Common name:** *Local timeline shuffling.* **Features constrained:** the sequence of local distributions of event durations on each link, $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}) = ([\tau^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$. **Reference:** Supplementary Note 3.

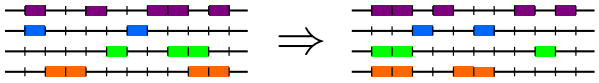


FIG. IV.12: P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$] redistributes the events uniformly inside each timeline, but not in-between them. The events are colored by the timeline they belong to.

P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$,$\mathbf{t}^1$,$\mathbf{t}^w$] **Common name:** *Activity-constrained timeline shuffling.* **Features constrained:** the sequence of local distributions of event durations on each link, $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}) = ([\tau^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$. the sequences of times of the first and last events on each link, $\mathbf{t}^1 = (t^1_{(i,j)})_{(i,j)\in\mathcal{L}}$ and $\mathbf{t}^w = (t^w_{(i,j)})_{(i,j)\in\mathcal{L}}$, respectively. **Reference:** Supplementary Note 3.

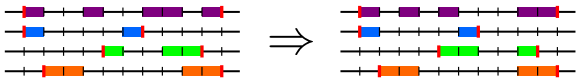


FIG. IV.13: P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$,$\mathbf{t}^1$,$\mathbf{t}^w$] redistributes the events at random inside each timeline, while constraining the start time of the first event and the end time of the last event in each timeline (i.e. the timelines' *activity* intervals). The colors of events mark the timeline they belong to and vertical red lines mark the start time of the first event and end time of last event on each timeline,

P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$, $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$] **Common name:** *Interval shuffling.* **Features constrained:** the sequences of local distributions of event and inter-event durations on each link, $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}) = ([\tau^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$ and $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}) = ([\Delta\tau^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$, respectively. **References:** [14] (*interval shuffling*); Supplementary Note 3.

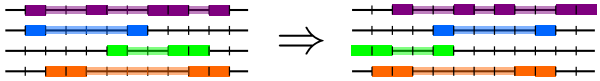


FIG. IV.14: P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$, $\boldsymbol{\pi}_{\mathcal{L}}\boldsymbol{\Delta\tau}$] shuffles the start time of the first event as well as the order of the event and inter-event durations on each link. The events and inter-event intervals are colored by the timeline they belong to.

P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$, $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$,$\mathbf{t}^1$] **Common name:** *Inter-event shuffling.* **Features constrained:** the sequences of local distributions of event and inter-event durations on each link, $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}) = ([\tau^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$ and $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}) = ([\Delta\tau^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$, respectively; the sequence of times of the first event on each link, $\mathbf{t}^1 = (t^1_{(i,j)})_{(i,j)\in\mathcal{L}}$. **Reference:** Supplementary Note 3.

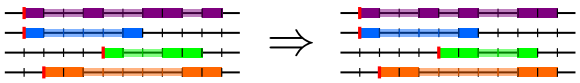


FIG. IV.15: P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$,$\mathbf{t}^1$] adds another constraint to P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$, $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$] so that it conserves the time of the first event each link. The events and inter-event intervals are colored by the timeline they belong to, and red vertical lines mark the start time of the first events in each timeline.

P[$\mathbf{per}(\boldsymbol{\Theta})$] **Common name:** *Timeline shifting.* **Features constrained:** The sequence of sets of all possible translations of each timeline with periodic boundary conditions, $\mathbf{per}(\boldsymbol{\Theta}) = (\text{per}(\Theta_{(i,j)}))_{(i,j)\in\mathcal{L}}$. **Reference:** [19] (*random offset*).

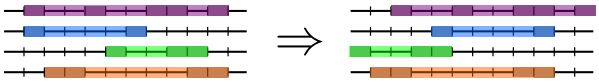


FIG. IV.16: P[**per**(**Θ**)] randomly translates the timelines on each link individually using periodic boundary conditions, randomizing the start time of the first event in each timeline but otherwise conserving their temporal order and placement. Colors highlight the intervals between the first and last events in each timeline before and after shuffling.

5. *Sequence shufflings*

Sequence shufflings randomize the sequence of snapshots while leaving the individual snapshots unchanged, They generally destroy temporal correlations inside timelines and in node activities.

We have identified the following two sequence shufflings in the literature. These are included in the Hasse diagram shown in Fig. IV.27(c).

P[$p_{\mathcal{T}}(\mathbf{\Gamma})$] **Common name:** *Sequence shuffling.* **Features constrained:** the distribution (Def. IV.4) of snapshot graphs $p_{\mathcal{T}}(\mathbf{\Gamma}) = [\Gamma^t]_{t\in\mathcal{T}}$. **References:** [57] (*reshuffled sequences*); [20, 42] (*random ordered*); [50] (*shuffled times*); [25] (*reshuffle*).

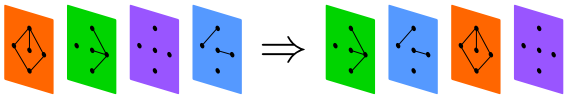


FIG. IV.17: P[$p_{\mathcal{T}}(\mathbf{\Gamma})$] randomly shuffles the order of the snapshots. Colors mark each individual snapshot graph, which the shuffling conserves.

P[$p_{\mathcal{T}}(\mathbf{\Gamma})$,$\boldsymbol{\chi}_{\mathbb{N}^+}(\mathbf{A})$] **Common name:** *Activity-constrained sequence shuffling.* **Features constrained:** the distribution (Def. IV.4) of snapshot graphs $p_{\mathcal{T}}(\mathbf{\Gamma}) = [\Gamma^t]_{t\in\mathcal{T}}$; the times during which events take place, formally defined as the indicator function for $A^t \in \mathbb{N}^+$ (i.e. $A^t \leq 1$) at each time, $\boldsymbol{\chi}_{\mathbb{N}^+}(\mathbf{A}) = (\chi_{\mathbb{N}^+}(A^t))_{t\in\mathcal{T}}$. **References:** [50] (*shuffled times*).

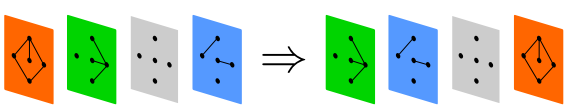


FIG. IV.18: P[$p_{\mathcal{T}}(\mathbf{\Gamma})$,$\boldsymbol{\chi}_{\mathbb{N}^+}(\mathbf{A})$] shuffles the timestamps between snapshots where at least one event takes place. Colors mark each individual snapshot graph, which the shuffling conserves. The grey snapshot contains no events and its placement is not shuffled.

6. *Snapshot shufflings*

Snapshot shufflings conserve the start times t of all events. They are typically implemented by randomizing the instantaneous snapshot graphs Γ^t corresponding to each time $t \in \mathcal{T}$. As a consequence, all snapshot shufflings found in the literature are instant-event shufflings, but they may also be implemented as event shufflings—we give one example in this section and two others in Section IV B 8.

Snapshot shufflings generally destroy temporal features of the links, but may conserve some temporal node features, such as the sequence of instantaneous degrees, **d**.

The snapshot shufflings listed below are ordered hierarchically in the Hasse diagram shown in Fig. IV.27(d).

a. *Instant-event shufflings*

P[**t**] **Common name:** *Snapshot shuffling.* **Features constrained:** the times of all instantaneous events, $\mathbf{t} = (t_q)_{q\in\{1,2,\ldots,E\}}$. **Reference:** [50] (*random network*).

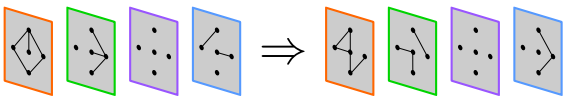


FIG. IV.19: P[**t**] randomly shuffles the instantaneous events inside each snapshot. This is equivalent to generating each snapshot Γ^t as an instance of an Erdős-Rényi graph with N nodes and $A^t = |\mathcal{E}^t|$ edges. Colored outlines mark the temporal placement of each snapshot, which is conserved.

P[**d**] **Common name:** *Degree-constrained snapshot shuffling.* **Features constrained:** sequence of instantaneous degrees, $\mathbf{d} = ((d_i^t)_{i\in\mathcal{V}})_{t\in\mathcal{T}}$. **References:** [42] (*time ordered and reshuffled networks*);[50] (*degree preserved network*); [63, 91].

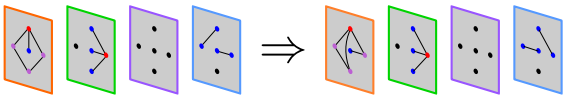


FIG. IV.20: P[**d**] shuffles the events inside each snapshot while constraining the instantaneous degree sequence $\mathbf{d}^t = (d_i^t)_{i\in\mathcal{V}}$, using e.g. the Maslov-Sneppen model. Outline colors mark the temporal placement of each snapshot and the nodes' colors mark their instantaneous degrees.

P[**iso**(**Γ**)] **Common name:** *Isomorphic snapshot shuffling.* **Features constrained:** the isomorphism class of each snapshot graph, $\mathbf{iso}(\mathbf{\Gamma}) = (\mathrm{iso}(\Gamma^t))_{t\in\mathcal{T}}$. **Reference:** [25] (*anonymize*).

FIG. IV.21: P[**iso**(**Γ**)] consists in randomizing the identity of the nodes in each time snapshot. This produces snapshot graphs, $(\Gamma^t)'$, that are isomorphic to those of the original network, $(\Gamma^t)' \simeq \Gamma^t$. Outline colors mark the temporal placement of the snapshots and link colors mark the isomorphism class of the snapshot graphs.

b. Event shufflings

P[$p(\mathbf{t},\boldsymbol{\tau})$] **Common name:** *Snapshot shuffling.* **Features constrained:** the timestamps and durations of the events, $p(\mathbf{t},\boldsymbol{\tau}) = [(t,\tau)]_{(i,j,t,\tau)\in\mathcal{C}}$. **References:** Supplementary Note 3.

P[$p(\mathbf{t},\boldsymbol{\tau})$] randomizes the values of i and j for each event $(i,j,t,\tau) \in \mathcal{C}$ while constraining the time t at which it occurs as well as its duration τ.

7. Intersections of link and timeline shufflings

Several shuffling methods constrain both the static graph $G^{\text{stat}} = (\mathcal{V},\mathcal{L})$ and the multiset of timelines, $p(\boldsymbol{\Theta}) = [\Theta_\ell]_{(i,j)\in\mathcal{L}}$, and are thus intersections of link and timeline shufflings.

The shufflings randomize temporal-topological correlations.

The shufflings are included in the Hasse diagrams in Figs. IV.27(a) and IV.27(b).

P[$\mathcal{L}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$] **Common name:** *Topology-constrained link shuffling.* **Features constrained:** the static graph $G^{\text{stat}} = (\mathcal{V},\mathcal{L})$; the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\boldsymbol{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$. **References:** [9, 13, 21] (*link-sequence shuffled*); [45] (*edge randomization*); [14, 24] (*link shuffling*); Supplementary Note 3.

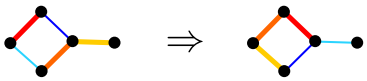


FIG. IV.22: P[$\mathcal{L}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$] randomly shuffles the timelines between all links while keeping the static graph G^{stat} fixed. Colors mark links corresponding to different timelines and are randomized by the shuffling. P[$\mathcal{L}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$] is the most random intersection between a link and a timeline shuffling.

P[**w**,$p_{\mathcal{L}}(\boldsymbol{\Theta})$] **Common name:** *Weight-constrained link shuffling.* **Features constrained:** the sequence of link weights, $\mathbf{w} = (w_{(i,j)})_{(i,j)\in\mathcal{L}}$ (cumulative duration of events on each link); the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\boldsymbol{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$. **References:** [9, 13, 21, 60] (*equal-weight link-sequence shuffled*); [45] (*equal-weight edge randomization (EWER)*).

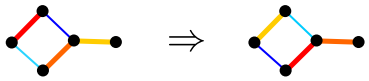


FIG. IV.23: P[**w**,$p_{\mathcal{L}}(\boldsymbol{\Theta})$] shuffles timelines $\Theta_{(i,j)}$ between links with the same cumulative event duration $w_{(i,j)}$ (for instant-event networks defined as the number of instantaneous events). Colors mark links corresponding to different timelines, while link thickness marks their weights $w_{(i,j)}$. The former are randomized under the constraint that the latter are conserved.

P[**n**,$p_{\mathcal{L}}(\boldsymbol{\Theta})$] **Common name:** *Weight-constrained link shuffling.* **Features constrained:** the sequence of link weights, $\mathbf{n} = (n_{(i,j)})_{(i,j)\in\mathcal{L}}$ (number of events on each link); the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\boldsymbol{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$. **Reference:** Supplementary Note 3.

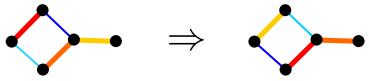


FIG. IV.24: P[**n**,$p_{\mathcal{L}}(\boldsymbol{\Theta})$] shuffles timelines between links with the same number of events, $n_{(i,j)}$. Colors mark links corresponding to different timelines, while link thickness marks their weights $n_{(i,j)}$. The former are randomized under the constraint that the latter are conserved.

8. Intersections of timeline and snapshot shufflings

Yet other shuffling methods constrain both the static graph $G^{\text{stat}} = (\mathcal{V},\mathcal{L})$ and the timestamps of each event **t** and are thus intersections of timeline and snapshot shufflings.

The shufflings are included in the Hasse diagrams in Figs. IV.27(b) and IV.27(d).

a. Instant-event shufflings.

P[**w**,**t**] **Common name:** *Timestamp shuffling.* **Features constrained:** the sequence of link weights, $\mathbf{w} = (w_{(i,j)})_{(i,j)\in\mathcal{L}}$ (number instantaneous events on each link); the times of all instantaneous events, $\mathbf{t} = (t_q)_{q\in\{1,2,\ldots,E\}}$. **References:** [23] (*time-stamp shuffling*); [52] (*permuted times*); [4, 7, 9, 13, 21, 60] (*time-shuffled* or *time-shuffling*); [15] (*time reshuffle*); [16, 45, 54, 56] (*randomly permuted times*); [10] (*random dynamic*); [19] (*random time shuffle*); [25] (*reconfigure*); [47] (*shuffled time stamps*); [11, 15, 49, 53, 62].

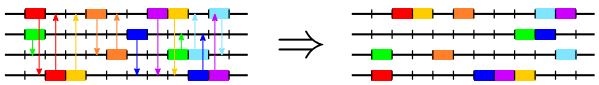


FIG. IV.25: P[**w**,**t**] randomly shuffles the timestamps t of all instantaneous events $(i,j,t) \in \mathcal{E}$, while keeping i and j fixed. In a completely equivalent manner, we may define the shuffling by constraining the timestamps **t** and permuting the pairs (i,j). The figure illustrates the procedure, where pairs of instantaneous events (of the same color) are swapped between timelines (red arrows) while conserving their timestamps. Due to the indistinguishability of networks obtained through permutation of event indices, both are equivalent to conserving **w** and **A**. For convenience, we choose the canonical name P[**w**,**t**] which conveys that it is both a timeline shuffling and a snapshot shuffling.

b. Event shufflings

P[$\mathcal{L}$,p(**t**,$\boldsymbol{\tau}$)] **Common name:** *Topology-constrained snapshot shuffling.* **Features constrained:** the static graph $G^{\text{stat}} = (\mathcal{V},\mathcal{L})$. **References:** Supplementary Note 3.

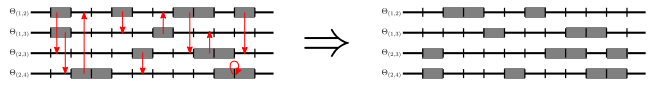


FIG. IV.26: P[$\mathcal{L}$,p(**t**,$\boldsymbol{\tau}$)] shuffles the events between existing links while constraining their starting times and their durations. Each event is moved a random timeline (red arrows—note that the new placement may be the same as the old).

9. Randomization based on metadata

The availability of metadata offers the possibility to impose additional constraints in the MRRMs. This allows to study effects that are not purely due to network structure and dynamics. For instance, in Ref. [7], the age, gender, and type of subscription of mobile phone users were known; in Ref. [10], the authors used shuffling methods respecting the bipartite structure of a sex worker-buyer interaction network, and Ref. [24] used a shuffling that rewired links between each pair of predefined node groups in face-to-face networks.

These metadata MRRMs are all a type of stochastic blockmodel[5]. They may be defined by assigning a *color* to each node, i.e. to which group it belongs among a set of R predefined groups. The node colors are fixed by the vector $\boldsymbol{\sigma} = (\sigma_1,\sigma_2,\ldots,\sigma_N)$, where $\sigma_i \in \{1,2,\ldots,R\}$. An $R \times R$ *group contact matrix*, $\Sigma_{\mathcal{L}}$ [with elements given by the number of links between groups, $(\Sigma_{\mathcal{L}})_{\sigma\sigma'} = \sum_{(i,j)\in\mathcal{L}}(\delta_{\sigma_i,\sigma}\delta_{\sigma_j,\sigma'} + \delta_{\sigma_j,\sigma}\delta_{\sigma_i,\sigma'})$], typically fixes the number of links between members of each group [we may alternatively fix the number of events instead using a matrix $\Sigma_{\mathcal{C}}$, with elements given by $(\Sigma_{\mathcal{C}})_{\sigma\sigma'} = \sum_{(i,j,t,\tau)\in\mathcal{C}}(\delta_{\sigma_i,\sigma}\delta_{\sigma_j,\sigma'} + \delta_{\sigma_j,\sigma}\delta_{\sigma_i,\sigma'})$]. These two additional constraints enables us to define MRRMs that impose structure or dynamics determined by the metadata.

We may also directly use this blockmodel MRRM construction to conserve the bipartite structure of a network as in [10] by imposing two groups and a perfectly antidiagonal $\Sigma_{\mathcal{L}}$, with $(\Sigma_{\mathcal{L}})_{11} = (\Sigma_{\mathcal{L}})_{22} = 0$ and $(\Sigma_{\mathcal{L}})_{12} = (\Sigma_{\mathcal{L}})_{21} = L$. We may finally allow both $\boldsymbol{\sigma}$ and Σ to vary over time in order to capture temporal changes in the group structure.

We describe below MRRMs relying on node metadata. These are all link shufflings, so they conserve the same network features as those.

P[$p_{\mathcal{L}}(\boldsymbol{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$] **Common name:** *block-constrained link shuffling.* **Features constrained:** the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\boldsymbol{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$; the sequence of node colors, $\boldsymbol{\sigma} = (\sigma_i)_{i\in\mathcal{V}}$; the group contact matrix, $\Sigma_{\mathcal{L}}$. **References:** [24] (*CM-shuffling*).

P[$p_{\mathcal{L}}(\boldsymbol{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$ shuffles the links in the static graph while constraining the group membership of each node, $\boldsymbol{\sigma}$, and the number of links between each group, $\Sigma_{\mathcal{L}}$.

P[**k**,$p_{\mathcal{L}}(\boldsymbol{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$] **Common name:** *degree- and block-constrained link shuffling.* **Features constrained:** the degree sequence $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$; the distribution (Def. IV.4) of timelines, $p_{\mathcal{L}}(\boldsymbol{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$; the sequence of node colors, $\boldsymbol{\sigma} = (\sigma_i)_{i\in\mathcal{V}}$; the group contact matrix, $\Sigma_{\mathcal{L}}$. **References:** [10] (*random topological*).

P[**k**,$p_{\mathcal{L}}(\boldsymbol{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$] randomizes G^{stat} while constraining the group structure, as P[$p_{\mathcal{L}}(\boldsymbol{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$] does, while additionally constraining the node degrees **k**.

P[G,p($\boldsymbol{\sigma}$)] **Common name:** *color shuffling.* **Features constrained:** the complete temporal network G; the distribution (Def. IV.4) of node colors, $p(\boldsymbol{\sigma}) = [\sigma_i]_{i\in\mathcal{V}}$. **References:** [7] (*node type shuffled data*).

P[G,p($\boldsymbol{\sigma}$)] shuffles the group affiliations (colors) of the nodes at random. It thus destroys all correlations between node color and network structure and dynamics, but conserves the network structure and dynamics completely.

[5] These node-grouped MRRMs can be seen as microcanonical variants of the stochastic block model [92]. However, typical stochastic block models found in the literature assign either the links at random inside each block (i.e. equivalent to P[$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$]) or while constraining the degree sequence (equivalent to P[**k**,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$]) [37], while the metadata MRRMs we consider here may impose any structural constraints.

TABLE IV.2: **Effects of MRRMs on features of temporal networks.** See Table IV.1 for definitions of features. Colored symbols show to what extent each feature is conserved. Informal definitions are found in the tablenotes (detailed definitions are found in Supplementary Table S1).

Canonical name	Common name	topological			weighted				temp.	node			link			
		G^{stat}	k_i	L	$a_i{}^{\dagger}$	s_i	$n_{(i,j)}{}^{\dagger}$	$w_{(i,j)}$	A^t	$\alpha_i^{m\,\dagger}$	$\Delta\alpha_i^m$	d_i^t	$\tau_{(i,j)}^m{}^{\dagger}$	$\Delta\tau_{(i,j)}^m$	$t_{(i,j)}^1$	$t_{(i,j)}^w$
P[E]	*Instant-event shuffling*	–	–	–	–	μ	–	–	μ	–	–	μ	–	–	–	–
P[$p(\boldsymbol{\tau})$]	*Event shuffling*	–	–	–	–	μ	–	–	μ	–	–	μ	p	–	–	–
Link shufflings (LS):																
P[$p_{\mathcal{L}}(\boldsymbol{\Theta})$]	*LS*	–	μ	x	μ	μ	p	p	x	–	–	$\mu_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
P[$\chi_\lambda, p_{\mathcal{L}}(\boldsymbol{\Theta})$]	*Connected LS*	χ_λ	μ	x	μ	μ	p	p	x	–	–	$\mu_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
P[$\mathbf{k}, p_{\mathcal{L}}(\boldsymbol{\Theta})$]	*Degree-constrained LS*	–	x	x	μ	μ	p	p	x	–	–	$\mu_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
P[$\mathbf{k}, \chi_\lambda, p_{\mathcal{L}}(\boldsymbol{\Theta})$]	*Connected, degree-constr. LS*	χ_λ	x	x	μ	μ	p	p	x	–	–	$\mu_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
Timeline shufflings (TS):																
P[$\mathcal{L}, E$]	*TS*	x	x	x	–	μ	–	μ	μ	–	–	μ	–	–	–	–
P[$\mathbf{w}$]	*Weight-constrained TS*	x	x	x	–	x	–	x	μ	–	–	μ	–	–	–	–
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}),\mathbf{t}^1$]	*Inter-event shuffling*	x	x	x	x	x	x	x	μ	–	–	μ	$\mu_{\mathcal{L}}$	$\boldsymbol{\pi}_{\mathcal{L}}$	x	x
P[$\mathcal{L}$,$p(\boldsymbol{\tau})$]	*TS*	x	x	x	μ	μ	μ	μ	μ	–	–	μ	p	–	–	–
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$]	*Local TS*	x	x	x	x	x	x	x	μ	–	–	μ	$\boldsymbol{\pi}_{\mathcal{L}}$	–	–	–
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}),\mathbf{t}^1,\mathbf{t}^w$]	*Activity-constrained TS*	x	x	x	x	x	x	x	μ	–	–	μ	$\boldsymbol{\pi}_{\mathcal{L}}$	$\boldsymbol{\mu}_{\mathcal{L}}$	x	x
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}),\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$]	*Interval shuffling*	x	x	x	x	x	x	x	μ	–	–	μ	$\boldsymbol{\pi}_{\mathcal{L}}$	$\boldsymbol{\pi}_{\mathcal{L}}$	–	–
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}),\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}),\mathbf{t}^1$]	*Inter-event shuffling*	x	x	x	x	x	x	x	μ	–	–	μ	$\boldsymbol{\pi}_{\mathcal{L}}$	$\boldsymbol{\pi}_{\mathcal{L}}$	x	x
P[**per**($\boldsymbol{\Theta}$)]	*Timeline shifting*	x	x	x	x	x	x	x	μ	–	–	μ	x	x	–	–
Sequence shufflings (SeqS):																
P[$p_{\mathcal{T}}(\boldsymbol{\Gamma})$]	*SeqS*	x	x	x	–	x	–	x	p	–	–	$p_{\mathcal{T}}$	–	–	–	–
P[$p_{\mathcal{T}}(\boldsymbol{\Gamma}),\boldsymbol{\chi}_{\mathbb{N}^+}(\mathbf{A})$]	*Activity-constrained SeqS*	x	x	x	–	x	–	x	p, H	–	–	$p_{\mathcal{T}}$	–	–	–	–
Snapshot shufflings (SnapS):																
P[$\mathbf{t}$]	*SnapS*	–	–	–	–	μ	–	–	x	–	–	$\mu_{\mathcal{T}}$	–	–	–	–
P[$\mathbf{d}$]	*Degree-constrained SnapS*	–	–	–	–	μ	–	–	x	x	x	x	–	–	–	–
P[**iso**($\boldsymbol{\Gamma}$)]	*Isomorphic SnapS*	–	–	–	–	μ	–	–	x	–	–	$\boldsymbol{\pi}_{\mathcal{T}}$	–	–	–	–
P[$p(\mathbf{t},\boldsymbol{\tau})$]	*SnapS*	–	–	–	–	μ	–	–	x	–	–	$\mu_{\mathcal{T}}$	$\boldsymbol{\pi}_{\mathcal{T}}$	–	–	–
Link-timeline intersections:																
P[$\mathcal{L}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$]	*Topology-constrained LS*	x	x	x	μ	μ	p	p	x	–	–	$\mu_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
P[$\mathbf{w}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$]	*Weight-constrained LS*	x	x	x	μ	x	p	x	x	–	–	$\mu_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
P[$\mathbf{n}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$]	*Weight-constrained LS*	x	x	x	x	μ	x	p	x	–	–	$\mu_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
Link-snapshot intersections:																
P[$\mathbf{w}$,$\mathbf{t}$]	*Timestamp shuffling*	x	x	x	–	x	–	x	x	–	–	$\mu_{\mathcal{T}}$	–	–	–	–
P[$\mathcal{L}$,$p(\mathbf{t},\boldsymbol{\tau})$]	*Topology-constrained SnapS*	x	x	x	μ	μ	μ	μ	x	–	–	$\mu_{\mathcal{T}}$	$\boldsymbol{\pi}_{\mathcal{T}}$	–	–	–

† Feature only defined for temporal networks with event durations.

x Feature completely conserved, typically the ordered sequence of individual features, e.g. $\mathbf{x} = (x_i)_{i\in\mathcal{V}}$ or $\mathbf{x} = ((\mathbf{x}_i^t)_{t\in\mathcal{T}})_{i\in\mathcal{V}}$.

$\boldsymbol{\pi}_{\mathcal{L}}$ Sequence of local distributions on links, $\boldsymbol{\pi}_{\mathcal{L}}(\mathbf{x}) = \left([x_{(i,j)}^m]_{m\in\mathcal{M}_{(i,j)}}\right)_{(i,j)\in\mathcal{L}}$.

$\boldsymbol{\pi}_{\mathcal{T}}$ Sequence of local distributions in snapshots, $\boldsymbol{\pi}_{\mathcal{T}}(\mathbf{x}) = \left([x_i^t]_{i\in\mathcal{V}}\right)_{t\in\mathcal{T}}$.

$p_{\mathcal{L}}$ Distribution of local sequences on links, $p_{\mathcal{L}}(\mathbf{x}) = \left[(x_{(i,j)}^m)_{m\in\mathcal{M}_{(i,j)}}\right]_{(i,j)\in\mathcal{L}}$.

$p_{\mathcal{T}}$ Distribution of local sequences in snapshots, $p_{\mathcal{T}}(\mathbf{x}) = \left[(x_i^t)_{i\in\mathcal{V}}\right]_{t\in\mathcal{T}}$.

p Distribution (i.e. the multiset) of individual scalar values in sequence.

$\boldsymbol{\mu}_{\mathcal{L}}$ Sequence of local means on links, $\boldsymbol{\mu}_{\mathcal{L}}(\mathbf{x}) = \left(\sum_{m\in\mathcal{M}_{(i,j)}} x_{(i,j)}^m / M_{(i,j)}\right)_{(i,j)\in\mathcal{L}}$.

$\boldsymbol{\mu}_{\mathcal{T}}$ Sequence of local means in snapshots links, $\boldsymbol{\mu}_{\mathcal{T}}(\mathbf{x}) = \left(\sum_{i\in\mathcal{V}} x_i^t / N\right)_{t\in\mathcal{T}}$.

μ Mean value of the individual scalar features in sequence.

– Feature not conserved.

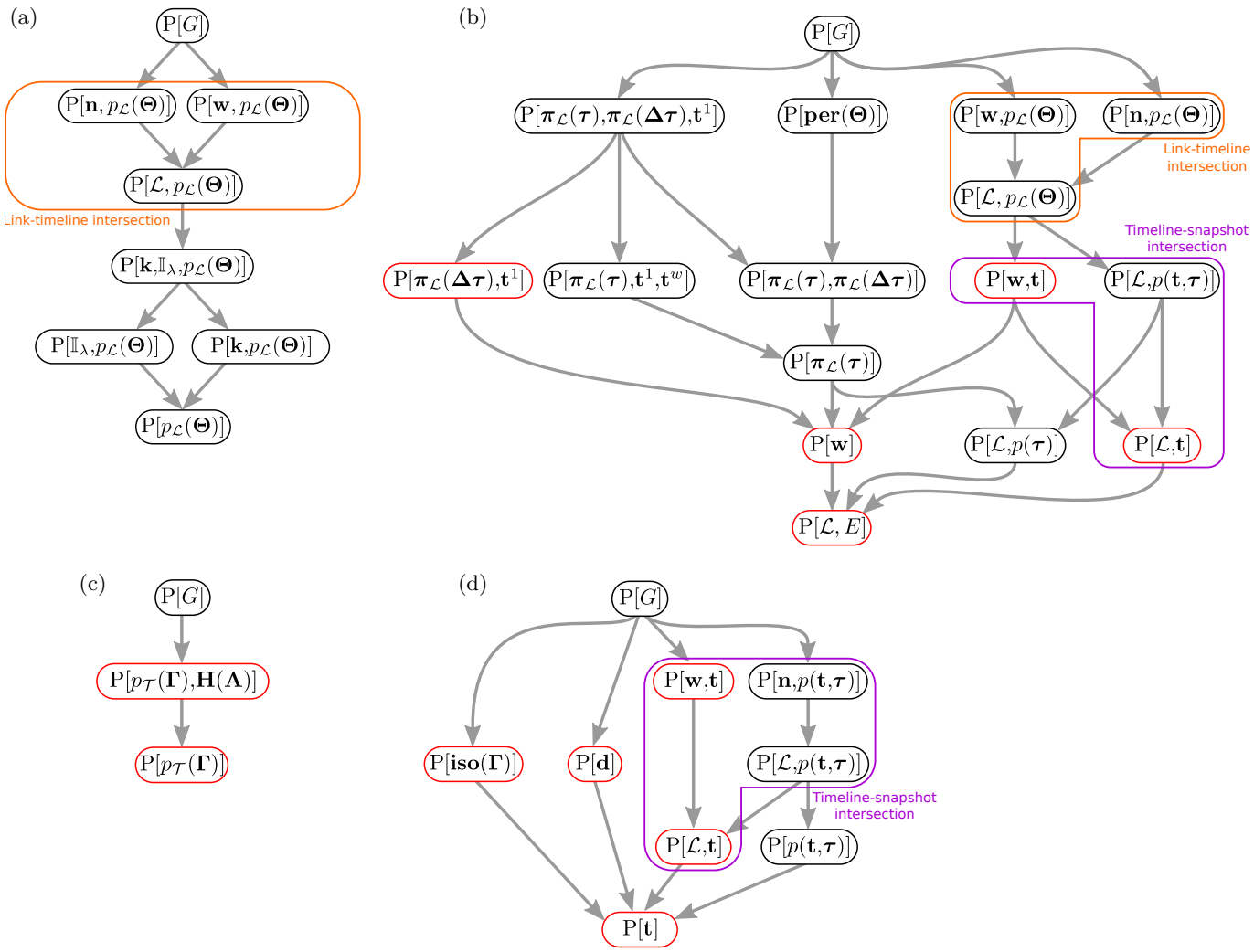


FIG. IV.27: **Hierarchies of shuffling methods.** (a) Link shufflings (Sec. IV B 3). (b) Timeline shufflings (Sec. IV B 4). (c) Sequence shufflings (Sec. IV B 5). (d) Snapshot shufflings (Sec. IV B 6). An arrow from a higher MRRM to a lower one indicates that the former MRRM is finer than the latter and thus randomizes less. Nodes with red outlines represent instant-event shufflings, black outlines mark event shufflings. The link-timeline intersections are defined in Section IV B 7 and the timeline-snapshot intersections in Section IV B 8.

V. GENERATING NEW MICROCANONICAL RANDOMIZED REFERENCE MODELS FROM EXISTING ONES

It is possible to combine two different MRRMs to form a new MRRM by applying the second MRRM to the state (i.e., a network) returned by the first. This defines a *composition* of the MRRMs and results in a model that randomizes more than either of the two original MRRMs. However, not all MRRMs are *compatible* in a way that their composition would produce another MRRM, and we here develop the theory needed to show that two MRRMs are compatible and to identify the MRRMs resulting from their composition. These theorems are instrumental for showing that several of the important classes of shuffling methods defined in Section II C are compatible, and thus that new MRRMs can be created by composing pairs of these shufflings.

In Subsection V A, we start by developing the theory needed to formally define the composition of two MRRMs, and we explore its properties. We next show that compatibility is equivalent to a certain form of conditional independence between the features constrained by the two MRRMs. We use this in the following subsection (Subsection V B) to show that certain of the classes of shuffling methods described in Section II C are compatible and to describe the MRRM that results from their composition. In Subsection V C, we finally list and characterize MRRMs found in the literature that are compositions of two other MRRMs.

TABLE IV.3: **Effects of metadata-dependent shufflings on features of temporal networks.** Special metadata symbols are the color (group affiliation) of a node, σ_i, and the group contact matrices $\Sigma_{\mathcal{L}}$ and $\Sigma_{\mathcal{E}}$ (see main text for their definition and Table IV.1 for other features). Colored symbols show to what extent each feature is conserved. Informal definitions are found in the tablenotes (detailed definitions are found in Supplementary Table S1).

Canonical name	Common name	meta			topological			weighted				temp.	node			link			
		σ_i	$\Sigma_{\mathcal{L}}$	$\Sigma_{\mathcal{E}}$	G^{stat}	k_i	L	$a_i^{\dagger}$	s_i	$n_{(i,j)}{}^{\dagger}$	$w_{(i,j)}$	A^t	$\alpha_i^{m\dagger}$	$\Delta\alpha_i^m$	d_i^t	$\tau_{(i,j)}^{m}{}^{\dagger}$	$\Delta\tau_{(i,j)}^m$	$t_{(i,j)}^1$	$t_{(i,j)}^w$
P[$p_{\mathcal{L}}(\boldsymbol{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$]	*Block LS*	x	x	–	–	μ	x	μ	μ	p	p	x	–	–	$\boldsymbol{\mu}_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
P[$\mathbf{k}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$,$\boldsymbol{\sigma}$,$\Sigma_{\mathcal{L}}$]	*Deg. + block LS*	x	x	–	–	x	x	μ	μ	p	p	x	–	–	$\boldsymbol{\mu}_{\mathcal{T}}$	$p_{\mathcal{L}}$	$p_{\mathcal{L}}$	p	p
P[G,$p(\boldsymbol{\sigma})$]	*Color shuffling*	p	x	–	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

† Feature only defined for temporal networks with event durations.
$^{\text{x}}$ Feature completely conserved, typically the ordered sequence of individual features, e.g. $\mathbf{x} = (x_i)_{i\in\mathcal{V}}$ or $\mathbf{x} = ((\mathbf{x}_i^t)_{t\in\mathcal{T}})_{i\in\mathcal{V}}$.
$^{p_{\mathcal{L}}}$ Distribution of local sequences on links, $p_{\mathcal{L}}(\mathbf{x}) = \left[(x_{(i,j)}^m)_{m\in\mathcal{M}_{(i,j)}}\right]_{(i,j)\in\mathcal{L}}$.
^{p} Distribution (i.e. the multiset) of individual scalar values in sequence.
$^{\boldsymbol{\mu}_{\mathcal{T}}}$ Sequence of local means in snapshots links, $\boldsymbol{\mu}_{\mathcal{T}}(\mathbf{x}) = \left(\sum_{i\in\mathcal{V}} x_i^t/N\right)_{t\in\mathcal{T}}$.
$^{\mu}$ Mean value of the individual scalar features in sequence.
$^{-}$ Feature not conserved.

A. Theory: Composition of MRRMs

In this section we explore how two different MRRMs may be combined in composition to generate another MRRM. This generates a RRM that is not necessarily microcanonical, but if it is, it shuffles more than either of the two. Especially the type of compositions that produce MRRMs are of practical interest as they provide a way of producing new MRRMs by combining existing shuffling algorithms.

The latter part of this section will be devoted to exploring under which conditions the composition of two MRRMs is microcanonical (we then say that the two MRRMs are *compatible*). We develop a concept called *conditional independence given a common coarsening* and show that it characterizes compatibility. Finally, we show that specific types of refinements of compatible MRRMs are also compatible and identify the way the resulting MRRM inherits the features of the two input models.

We here present propositions and theorems without proofs. They can be found in Appendix VIII.

1. *Composition of two MRRMs*

In practice, the composition of two MRRMs involves first applying one shuffling method to the input network G^*, and then applying the second shuffling method to the outputs of the first [Fig. V.1(a)]. This thus defines a composition of the two shuffling methods:

Definition V.1. *Composition of MRRMs.* Consider two MRRMs P[$\mathbf{x}$] and P[$\mathbf{y}$] and an input network $G^* \in \mathcal{G}$. The composition of P[$\mathbf{y}$] on P[$\mathbf{x}$], denoted P[$\mathbf{y}\circ\mathbf{x}$] is defined by the conditional probability:

$$P_{\mathbf{y}\circ\mathbf{x}}(G|G^*) = \sum_{G'\in\mathcal{G}} P_{\mathbf{y}}(G|G')P_{\mathbf{x}}(G'|G^*) = \sum_{G'\in\mathcal{G}} \frac{\delta_{\mathbf{y}(G),\mathbf{y}(G')}}{\Omega_{\mathbf{y}}(G')} \frac{\delta_{\mathbf{x}(G'),\mathbf{x}(G^*)}}{\Omega_{\mathbf{x}(G^*)}} \quad . \tag{6}$$

For a given indexing of the state space $\mathcal{G}$, Eq. (6) shows that the transition matrix (Def. II.11) for the composition of P[$\mathbf{y}$] on P[$\mathbf{x}$] is simply the matrix product of the individual transition matrices, $\mathbf{P}^{\mathbf{y}\circ\mathbf{x}} = \mathbf{P}^{\mathbf{y}}\mathbf{P}^{\mathbf{x}}$.

Definition V.2. *Compatibility.* We say that two MRRMs P[$\mathbf{x}$] and P[$\mathbf{y}$] are *compatible* if their composition P[$\mathbf{y}\circ\mathbf{x}$] is also a MRRM.

The notion of compatibility is central as it defines which MRRMs we may combine through composition to define a new MRRM.

Proposition V.1. *Compatible randomized reference models commute.* If two MRRMs, P[$\mathbf{x}$] and P[$\mathbf{y}$], are compatible then P[$\mathbf{y}\circ\mathbf{x}$] = P[$\mathbf{x}\circ\mathbf{y}$].

Proposition V.1 means that it does not matter in which order we apply two compatible MRRMs in the composition, and consequently that P[$\mathbf{y}\circ\mathbf{x}$] and P[$\mathbf{x}\circ\mathbf{y}$] define the same MRRM if P[$\mathbf{x}$] and P[$\mathbf{y}$] are compatible. It also means that in order to show that two MRRMs are not compatible, it suffices to show that they do not commute.

Example V.1. Let the state space $\mathcal{G}$ be all static graphs with 3 nodes. As in Example II.7, we number the 8 graphs such that G_1 is the graph with 0 links, G_2, G_3, and G_4 are the graphs with 1 link, G_5, G_6, and G_7 are the graphs with 2 links and G_8 is the graph with 3 links. Let us now define two MRRMs for this state space [Fig. V.1(b)]:

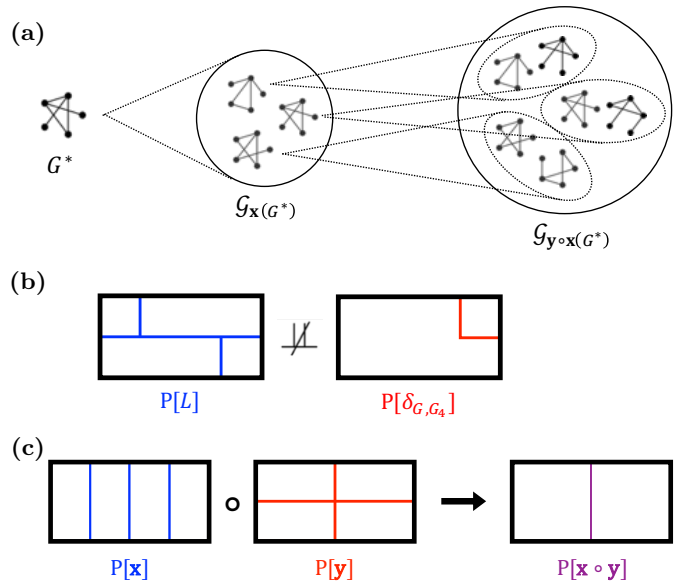


FIG. V.1: **Composition of MRRMs.** (a) The composition P[$\mathbf{x}\circ\mathbf{y}$] of two shuffling methods is implemented by applying P[$\mathbf{x}$] to the outputs of P[$\mathbf{y}$]. (b) Example of incompatible partitions of a state space of 8 states (e.g. all simple graphs with 3 nodes). (c) The composition P[$\mathbf{x}\circ\mathbf{y}$] of two compatible MRRMs, P[$\mathbf{x}$] and P[$\mathbf{y}$], shuffles more than (or as much as) either of the two. In terms of partitions, P[$\mathbf{x}\circ\mathbf{y}$] produces the least upper bound of P[$\mathbf{x}$] and P[$\mathbf{y}$].

1. The MRRM P[L], defined by the number of links L in the graph, partitions the state space in 4 sets $\mathcal{G}_0 = \{G_1\}$, $\mathcal{G}_1 = \{G_2, G_3, G_4\}$, $\mathcal{G}_2 = \{G_5, G_6, G_7\}$, and $\mathcal{G}_3 = \{G_8\}$.

2. The MRRM P[δ_{G,G_4}] keeps the graph G_4 unchanged and shuffles all the others. It is defined by the feature δ_{G,G_4} which returns 1 when applied to G_4 and 0 otherwise. This MRRM partitions $\mathcal{G}$ into two partitions $\mathcal{G}_1 = \{G_4\}$ and $\mathcal{G}_0 = \{G_1, G_2, G_3, G_5, G_6, G_7, G_8\}$.

With these definitions P[$L\circ\delta_{G,G_4}$] $\neq$ P[$\delta_{G,G_4}\circ L$] . For example, P[$L\circ\delta_{G,G_4}$] applied to G_4 can return the states $\{G_2, G_3, G_4\}$, while the application of P[$\delta_{G,G_4}\circ L$] can return the entire $\mathcal{G}$. So the two MRRMs do not commute and are thus not compatible. Consequently, the RRM obtained by composition of P[δ_{G,G_4}] and P[L] is not microcanonical. It is in the above example also easy to verify e.g. that the states generated by P[$\delta_{G,G_4}\circ L$] are not equiprobable (e.g. $P_{\delta_{G,G_4}\circ L}(G_4) = 1/3$ while $P_{\delta_{G,G_4}\circ L}(G_i) = 2/21$ for all other graphs).

2. *Comparability and compatibility*

The composition of two compatible MRRMs produces a MRRM that is comparable to the two and, in particular, one that randomizes more than each of them individually (i.e. one that is *coarser*).

Proposition V.2. *Composition of two compatible MRRMs always results in a MRRM which does not shuffle less.* Consider two compatible MRRMs, P[$\mathbf{x}$] and P[$\mathbf{y}$]. Their composition, P[$\mathbf{y}\circ\mathbf{x}$], is coarser than (or equal to) both P[$\mathbf{y}$] and P[$\mathbf{x}$], i.e. P[$\mathbf{y}\circ\mathbf{x}$] $\geq$ P[$\mathbf{x}$] and P[$\mathbf{y}\circ\mathbf{x}$] $\geq$ P[$\mathbf{y}$], even if P[$\mathbf{x}$] and P[$\mathbf{y}$] are not comparable.

In order for the concept of compatibility to be practically useful we need to be able to find out which MRRMs are compatible and what the result of their compositions is. Comparable MRRMs are an easy special case in this regard, as all comparable MRRMs turn out to be compatible and their composition simply yields the MRRM that shuffles more.

Proposition V.3. *Comparable microcanonical randomized reference models are compatible.* Let P[$\mathbf{x}$] and P[$\mathbf{y}$] be two MRRMs satisfying P[$\mathbf{x}$] $\leq$ P[$\mathbf{y}$]. Then they are compatible and their composition gives P[$\mathbf{y}\circ\mathbf{x}$] = P[$\mathbf{y}$].

Example V.2. Consider again the MRRMs P[L] and P[$\mathbf{k}$] from Example IV.1. Since they are comparable, they are compatible according to Proposition V.3. Consequently they also commute (Proposition V.1) and P[$\mathbf{k}\circ L$] = P[$L\circ\mathbf{k}$] = P[L].

The effect of the composition operation works in the opposite manner to the intersection of MRRMs (Def. II.9). This is also seen in the effect of composition with the zero and unity elements (which are compatible with all MRRMs by Proposition V.3): here zero is the neutral element P[$0\circ\mathbf{x}$] = P[$\mathbf{x}$] and unity is the absorbing element P[$1\circ\mathbf{x}$] = P[1], whereas P[0] is the neutral element and P[1] the absorbing element for intersection (Sec. IV A 1). Furthermore, by Proposition V.2, the composition gives an upper bound for the two MRRMs (Fig. V.1). In fact, the bound is the least upper bound, and any set of compatible MRRMs forms a *lattice* [89], but this connection to the theory of partially ordered sets is not pursued further here.

3. *Conditional independence and compatibility*

Our aim in this section is to be able to compose MRRMs to produce new ones, and even though comparable MRRMs are always compatible they are not useful for this purpose as they do not produce a new MRRM. There are more interesting compositions, but in order to be able to access these we need a way of characterizing which pairs of MRRMs are compatible outside of comparable ones. We will next define the concept of conditional independence between two features given a common coarsening of these and show in Theorem 1 that it is equivalent to compatibility. We next show in Theorem 2 that certain refinements of compatible MRRMs (termed *adapted*

refinements) are themselves compatible. Theorem 2 furthermore shows which features their composition inherits from the original MRRMs.

Before we can define the concept of conditional independence given a common coarsening, we first need to define the concepts of conditional probability of a feature and conditional independence between features.

Definition V.3. *Conditional probability of a feature.* The conditional probability of a feature $\mathbf{y}$ given another feature $\mathbf{x}$ is the probability $P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^\dagger|\mathbf{x}^*)$ that the feature $\mathbf{y}$ takes the value $\mathbf{y}^\dagger$ conditioned on the value $\mathbf{x}^*$ of the feature $\mathbf{x}$. It is given by

$$P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^\dagger|\mathbf{x}^*) = \sum_{G'\in\mathcal{G}} \delta_{\mathbf{y}^\dagger,\mathbf{y}(G')} P_{\mathbf{x}}(G'|\mathbf{x}^*) = \frac{\Omega_{(\mathbf{y}^\dagger,\mathbf{x}^*)}}{\Omega_{\mathbf{x}^*}} \ . \tag{7}$$

The conditional probability of a feature satisfies all properties of usual conditional probabilities. We may notably relate the composition of two MRRMs to the conditional probability of their features using the law of total probability as $P_{\mathbf{y}\circ\mathbf{x}}(G|\mathbf{x}^*) = \sum_{\mathbf{y}^\dagger} P_{\mathbf{y}}(G|\mathbf{y}^\dagger) P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^\dagger|\mathbf{x}^*)$. It also allows us to define the conditional independence in the usual sense as when $P_{\mathbf{y}|\mathbf{x},\mathbf{z}}(\mathbf{y}^\dagger|\mathbf{x}^\ddagger,\mathbf{z}^*) = P_{\mathbf{y}|\mathbf{z}}(\mathbf{y}^\dagger|\mathbf{z}^*)$ for a given a third feature $\mathbf{z}$. We shall here be concerned with a stricter version of conditional independence which is satisfied when the feature $\mathbf{z}$ is coarser than both $\mathbf{y}$ and $\mathbf{x}$. This *conditional independence given a common coarsening* is equivalent to $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ being compatible.

Definition V.4. *Conditional independence given a common coarsening.* If there exist a feature $\mathbf{z}$ that is a common coarsening of both $\mathbf{x}$ and $\mathbf{y}$ (i.e. $\mathbf{z} \geq \mathbf{x}$ and $\mathbf{z} \geq \mathbf{y}$) such that $P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^\dagger|\mathbf{x}(G^*)) = P_{\mathbf{y}|\mathbf{z}}(\mathbf{y}^\dagger|\mathbf{z}(G^*))$ for all $G^* \in \mathcal{G}$, we will say that $\mathbf{y}$ is conditionally independent of $\mathbf{x}$ given their common coarsening $\mathbf{z}$.

As for the usual conditional independence, the conditional independence given a common coarsening defined above is symmetric in $\mathbf{x}$ and $\mathbf{y}$.

Proposition V.4. *Symmetry of the conditional independence given a common coarsening.* If $\mathbf{x}$ is conditionally independent of $\mathbf{y}$ given a common coarsening $\mathbf{z}$ then $\mathbf{y}$ is conditionally independent of $\mathbf{x}$ given $\mathbf{z}$

Because of the symmetry, we can simply say that $\mathbf{x}$ and $\mathbf{y}$ are conditionally independent given the common coarsening $\mathbf{z}$.

As we stated above, the concept of conditional independence given a common coarsening is important because it is a characterization of compatibility. The following theorem proves this.

Theorem 1. *Conditional independence given a common coarsening is equivalent to compatibility.* $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ are compatible if and only if they are conditionally independent given the common coarsening $\mathbf{z} = \mathbf{x} \circ \mathbf{y}$.

Conditional independence is important in practice for designing MRRMs that randomize both the topology and the time-domain of a temporal network by implementing them as compositions of MRRMs that individually randomize either the topological or temporal aspects of a network (see Section V B below). The following example illustrates the concepts of conditional independence and compatibility in terms of the abstract state space.

Example V.3. Consider a state space with 8 states, $\mathcal{G} = \{G_1, \ldots, G_8\}$, which are placed into a square formation such that the states 1 to 4 are in the first row and 5 to 8 are in the third row. Now we can define two features: f_c that returns the column number, and f_q that returns the quadrant that the state is in. The partitions these two features induce are $\mathcal{G}_{f_c} = \{\{G_1, G_5\}, \{G_2, G_6\}, \{G_3, G_7\}, \{G_4, G_8\}\}$ for f_c and $\mathcal{G}_{f_c} = \{\{G_1, G_2\}, \{G_3, G_4\}, \{G_5, G_6\}, \{G_7, G_8\}\}$ for f_q [Fig. V.1(c)]. The two features f_c and f_q are conditionally independent given the function f_h that indicates which half of the state space the state is in. Thus, the corresponding MRRMs, $\mathrm{P}[f_c]$ and $\mathrm{P}[f_q]$, are compatible and their composition is the MRRM $\mathrm{P}[f_h]$ [Fig. V.1(c)].

Our main aim when defining compositions has been to be able to produce new useful MRRMs. With the help of the concept of conditional independence we are now ready to write down a theorem that will allow us to compose non-comparable MRRMs and know the features of the resulting model. To do so we define a special type of *adapted* refinements of compatible MRRMs which we can show are also compatible.

Definition V.5. *Adapted refinement.* Consider two compatible MRRMs, $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$. Any refinement of $\mathrm{P}[\mathbf{x}]$ of the form $\mathrm{P}[\mathbf{x}, \mathbf{f}(\mathbf{y})]$, where $\mathbf{f}$ is any function of $\mathbf{y}$, is said to be *adapted* to $\mathrm{P}[\mathbf{y}]$. We will refer to $\mathrm{P}[\mathbf{x}, \mathbf{f}(\mathbf{y})]$ as an *adapted refinement of* $\mathrm{P}[\mathbf{x}]$ *with respect to* $\mathrm{P}[\mathbf{y}]$.

In the following theorem we will now demonstrate that all adapted refinements of compatible MRRMs are themselves compatible, as well as showing which features the composition of such MRRMs inherits from the individual MRRMs. This theorem will be very useful in practice: if we can show that a given pair of MRRMs are compatible (i.e. using Theorem 1), we get for free that a whole class of MRRMs, consisting of all adapted refinements of the original MRRMs, are also compatible.

Theorem 2. *Adapted refinements of compatible MRRMs are compatible.* Consider two compatible MRRMs $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$, and any adapted refinements of these, $\mathrm{P}[\mathbf{y}, \mathbf{f}(\mathbf{x})]$ and $\mathrm{P}[\mathbf{x}, \mathbf{g}(\mathbf{y})]$. Then $\mathrm{P}[\mathbf{y}, \mathbf{f}(\mathbf{x})]$ and $\mathrm{P}[\mathbf{x}, \mathbf{g}(\mathbf{y})]$ are compatible, and their composition is given by $\mathrm{P}[\mathbf{x} \circ \mathbf{y}, \mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{y})]$.

B. Compositions of shuffling methods

Theorems 1 and 2 let us show with the two following propositions that all link shufflings are compatible with

any timeline shuffling, and likewise that all sequence shufflings are compatible with any snapshot shuffling. We may thus combine pairs of such shufflings to generate new MRRMs that randomize both the topological and temporal domains of a temporal network in different ways.

Formal proofs for the propositions are given in Appendix VIII. In essence they boil down to first using Theorem 1 to show that the coarsest (i.e. most random) link and timeline (sequence and snapshot) shufflings are compatible. Second, by noting that all link and timeline (sequence and snapshot) shufflings are adapted refinements (Def. V.5) of these, Theorem 2 gives that these are also compatible.

Proposition V.5. *Link shufflings and timeline shufflings are compatible.* Any link shuffling $\mathrm{P}[\mathbf{f}(\mathcal{L}), \mathbf{\Theta}]$ and timeline shuffling $\mathrm{P}[\mathcal{L}, \mathbf{g}(\mathbf{\Theta})]$ are compatible and their composition is given by $\mathrm{P}[L, \mathbf{f}(\mathcal{L}), \mathbf{g}(\mathbf{\Theta})]$.

Example V.4. The composition of the coarsest link shuffling $\mathrm{P}[p_{\mathcal{L}}(\mathbf{\Theta})]$ with the coarsest timeline shuffling $\mathrm{P}[\mathcal{L}, E]$ results in the MRRM $\mathrm{P}[L, E]$ which randomizes both the static topology and the temporal order of events while conserving the number of links L in the static graph

Proposition V.6. *Sequence shufflings and snapshot shufflings are compatible.* Any sequence shuffling $\mathrm{P}[\mathbf{f}(\mathbf{t}), p_{\mathcal{T}}(\mathbf{\Gamma})]$ and snapshot shuffling $\mathrm{P}[\mathbf{t}, \mathbf{g}(p_{\mathcal{T}}(\mathbf{\Gamma}))]$ are compatible and their composition is given by $\mathrm{P}[p(\mathbf{A}), \mathbf{f}(\mathbf{t}), \mathbf{g}(p_{\mathcal{T}}(\mathbf{\Gamma}))]$.

Example V.5. The composition of the coarsest sequence shuffling $\mathrm{P}[p_{\mathcal{T}}(\mathbf{\Gamma})]$ with the coarsest snapshot shuffling $\mathrm{P}[\mathbf{t}]$ results in the MRRM $\mathrm{P}[p(\mathbf{A})]$ which randomizes both the topology of snapshots and their temporal order while conserving the multiset of the numbers of events per snapshot, $p(\mathbf{A}) = [|\mathcal{E}^t|]_{t\in\mathcal{T}}$.

C. Examples of compositions of temporal network MRRMs

As we showed in the previous subsection, link shufflings are compatible with timeline shufflings (Subsec. IV B 4), and sequence shufflings (Subsec. IV B 5) are compatible with snapshot shufflings (Subsec. IV B 6).

We list here examples we have found in the literature of MRRMs that are compositions of two compatible MRRMs. Table V.1) lists the effects of these MRRMs on a selection of important temporal network features (see Table IV.1 for formal definitions of the features).

Many more MRRMs than those listed here may be generated directly from the MRRMs surveyed in Section IV B as any composition of a pair of a link and a timeline shuffling or of a sequence and a snapshot shuffling forms a new MRRM.

1. Instant-event shufflings

$\mathrm{P}[L, E]$ **Composition of:** $\mathrm{P}[p_{\mathcal{L}}(\mathbf{\Theta})]$ (Sec. IV B 3) and $\mathrm{P}[\mathcal{L}, E]$ (Sec. IV B 4). **Features constrained:** the number of links in static graph, L; the number of instantaneous events, E. **Reference:** Ref. [52] (*all random*).

$\mathrm{P}[\mathbf{k}, E]$ **Composition of:** $\mathrm{P}[\mathbf{k}, p_{\mathcal{L}}(\mathbf{\Theta})]$ (Sec. IV B 3) and $\mathrm{P}[\mathcal{L}, E]$ (Sec. IV B 4). **Features constrained:** the static degree sequence $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$; the number of instantaneous events, E. **Reference:** Ref. [59].

$\mathrm{P}[\mathbf{k},p(\mathbf{w}),\mathbf{t}]$ **Cothe mposition of:** $\mathrm{P}[\mathbf{k},p_{\mathcal{L}}(\mathbf{\Theta})]$ (Sec. IV B 3) and $\mathrm{P}[\mathbf{w},\mathbf{t}]$ (Sec. IV B 8). **Features constrained:** the static degree sequence $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$; the multiset of link weights, $p(\mathbf{w}) = [w_{(i,j)}]_{(i,j)\in\mathcal{L}}$; the sequence of event timestamps, $\mathbf{t} = [t]_{(i,j,t)\in\mathcal{E}}$. **References:** [45, 56] (*randomized edges with randomly permuted times*).

$\mathrm{P}[\mathbf{k},\chi_\lambda,p(\mathbf{w}),\mathbf{t}]$ **Composition of:** $\mathrm{P}[\mathbf{k},\chi_\lambda,p_{\mathcal{L}}(\mathbf{\Theta})]$ (Sec. IV B 3) and $\mathrm{P}[\mathbf{w},\mathbf{t}]$ (Sec. IV B 8). **Features constrained:** the static degree sequence $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$; the connectedness of static graph, χ_λ; the multiset of link weights, $p(\mathbf{w}) = [w_{(i,j)}]_{(i,j)\in\mathcal{L}}$; the sequence of event timestamps, $\mathbf{t} = [t]_{(i,j,t)\in\mathcal{E}}$. **References:** [9, 21] (*configuration model*).

$\mathrm{P}[\mathcal{L}, p(\mathbf{w})]$ **Composition of:** $\mathrm{P}[\mathcal{L},p_{\mathcal{L}}(\mathbf{\Theta})]$ (Sec. IV B 7) and $\mathrm{P}[\mathbf{w}]$ (Sec. IV B 4). **Features constrained:** the static topology $G^{\mathrm{stat}} = (\mathcal{V}, \mathcal{L})$; the multiset of link weights, $p(\mathbf{w}) = [w_{(i,j)}]_{(i,j)\in\mathcal{L}}$; **Reference:** Section VII.

$\mathrm{P}[\mathcal{L}, p(\mathbf{w}), \mathbf{t}]$ **Composition of:** $\mathrm{P}[\mathcal{L},p_{\mathcal{L}}(\mathbf{\Theta})]$ (Sec. IV B 7) and $\mathrm{P}[\mathbf{w}, \mathbf{t}]$ (Sec. IV B 8). **Features constrained:** the static topology $G^{\mathrm{stat}} = (\mathcal{V}, \mathcal{L})$; the multiset of link weights, $p(\mathbf{w}) = [w_{(i,j)}]_{(i,j)\in\mathcal{L}}$; the sequence of event timestamps, $\mathbf{t} = [t]_{(i,j,t)\in\mathcal{E}}$. **Reference:** Section VII.

2. Metadata-dependent shufflings

$\mathrm{P}[\mathbf{k},p(\mathbf{w}),\mathbf{t},\boldsymbol{\sigma},\Sigma_{\mathcal{L}}]$ **Composition of:** $\mathrm{P}[p_{\mathcal{L}}(\mathbf{\Theta}),\mathbf{k},\boldsymbol{\sigma},\Sigma_{\mathcal{L}}]$ with $\mathrm{P}[\mathbf{w},\mathbf{t}]$. **Features constrained:** the static degree sequence $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$; the multiset of link weights, $p(\mathbf{w}) = [w_{(i,j)}]_{(i,j)\in\mathcal{L}}$; the sequence of event timestamps, $\mathbf{t} = [t]_{(i,j,t)\in\mathcal{E}}$; the sequence of node colors, $\boldsymbol{\sigma} = (\sigma_i)_{i\in\mathcal{V}}$; the the group contact matrix, $\Sigma_{\mathcal{L}}$. **Reference:** Ref. [10] (*random dynamic topological*).

VI. OTHER REFERENCE MODELS

We have restricted this review to microcanonical RRMs as they are the only maximum entropy reference models

TABLE V.1: **Effects of compositions of two MRRMs on the features of a temporal network.** See Table IV.1 for definitions of features. Colored symbols show to what extent each feature is conserved. Informal definitions are found in the tablenotes (detailed definitions are found in Supplementary Table S1).

Canonical name	Composition of	meta			topological			weighted				temp.	node			link			
		σ_i	$\Sigma_{\mathcal{L}}$	$\Sigma_{\mathcal{E}}$	G^{stat}	k_i	L	$a_i{}^{\dagger}$	s_i	$n_{(i,j)}{}^{\dagger}$	$w_{(i,j)}$	A^t	$\alpha_i^{m\dagger}$	$\Delta\alpha_i^m$	d_i^t	$\tau_{(i,j)}^{m}{}^{\dagger}$	$\Delta\tau_{(i,j)}^m$	$t_{(i,j)}^1$	$t_{(i,j)}^w$
$\mathrm{P}[L, E]$	$\mathrm{P}[p_{\mathcal{L}}(\boldsymbol{\Theta})] \circ \mathrm{P}[\mathcal{L}, E]$	–	–	–	–	μ	x	–	μ	–	μ	μ	–	–	μ	–	–	–	–
$\mathrm{P}[\mathbf{k}, p(\mathbf{w}), \mathbf{t}]$	$\mathrm{P}[\mathbf{k}, p_{\mathcal{L}}(\boldsymbol{\Theta})] \circ \mathrm{P}[\mathbf{w}, \mathbf{t}]$	–	–	–	–	x	x	–	μ	–	p	x	–	–	$\boldsymbol{\mu}_{\mathcal{T}}$	–	–	–	–
$\mathrm{P}[\mathbf{k}, \chi_\lambda, p(\mathbf{w}), \mathbf{t}]$	$\mathrm{P}[\mathbf{k}, \chi_\lambda, p_{\mathcal{L}}(\boldsymbol{\Theta})] \circ \mathrm{P}[\mathbf{w}, \mathbf{t}]$	–	–	–	χ_λ	x	x	–	μ	–	p	x	–	–	$\boldsymbol{\mu}_{\mathcal{T}}$	–	–	–	–
$\mathrm{P}[\mathcal{L}, p(\mathbf{w})]$	$\mathrm{P}[\mathcal{L}, p_{\mathcal{L}}(\boldsymbol{\Theta})] \circ \mathrm{P}[\mathbf{w}]$	–	–	–	x	x	x	–	μ	–	p	μ	–	–	μ	–	–	–	–
$\mathrm{P}[\mathcal{L}, p(\mathbf{w}), \mathbf{t}]$	$\mathrm{P}[\mathcal{L}, p_{\mathcal{L}}(\boldsymbol{\Theta})] \circ \mathrm{P}[\mathbf{w}, \mathbf{t}]$	–	–	–	x	x	x	–	μ	–	p	x	–	–	$\boldsymbol{\mu}_{\mathcal{T}}$	–	–	–	–
Metadata dependent:																			
$\mathrm{P}[\mathbf{k}, p(\mathbf{w}), \mathbf{t}, \boldsymbol{\sigma}, \Sigma_{\mathcal{L}}]$	$\mathrm{P}[\mathbf{k}, p_{\mathcal{L}}(\boldsymbol{\Theta}), \boldsymbol{\sigma}, \Sigma_{\mathcal{L}}] \circ \mathrm{P}[\mathbf{w}, \mathbf{t}]$	x	x	–	–	x	x	–	μ	–	p	x	–	–	$\boldsymbol{\mu}_{\mathcal{T}}$	–	–	–	–

† Feature only defined for temporal networks with event durations.
x Feature completely conserved, typically the ordered sequence of individual features, e.g. $\mathbf{x} = (x_i)_{i\in\mathcal{V}}$ or $\mathbf{x} = ((\mathbf{x}_i^t)_{t\in\mathcal{T}})_{i\in\mathcal{V}}$.
$p_{\mathcal{L}}$ Distribution of local sequences on links, $p_{\mathcal{L}}(\mathbf{x}) = \left[(x_{(i,j)}^m)_{m\in\mathcal{M}_{(i,j)}}\right]_{(i,j)\in\mathcal{L}}$.
p Distribution (i.e. the multiset) of individual scalar values in sequence.
$\boldsymbol{\mu}_{\mathcal{T}}$ Sequence of local means in snapshots links, $\boldsymbol{\mu}_{\mathcal{T}}(\mathbf{x}) = \left(\sum_{i\in\mathcal{V}} x_i^t/N\right)_{t\in\mathcal{T}}$.
μ Mean value of the individual scalar features in sequence.
– Feature not conserved.

that can be generated by shuffling elements of an empirical temporal network and they constitute the largest part of RRMs for temporal networks found in the literature.

In this section, we briefly discuss other types of reference models for temporal networks. These models can be divided into three general classes: (1) *canonical* RRMs, which correspond to generalized canonical ensembles of random networks; (2) data-driven reference models that do not maximize entropy; (3) bootstrap methods, which are a particular type of reference models that do not maximize entropy.

A. Canonical randomized reference models

Canonical RRMs present alternatives that are very close in spirit to the microcanonical RRMs considered here. They permit to sample canonical ensembles of networks, i.e. ensembles where selected features are constrained only on average, $\langle \mathbf{x}(G) \rangle = \mathbf{x}(G^*)$, instead of exactly, $\mathbf{x}(G) = \mathbf{x}(G^*)$, as is the case for MRRMs. (One often talks of *soft* constraints for the canonical ensemble and *hard* constraints for the microcanonical ensemble). Such canonical generative models are also known as *exponential random graph models* (ERGMs) [2, 93] and allow to model the expected variability between samples (see discussion in [94, Section 4]). Additionally, due their soft constraints, canonical models are typically more amenable to analytical treatment than their microcanonical counterparts [95].

Conversely, the main advantage of MRRMs is that they are usually defined as data shuffling methods, which are often easier to construct than methods that generate networks from scratch. They are thus generally the only type of models that realistically capture many of the temporal and topological correlations present in empirical networks, which explains their popularity for analyzing temporal networks. In particular, it is easy to generate microcanonical RRMs that impose features such as the global activity timeline $\mathbf{A}$ or temporal correlations in individual timelines. Perhaps due to the difficulty in defining generative reference models that capture empirical temporal correlations, these are currently almost exclusively defined for static networks or to model either memoryless dynamics [9, 22, 51, 84] or dynamics with limited temporal correlations [96–100] A notable exception is a recent study combining Markov chains with change point detection to model multiscale temporal dynamics [101]. We shall not discuss canonical RRMs in more detail here, but refer to [93] for a recent review of ERGMs for temporal networks and to [94, 102] for recent developments in such models for static networks.

B. Reference models that do not maximize entropy

Several reference models exist that impose a constraint that is not justified solely by the data (the empirical temporal network) in conjunction with the maximum entropy principle [66]. Such reference models thus introduce new order that is not found in the original network. Here we discuss different types of such reference models and give examples.

Delta function constraints. Some studies have considered reference models where what we may call a *delta function constraint* was imposed on a set of features of the temporal network. Specifically they constrained all instances of this feature to have the same value, i.e. to

follow a delta distribution. This is different from (and more constrained than) the maximum entropy distribution. The *SStat* method introduced in Ref. [12] imposes a fixed number of events in each snapshot (equal to the mean number of events per snapshot in the empirical network). Holme [17] introduced three reference models that all three impose a delta-function constraint (referred to as *poor man's reference models* since they do not satisfy the maximum entropy principle and provide only a single reference network instead of an ensemble [47]): equalizing the inter-event durations $\Delta\tau^m_{(i,j)}$ while constraining $t^1_{(i,j)}$, $t^w_{(i,j)}$ and $w_{(i,j)}$ for each link $(i,j) \in \mathcal{L}$; shifting the whole sequence of events (sequences of event and inter-event times) on each link in order to make $t^1_{(i,j)} = t_{\min}$ or to make $t^w_{(i,j)} = t_{\max}$ for all $(i,j) \in \mathcal{L}$.

Biased sampling. Kovanen et al. [4] proposed a biased version of P[**w**,**t**], where instead of swapping timestamps of events at random, for each instantaneous event (i,j,t) they drew m other events at random from the set of instantaneous events $\mathcal{E}$ and swapped the timestamps of (i,j,t) and the other event (i',j',t') among the m drawn for which t' was closest to t. This reference model thus retains some temporal correlations due to the biased sampling, where the parameter m controls the force of this bias and thus of temporal correlations (for $m = 1$ the reference model is equal to P[**w**,**t**]). The same method was also employed in Refs. [7, 46]. Valdano et al. [25] considered a heuristic variant of P[$p_{\mathcal{T}}(\mathbf{g})$] (called *reshuffle-social*, where they only permuted snapshots inside intervals where nodes showed approximately the same median *social strategy* [103], where the social strategy of a node i is defined as the ratio $\gamma^t_i = k^{\delta,t}_i / s^{\delta,t}_i$ of its degree $k^{\delta,t}_i$ and its strength $s^{\delta,t}_i$ in a network aggregated over $\delta = 20$ consecutive snapshots from $t - \delta\Delta t$ to t. The empirical temporal network that they investigated showed very clear spikes in γ^t_i separated by low-γ^t_i intervals, referred to as γ-slices, which allowed them to permute snapshots within each γ-slice only.

Time reversal. A quick but informal way to gain insight into the role of causality in the interaction dynamics is to reverse the order of the snapshots [4, 45, 47, 56]. This method obviously does not increase entropy as the time-reversed network is unique, but it may be used as a simple way to study the importance of causality in the temporal network. (Note that a *time-reversal* MRRM may in principle be defined as one that returns an input temporal network and its time-reversed version with equal probability.)

C. Bootstrap methods

Bootstrap methods are based on sampling with replacement, whereas MRRMs are based on sampling without replacement (i.e. shuffling). Resampling with replacement means that network features are not constrained exactly as for shuffling methods. This means that bootstrapping algorithms may be easier to implement than shuffling methods when the exact constraints are hard to satisfy. The hope when using bootstrapping can additionally be to capture some of the expected out-of-sample variability. The set of states that may be generated is strongly constrained by the particular dataset however, so bootstrapping does not generate a maximum entropy model. Though it may be seen as a means to approximate one, it does not come with the same statistical guarantees as microcanonical and canonical RRMs do. So the theoretical results and guarantees that exist for microcanonical RRMs do not hold for bootstrapping, and additional care is advised when analyzing results obtained using bootstrapping.

Two bootstrap methods used in the literature are described below. The method called *time shuffling* in Ref. [14] constrains the number of events per link **n** exactly and resamples the event durations τ from the global distribution $p(\boldsymbol{\tau})$ with replacement. The method called *time shuffling* in Ref. [24] constrains the static network G^{stat} and bootstraps $n_{(i,j)}$, $t^1_{(i,j)}$ for all links from the global distributions $p(\mathbf{n})$ and $p(\mathbf{t}^1)$, respectively, and then bootstraps the $n_{(i,j)}$ event durations $\tau^m_{(i,j)}$ and $(n_{(i,j)} - 1)$ of inter-event durations $\Delta\tau^m_{(i,j)}$ for each link $(i,j) \in \mathcal{L}$ from the global distributions $p(\boldsymbol{\tau})$ and $p(\boldsymbol{\Delta\tau})$, respectively.

VII. ANALYZING TEMPORAL DISTANCES IN A COMMUNICATION NETWORK USING A SERIES OF MRRMS

In this section we go through a walk-through example in which we use the hierarchy of MRRMs to investigate how different features of a temporal communication network affects the temporal distances between nodes in the network (defined as the minimal times required for any contagion process to spread between the nodes). This example additionally serves to showcase a graphical representation that incorporates both the hierarchy of the MRRMs and their effects on a scalar feature, and which provides an intuitive way to interpret the results (Fig. VII.1). As discussed in Section III, understanding how different features affect spreading was the starting point of some of the early studies employing MRRMs in temporal networks, and here we reproduce some of those results with a different data set. However, the analysis pipeline introduced here does not only work for temporal distances, but can be used for any other scalar-valued feature.

The dataset is a publicly available temporal mobile phone communication network published by Wu et al. [104]. Here we focus on the first company with 44431 nodes and around 5.5×10^5 instantaneous events taking place over 30 days. Distances in temporal networks is a multifaceted topic [60], but here we quantify the distances in a network by a single number describing the typical temporal distance in the network. More specifically, we calculate the expected temporal distance to

TABLE VII.1: Effects of selected MRRMs on temporal network features. The features considered are: the static graph $G^{\text{stat}} = (\mathcal{V}, \mathcal{L})$; the link weights and their configuration in G^{stat}, $\mathbf{w} = (w_{(i,j)})_{(i,j)\in\mathcal{L}}$; the activity timeline $\mathbf{A} = (|\mathcal{E}^t|)_{t\in\mathcal{T}}$; higher order temporal correlations in and between timelines, i.e. $\mathbf{\Theta} = (\Theta_{(i,j)})_{(i,j)\in\mathcal{L}}$; the inter-event durations on the links, $\mathbf{\Delta\tau} = ((\Delta\tau^m_{(i,j)})_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$, where $\mathcal{M}_{(i,j)}$ is a temporal index; and the timing of the first event on each link $\mathbf{t}^1 = (t^1)_{(i,j)\in\mathcal{L}}$ (Table IV.1 provides detailed definitions each feature).

Model		Features					
Name	Common name	G^{stat}	$\mathbf{w}$	$\mathbf{A}$	$\mathbf{\Theta}$	$\mathbf{\Delta\tau}$	$\mathbf{t}^1$
$\mathrm{P}[\mathbf{w},p_{\mathcal{L}}(\mathbf{\Theta})]$	*Weight-constrained LS*	✓	✓	✓	p	p	p
$\mathrm{P}[\mathcal{L},p_{\mathcal{L}}(\mathbf{\Theta})]$	*Topology-constrained LS*	✓	p	✓	p	p	p
$\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\mathbf{\Delta\tau}),\mathbf{t}^1]$	*Inter-event shuffling*	✓	✓	μ	✗	p	✓
$\mathrm{P}[\mathbf{w},\mathbf{t}]$	*Timestamp shuffling*	✓	✓	✓	✗	✗	✗
$\mathrm{P}[\mathbf{w}]$	*Weight-constrained TS*	✓	✓	μ	✗	✗	✗
$\mathrm{P}[\mathcal{L},p(\mathbf{w}),\mathbf{t}]$	$\mathrm{P}[\mathcal{L},p_{\mathcal{L}}(\mathbf{\Theta})]\circ\mathrm{P}[\mathbf{w},\mathbf{t}]$	✓	p	✓	✗	✗	✗
$\mathrm{P}[\mathcal{L},p(\mathbf{w})]$	$\mathrm{P}[\mathcal{L},p_{\mathcal{L}}(\mathbf{\Theta})]\circ\mathrm{P}[\mathbf{w}]$	✓	p	μ	✗	✗	✗

✓ Feature completely conserved.
^{p} Distribution (i.e. the multiset) of individual values in sequence conserved.
$^\mu$ Mean value of the individual features in sequence conserved.
✗ Feature not conserved.

reach half of the nodes in the network, i.e. the *expected median temporal distance* $\langle d_{1/2}(G)\rangle$, where the expectation is evaluated over all nodes and all times as source points. The temporal distance from one node i to another node j is defined as the time required for the fastest possible spreading process starting at a given time t to spread from i to j[6]. Formally, this fastest possible spreading is modeled by a deterministic susceptible-infectious (SI) process where susceptible nodes are infected immediately by contact with an infectious node. When evaluating the distances we use periodic boundary conditions in time to remove boundary effects [9].

The MRRMs explored are listed in Table VII.1. The table also lists to which extent the MRRMs conserve selected network features. It is constructed formally by showing for feature and each MRRM whether or not the feature can be defined as function of the features constrained by the MRRM. For example, because the activity timeline $\mathbf{A} = (|\mathcal{E}^t|)_{t\in\mathcal{T}}$ can be calculated from the multiset of timelines $p_{\mathcal{L}}(\mathbf{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$, the MRRM $\mathrm{P}[\mathcal{L},p_{\mathcal{L}}(\mathbf{\Theta})]$ conserves $\mathbf{A}$. Similarly, $p(\mathbf{w}) = [w_{(i,j)}]_{(i,j)\in\mathcal{L}}$ is known from $p_{\mathcal{L}}(\mathbf{\Theta})$, but $\mathbf{w} = (w_{(i,j)})_{(i,j)\in\mathcal{L}}$ is not, so $\mathrm{P}[\mathcal{L},p_{\mathcal{L}}(\mathbf{\Theta})]$ conserves $p(\mathbf{w})$.

Figure VII.1 displays $\langle d_{1/2}(G)\rangle$ for the original data and for several MRRMs. The figure is organized in a way that the hierarchies (see Section IV) are visible similar to Figure IV.27. Reading the figure from top to bottom now yields a picture of what happens when the original data is shuffled more and more, i.e., when the temporal features present in the data are destroyed one by one by the MRRMs. All of the arrows are pointing either almost directly downwards or down and left, which means that, for this network and this set of MRRMs, randomizing more never leads to longer temporal distances.

The overall activity sequence $\mathbf{t}$, including the daily and weekly changes in the activity, does not have a noticeable effect on the temporal distances on these MRRMs: Removing the constraint on $\mathbf{t}$ when going from $P[\mathbf{w},\mathbf{t}]$ to $P[\mathbf{w}]$ and from $P[\mathcal{L},p_{\mathcal{L}}(\mathbf{w}),\mathbf{t}]$ to $P[\mathcal{L},p_{\mathcal{L}}(\mathbf{w})]$ almost does not change the temporal distances. Similarly, shuffling the inter-event times, $\mathbf{\Delta\tau}$, while keeping the first activation time of each link, $\mathbf{t}^1$, with $P[\boldsymbol{\pi}_{\mathcal{L}}(\mathbf{\Delta\tau}),\mathbf{t}^1]$ barely changes $\langle d_{1/2}(G)\rangle$, showing that higher-order temporal correlations between events over the same link has a very small effect on the temporal distances of the original data.

Adding the shuffling of the weights of the network – i.e. replacing the feature that keeps the weights of the links, $\mathbf{w}$, with the one only keeping the links and the weight distribution, $\mathcal{L}$ and $p(\mathbf{w})$ – makes the temporal paths around 7–9 days faster. The pairs of MRRMs corresponding to this replacement are $P[\mathbf{w},\mathbf{t}]$ to $P[\mathcal{L},p_{\mathcal{L}}(\mathbf{w}),\mathbf{t}]$, $P[\mathbf{w}]$ to $P[\mathcal{L},p_{\mathcal{L}}(\mathbf{w})]$, and $P[\mathbf{w},p_{\mathcal{L}}(\Theta)]$ to $P[\mathcal{L},p_{\mathcal{L}}(\Theta)]$. Note that in the MRRM $P[\mathcal{L},p_{\mathcal{L}}(\Theta)]$ the weight distribution $p_{\mathcal{L}}(\mathbf{w})$ is kept implicitly by the link sequence distribution $p_{\mathcal{L}}(\Theta)$, because $p_{\mathcal{L}}(\Theta) \leq p_{\mathcal{L}}(\mathbf{w})$.

Finally the largest change in the temporal distance are seen when the times of events in the timelines $\Theta_{(i,j)}$ are shuffled such that they simply follow the overall activity sequence $\mathbf{t}$. In these transitions, from $P[\mathbf{w},p_{\mathcal{L}}(\Theta)]$ to $P[\mathbf{w},\mathbf{t}]$ and from $P[\mathcal{L},p_{\mathcal{L}}(\Theta)]$ to $P[\mathcal{L},p_{\mathcal{L}}(\mathbf{w}),\mathbf{t}]$, the temporal distances are reduced on average by around 12–14 days.

Almost no combination effects were observed for these data: removing each feature had a very similar effect – with variations of around 2 days – independently of the other features that were kept. This allows a very simple summarization of the results: The typical temporal distance in the data is around 73 days and in the most random MRRM applied here around 48 days. Out of that difference, around 12–14 days is explained by link activation sequence features (such as bursts [9] and the timings of the links' first activations [17]), 7–9 days are by weight-topology correlations (such as weak links located in bridge positions [41, 79]), and 4 days by link-timeline-topology correlations (such as correlations in times at which two neighbors of a node are communicated with [9, 13]).

This analysis can be made more detailed by adding more fine-scaled features related to timings of events or link weights. Alternatively, the analysis could be expanded by including topological MRRMs such as the con-

[6] Note that the temporal distance is not a metric distance as the temporal distance from i to j generally differs from the temporal distance from j to i.

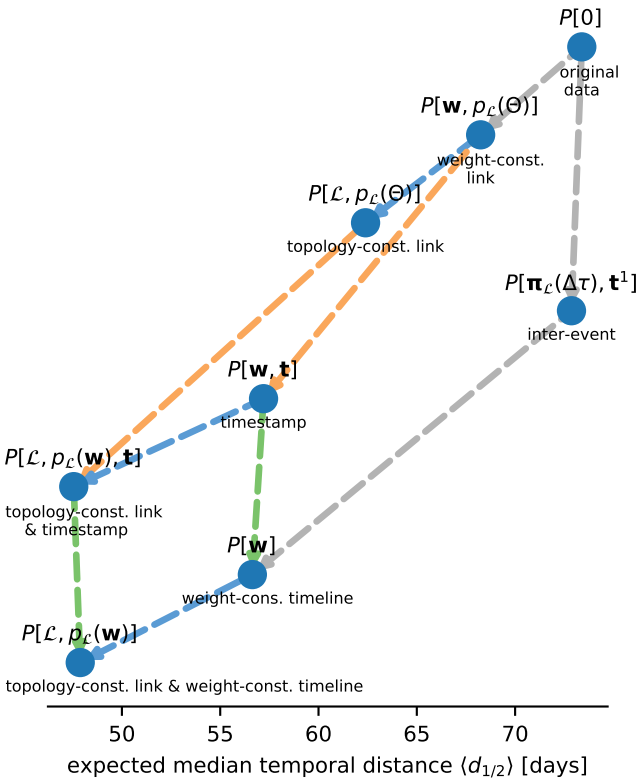


FIG. VII.1: **Expected median temporal distance for a hierarchy of MRRMs.** Each circle in the figure represents a single MRRM. The horizontal location of the circle reports the expected median temporal distance $\langle d_{1/2}(G) \rangle$ of the MRRM applied to the empirical network. An arrow from a MRRM at a higher location to a lower one means that the former shuffles less than the latter. (Note that the absolute positions of the nodes along the vertical axis are arbitrary and are thus not indicative of how random the MRRMs are. Only MRRMs linked by are directed path in the diagram can be formally compared.) A canonical name of each MRRM is given above each circle and a common name below it (see Section IV B). In the common names the word *shuffling* is always removed for brevity, and the &-sign denotes that both MRRMs are applied to the data in composition (Def. V.1). Colored links indicate that the same features were removed: $\mathbf{t}$ for green links, $\mathbf{w} \hookrightarrow \mathcal{L}, p_{\mathcal{L}}(\mathbf{w})$ for blue links, and $p_{\mathcal{L}}(\Theta) \hookrightarrow \mathbf{t}$ for orange links.

figuration model $\mathrm{P}[\mathbf{k}, p_{\mathcal{L}}(\boldsymbol{\Theta})]$.

VIII. CONCLUSION

Microcanonical randomized reference models (MRRMs) provide a versatile and generally applicable toolbox for the analysis of dynamical networked systems. Their main advantages are their wide applicability and relative ease of implementation: they only require the definition of a corresponding unbiased shuffling method, which is often easily implemented as a purely numerical randomization scheme. This means that they can be used to test the importance of any given feature provided a corresponding shuffling method, and may in principle be used to generate model networks that are arbitrarily close to empirical ones.

Shuffling methods provide an interesting alternative to more elaborate generative models, and can be seen as a *top-down* approach to modeling by progressively randomizing features of an empirical network, as opposed to the *bottom-up* approach of generative models. Each approach has its strengths and weaknesses (as discussed in Sec. VI A). We believe that shuffling methods are best used as exploratory tools to identify important qualitative features and effects. Generative models can then be used to explore them quantitatively and to perform model selection in order to identify potential underlying generative mechanisms.

We here introduced a fundamental framework for MRRMs. This enabled consistent naming, analysis, and classification of MRRMs for temporal networks. We have used this framework to describe numerical shuffling procedures found in the literature rigorously in terms of microcanonical RRMs, built a taxonomy of these RRMs, and surveyed their applications to the study of temporal networks. This framework also allowed us to define conditions for when we may combine two MRRMs in a composition to generate a new MRRM and to derive which features it inherits from them. Such compositions of compatible MRRMs make it possible to easily generate new MRRMs from the existing ones.

We have focused on undirected and unweighted temporal networks, but the extension of the MRRM framework introduced in Section II to any other types of network is trivial. Such extensions may require defining new ways of representing the structure and defining appropriate features. This is straightforward for temporal networks with directed (see Supplementary Table S2) or weighted events as well as networks with bipartite or even multipartite structure (Sec. IV B 9). Furthermore, a MRRM-based framework can be developed for any other types of multilayer networks [105], such as multiplex networks or networks of networks, and even for structures beyond networks such as hypergraphs. Finally, it should be helpful to define a similar framework for canonical reference models (see Section VI A) as more of such models are emerging.

It is our intention that this framework and collection of MRRMs will serve as a reference for researchers who want to employ MRRMs to analyze the dynamics of networks and if processes take place on them. It is straightforward to incorporate many more temporal network MRRMs into the framework than those presented here, e.g. models that constrain correlations between features in-

stead of only their marginal values. With the foundations for MRRMs laid here we are thus ready to repeat the success stories of RRMs for static networks, and may even go much further.

Notable important challenges remain which may now be addressed using the formalism defined here. For example, how to automatize the definition and classification of new MRRMs, which would allow a user to simply state the set of features she wants to constrain to generate a corresponding set of networks. How to automatize the choice of MRRMs in order to most efficiently infer which features of an empirical temporal network control a given dynamical phenomenon, i.e. identifying which models best divide the space of network features. How to compare MRRMs (e.g. in terms of sizes or overlap of partitions) that are neither comparable nor compatible. Being able to do this would notably make it possible to characterize automatically the effects of a MRRM on temporal network features that are not comparable to nor independent of the features constrained by the MRRM.

We note finally that it is a difficult problem to design unbiased shuffling methods for MRRMs that take higher order topological correlations into account [8, 106]. This may put natural barriers on the possible resolution of exact MRRMs. Instead, approximate procedures for generating such MRRMs would have to be considered, and their accuracy may be gauged by how closely they reproduce features which we know they should constrain.

ACKNOWLEDGMENTS

L.G. acknowledges support from the Lagrange Laboratory of the ISI Foundation funded by the CRT Foundation. M.G. was supported by the French ANR HARMS-flu, ANR-12-MONU-0018. M. Karsai acknowledges support from the DylNet (ANR-16-CE28-0013) and SoSweet (ANR-15-CE38-0011) ANR projects and the MOTIf STIC-AmSud project. T.T. was supported by the JST ERATO Grant Number JPMJER1201, Japan. C.L.V. was supported by the EU FET project MULTIPLEX 317532, the french ANR SiNCoBe, ANR-20-CE45-0021, and by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). M.G, T.T., and C.L.V. acknowledge financial support through the Bilateral Joint Research Program between MAEDI, France, and JSPS, Japan (SAKURA Program).

AUTHOR CONTRIBUTIONS

C.L.V. conceived and directed the study. All authors contributed to the definition of the classification system, to classifying existing MRRMs found in the literature, and to writing. M. Kivelä, T.T., and C.L.V. performed theoretical calculations. M.G. and M. Kivelä wrote the Python/C++ software packages. M.G., M. Karsai, M. Kivelä, T.T., and C.L.V. drafted the final manuscript.

APPENDIX: FORMAL DEFINITIONS AND PROOFS

This appendix provides formal definitions of the different classes of shuffling methods as well as proofs for the propositions and and theorems given in Sections IV A and V A.

1. Formal definitions of classes of shuffling methods

To formalize the set of possible constraints imposed by a class of shuffling methods, we shall use functions that take network features as input.

Definition A.1. *Function of a network feature.* A function of a temporal network feature $\mathbf{x}$ (Def. II.4) is any function $\mathbf{f}$ that has as domain the entire co-domain of $\mathbf{x}$, i.e. a function that takes $\mathbf{x}(G)$ as input and returns a value $\mathbf{f}(\mathbf{x}(G))$ for all $G \in \mathcal{G}$.

Example A.1. The sequence of static degrees $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$ of a temporal network G can be written as a function of the static graph $G^{\text{stat}}(G) = (\mathcal{V}(G), \mathcal{L}(G))$ (Def. II.5). In particular, it is a function of the configuration of links, $\mathcal{L}$, with each $k_i(G) \in \mathbf{k}(G)$ given by $k_i(G) = |\{i : (i,j) \in \mathcal{L}(G)\}|$.

Definition A.1 lets us formally define classes of shuffling methods using the naming convention developed in Def. II.10.

Definition A.2. *Event shuffling* $\mathrm{P}[p(\boldsymbol{\tau}), \mathbf{f}(\mathcal{C})]$. We define an *event shuffling* as a shuffling method that generates networks from an input temporal network by randomizing one or multiple of the indices i, j, t in all of the events $(i,j,t,\tau) \in \mathcal{C}$. Formally, any event shuffling is of the form $\mathrm{P}[p(\boldsymbol{\tau}), \mathbf{f}(\mathcal{C})]$. It thus constrains the multiset of the event durations of the network, $p(\boldsymbol{\tau}) = [\tau_q]_{q=1}^{C}$, as well as any additional constraint that can be written as a function $\mathbf{f}$ of the set of events $\mathcal{C}$.

Definition A.3. *Instant-event shuffling* $\mathrm{P}[E, \mathbf{f}(\mathcal{E})]$. We define an *instant-event shuffling* as a shuffling method that generates networks by randomizing one or multiple of the indices i, j, t in all of the instantaneous events $(i,j,t) \in \mathcal{E}$ of an instant-event temporal network. Formally, any instant-event shuffling is of the form $\mathrm{P}[E, \mathbf{f}(\mathcal{E})]$. It thus constrains the number of events, E, as well as any additional constraint that can be written as a function $\mathbf{f}$ of the set of instantaneous events $\mathcal{E}$.

a. Link and timeline shufflings

Definition A.4. *Link shuffling* $\mathrm{P}[\mathbf{f}(\mathcal{L}), \boldsymbol{\Theta}]$. A *link shuffling* constrains all the individual timelines, i.e. the multiset $p_{\mathcal{L}}(\boldsymbol{\Theta}) = [\Theta_{(i,j)}]_{(i,j)\in\mathcal{L}}$. It randomizes the links in the static graph, i.e. the values of i and j for each link $(i,j) \in \mathcal{L}$, while respecting a constraint given by any function $\mathbf{f}$ of the configuration of links $\mathcal{L}$.

Definition A.5. *Timeline shuffling* $\mathrm{P}[\mathcal{L}, \mathbf{f}(\boldsymbol{\Theta})]$. A *timeline shuffling* constrains the network's static topology, i.e. $G^{\mathrm{stat}} = (\mathcal{V}, \mathcal{L})$. It shuffles the events in the timelines while respecting a constraint given by any function $\mathbf{f}$ of the timelines $\boldsymbol{\Theta} = (\Theta_{(i,j)})_{(i,j)\in\mathcal{L}}$.

b. Sequence and snapshot shufflings

Definition A.6. *Sequence shuffling.* $\mathrm{P}[\mathbf{f}(\mathbf{t}), p_{\mathcal{T}}(\boldsymbol{\Gamma})]$. A *sequence shuffling* constrains the multiset of instantaneous snapshot graphs, $p_{\mathcal{T}}(\boldsymbol{\Gamma}) = [\Gamma^t]_{t\in\mathcal{T}}$. It randomizes the order of snapshots in a manner that may depend on any function $\mathbf{f}$ of the times of the events, $\mathbf{t} = (t)_{(i,j,t)\in\mathcal{E}}$, but otherwise conserves the individual snapshot graphs.

Definition A.7. *Snapshot shuffling* $\mathrm{P}[\mathbf{t}, \mathbf{f}(p_{\mathcal{T}}(\boldsymbol{\Gamma}))]$. A *snapshot shuffling* constrains the time of each event, i.e. $\mathbf{t} = (t)_{(i,j,t)\in\mathcal{E}}$. It randomizes each snapshot graph Γ^t individually in a manner that may be constrained by any function $\mathbf{f}$ of $p_{\mathcal{T}}(\boldsymbol{\Gamma})$.

2. Proofs of propositions in Section IV A

Proof of Proposition IV.1: Two MRRMs $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ are comparable and $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$ if and only if there exists a function $\mathbf{f}$ for which $\mathbf{y}(G) = \mathbf{f}(\mathbf{x}(G))$ for all states $G \in \mathcal{G}$.

Proof. If there exists a function $\mathbf{f}$ that allows to calculate $\mathbf{y}$ solely from $\mathbf{x}$ then all networks in a given $\mathbf{x}$-equivalence class $\mathcal{G}_{\mathbf{x}^*}$ (Def. II.11) correspond to the same value of $\mathbf{y}$. Thus $\mathcal{G}_{\mathbf{x}(G^*)} \subseteq \mathcal{G}_{\mathbf{y}(G^*)}$ for all $G^* \in \mathcal{G}$. Conversely, if the $\mathbf{x}$-equivalence class $\mathcal{G}_{\mathbf{x}(G^*)}$ is contained in the $\mathbf{y}$-equivalence class $\mathcal{G}_{\mathbf{y}(G^*)}$ for all $G^* \in \mathcal{G}$, it means that a unique value of $\mathbf{y}$ corresponds to each possible value of $\mathbf{x} \in \mathcal{X} = \{\mathbf{x}(G) : G \in \mathcal{G}\}$, thus defining a functional relation from $\mathcal{X}$ to $\mathcal{Y} = \{\mathbf{y}(G) : G \in \mathcal{G}\}$. □

Proof of Proposition IV.2: $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$ if and only if the partition $\mathcal{G}_{\mathbf{x}}$ is finer than $\mathcal{G}_{\mathbf{y}}$.

Proof. The propositions follows directly from the fact that the definition of comparability (Def. IV.1) is exactly the definition of the refinement relation for partitions [73], i.e. that for all $G \in \mathcal{G}$, and thus for all sets $\mathcal{G}_{\mathbf{x}(G)}$ and $\mathcal{G}_{\mathbf{y}(G)}$ in the partitions generated by $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$, respectively, we have $\mathcal{G}_{\mathbf{x}(G)} \subseteq \mathcal{G}_{\mathbf{y}(G)}$. □

3. Proofs of propositions and theorems in Section V A

Proof of Proposition V.1: If two MRRMs, $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$, are compatible then $\mathrm{P}[\mathbf{x} \circ \mathbf{y}] = \mathrm{P}[\mathbf{y} \circ \mathbf{x}]$.

Proof. This can be shown by direct calculation by noting that the transition matrices of MRRMs are symmetric. For any two compatible MRRMs, $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$, their compositions $\mathrm{P}[\mathbf{x} \circ \mathbf{y}]$ and $\mathrm{P}[\mathbf{y} \circ \mathbf{x}]$ both define a MRRM. So the associated transition matrices must be symmetric (Def. II.11), and

$$\mathbf{P}^{\mathbf{y}}\mathbf{P}^{\mathbf{x}} = (\mathbf{P}^{\mathbf{y}}\mathbf{P}^{\mathbf{x}})^{\intercal} = (\mathbf{P}^{\mathbf{x}})^{\intercal}(\mathbf{P}^{\mathbf{y}})^{\intercal} = \mathbf{P}^{\mathbf{x}}\mathbf{P}^{\mathbf{y}} \ . \tag{A8}$$

□

Proof of Proposition V.2: Consider two compatible MRRMs, $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$. Their composition, $\mathrm{P}[\mathbf{y} \circ \mathbf{x}]$, is coarser (or equal) than both $\mathrm{P}[\mathbf{y}]$ and $\mathrm{P}[\mathbf{x}]$, i.e. $\mathrm{P}[\mathbf{y} \circ \mathbf{x}] \geq \mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y} \circ \mathbf{x}] \geq \mathrm{P}[\mathbf{y}]$, even if $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ are not comparable.

Proof. Since $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ are compatible, $\mathrm{P}[\mathbf{y} \circ \mathbf{x}]$ is a MRRM by the definition of compatibility (Def. V.2). Since G is itself by definition in the set obtained by applying any MRRM to G, the target set of $\mathrm{P}[\mathbf{y} \circ \mathbf{x}]$ can never be smaller than the target set of $\mathrm{P}[\mathbf{x}]$. Thus, $\mathrm{P}[\mathbf{y} \circ \mathbf{x}] \geq \mathrm{P}[\mathbf{x}]$. By the same reasoning we obtain that $\mathrm{P}[\mathbf{x} \circ \mathbf{y}] \geq \mathrm{P}[\mathbf{y}]$, and since $\mathrm{P}[\mathbf{x} \circ \mathbf{y}] = \mathrm{P}[\mathbf{y} \circ \mathbf{x}]$ (Proposition V.1), that $\mathrm{P}[\mathbf{y} \circ \mathbf{x}] \geq \mathrm{P}[\mathbf{y}]$. □

Proof of Proposition V.3: Let $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ be two MRRMs and $\mathrm{P}[\mathbf{x}] \leq \mathrm{P}[\mathbf{y}]$. Then they are compatible and their composition gives $\mathrm{P}[\mathbf{y} \circ \mathbf{x}] = \mathrm{P}[\mathbf{y}]$.

Proof. Since $\mathbf{y}$ is a function of $\mathbf{x}$, its value is the same for all G which satisfy $\mathbf{x}(G) = \mathbf{x}^*$ and is equal to $\mathbf{y}^* = \mathbf{y}(G^*)$. This means that $P_{\mathbf{y}}(G|G') = P_{\mathbf{y}}(G|G^*)$ for all $G' \in \mathcal{G}_{\mathbf{x}^*}$, and Eq. (6) reduces to:

$$\begin{aligned} P_{\mathbf{y}\circ\mathbf{x}}(G|G^*) &= P_{\mathbf{y}}(G|G^*) \sum_{G'\in\mathcal{G}_{\mathbf{x}^*}} P_{\mathbf{x}}(G'|\mathbf{x}^*) \\ &= P_{\mathbf{y}}(G|G^*) \ , \end{aligned} \tag{A9}$$

where the second equality is obtained from the requirement that $P_{\mathbf{x}}$ must be normalized on $\mathcal{G}_{\mathbf{x}^*}$. So $\mathrm{P}[\mathbf{y} \circ \mathbf{x}] = \mathrm{P}[\mathbf{y}]$, which is a MRRM, showing that $\mathrm{P}[\mathbf{x}]$ and $\mathrm{P}[\mathbf{y}]$ are compatible. □

Proof of Proposition V.4: If $\mathbf{x}$ is conditionally independent of $\mathbf{y}$ given a common coarsening $\mathbf{z}$ then $\mathbf{y}$ is conditionally independent of $\mathbf{x}$ given $\mathbf{z}$

Proof. We note that since $\mathbf{z}$ is coarser than $\mathbf{x}$, it follows that $P_{\mathbf{y}|\mathbf{x}} = P_{\mathbf{y}|\mathbf{x},\mathbf{z}}$ (since $\mathcal{G}_{\mathbf{x}^*} \subseteq \mathcal{G}_{\mathbf{z}^*}$, conditioning only on $\mathbf{x}^*$ is equivalent to conditioning on both $\mathbf{x}^*$ and $\mathbf{z}^*$), and in the same manner that $P_{\mathbf{x}|\mathbf{y}} = P_{\mathbf{x}|\mathbf{y},\mathbf{z}}$. Thus, the symmetry of the conditional independence given a common

coarsening follows directly from the symmetry of the traditional conditional independence. For completeness we demonstrate the symmetry of conditional independence below.

Consider that $\mathbf{x}$ is independent of $\mathbf{y}$ conditioned on $\mathbf{z}$, i.e. $P_{\mathbf{x}|\mathbf{z}}(\mathbf{x}^\dagger|\mathbf{z}^*) = P_{\mathbf{x}|\mathbf{y},\mathbf{z}}(\mathbf{x}^\dagger|\mathbf{y}^\ddagger,\mathbf{z}^*)$. To show that this implies the symmetric relation, we use the definition of the conditional probability [Eq. (7)]:

$$\begin{aligned}\frac{P_{\mathbf{x}|\mathbf{y},\mathbf{z}}(\mathbf{x}^\dagger|\mathbf{y}^\ddagger,\mathbf{z}^*)}{P_{\mathbf{x}|\mathbf{z}}(\mathbf{x}^\dagger|\mathbf{z}^*)} &= \frac{\Omega_{(\mathbf{x}^\dagger,\mathbf{y}^\ddagger,\mathbf{z}^*)}}{\Omega_{(\mathbf{y}^\ddagger,\mathbf{z}^*)}}\frac{\Omega_{\mathbf{z}^*}}{\Omega_{(\mathbf{x}^\dagger,\mathbf{z}^*)}} \\ &= \frac{\Omega_{(\mathbf{x}^\dagger,\mathbf{y}^\ddagger,\mathbf{z}^*)}}{\Omega_{(\mathbf{x}^\dagger,\mathbf{z}^*)}}\frac{\Omega_{\mathbf{z}^*}}{\Omega_{(\mathbf{y}^\ddagger,\mathbf{z}^*)}} \\ &= \frac{P_{\mathbf{y}|\mathbf{x},\mathbf{z}}(\mathbf{y}^\ddagger|\mathbf{x}^\dagger,\mathbf{z}^*)}{P_{\mathbf{y}|\mathbf{z}}(\mathbf{y}^\ddagger|\mathbf{z}^*)} \ . \end{aligned} \tag{A10}$$

This relation (which must be true) is only satisfied if $P_{\mathbf{y}|\mathbf{x},\mathbf{z}}(\mathbf{y}^\ddagger|\mathbf{x}^\dagger,\mathbf{z}^*) = P_{\mathbf{y}|\mathbf{z}}(\mathbf{y}^\ddagger|\mathbf{z}^*)$, i.e. if $\mathbf{y}$ is independent of $\mathbf{x}$ conditioned on $\mathbf{z}$, thus completing the proof. □

Proof of Theorem 1: P[$\mathbf{x}$] and P[$\mathbf{y}$] are compatible if and only if they are conditionally independent given the common coarsening $\mathbf{z} = \mathbf{x} \circ \mathbf{y}$.

Proof. To show that conditional independence given a common coarsening is equivalent to compatibility, we first show that the former implies the latter and then that the latter implies the former. To avoid clutter, we will in the following use the notation $\mathbf{y}'$ as short for $\mathbf{y}(G')$, as well as $\mathbf{y}''$ for $\mathbf{y}(G'')$ and $\mathbf{x}''$ for $\mathbf{x}(G'')$.

Conditional independence given a common coarsening implies compatibility. We will use the law of total probability and Proposition V.3 to show this. We first show that a law of total probability applies to the probabilities of features. Taking the probability distribution defining the composition of P[$\mathbf{y}$] on P[$\mathbf{x}$] [Eq. (6)] and multiplying by the term $\sum_{\mathbf{y}^\dagger}\delta_{\mathbf{y}^\dagger,\mathbf{y}'} = 1$, we get:

$$\begin{aligned}P_{\mathbf{y}\circ\mathbf{x}}(G|\mathbf{x}^*) &= \sum_{G'\in\mathcal{G}}\sum_{\mathbf{y}^\dagger}\delta_{\mathbf{y}^\dagger,\mathbf{y}'}\frac{\delta_{\mathbf{y}(G),\mathbf{y}'}}{\Omega_{\mathbf{y}}(G)}P_{\mathbf{x}}(G'|\mathbf{x}^*) \\ &= \sum_{\mathbf{y}^\dagger}\frac{\delta_{\mathbf{y}(G),\mathbf{y}^\dagger}}{\Omega_{\mathbf{y}}(G)}\sum_{G'\in\mathcal{G}}\delta_{\mathbf{y}^\dagger,\mathbf{y}'}P_{\mathbf{x}}(G'|\mathbf{x}^*) \\ &= \sum_{\mathbf{y}^\dagger}P_{\mathbf{y}}(G|\mathbf{y}^\dagger)P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^\dagger|\mathbf{x}^*) \ . \end{aligned} \tag{A11}$$

To obtain the second equality above we used the property of Kronecker delta functions that $\delta_{a,b}\delta_{b,c} = \delta_{a,c}\delta_{b,c}$, and the last equality was obtained from the definitions of $P_{\mathbf{y}}(G|\mathbf{y}^\dagger)$ (Def. II.8) and $P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^\dagger|\mathbf{x}^*)$ (Def. V.3). Using the law of total probability [Eq. (A11)], we now expand $P_{\mathbf{y}\circ\mathbf{x}}(G|\mathbf{x}^*)$ to get:

$$\begin{aligned}P_{\mathbf{y}\circ\mathbf{x}}(G|\mathbf{x}^*) &= \sum_{\mathbf{y}^\dagger}P_{\mathbf{y}}(G|\mathbf{y}^\dagger)P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^\dagger|\mathbf{x}^*) \\ &= \sum_{\mathbf{y}^\dagger}P_{\mathbf{y}}(G|\mathbf{y}^\dagger)P_{\mathbf{y}|\mathbf{z}}(\mathbf{y}^\dagger|\mathbf{z}^*) \\ &= P_{\mathbf{y}\circ\mathbf{z}}(G|\mathbf{z}^*) \\ &= P_{\mathbf{z}}(G|\mathbf{z}^*) \ . \end{aligned} \tag{A12}$$

Here, the second equality follows from the conditional independence of $\mathbf{x}$ and $\mathbf{y}$ (Def. V.4), the second-to-last equality follows from the law of total probability, and the last from Proposition V.3 since $\mathbf{z} \geq \mathbf{y}$.

Compatibility implies conditional independence given a common coarsening. Because $\mathbf{x}$ and $\mathbf{y}$ are compatible their composition is a MRRM and we can choose $\mathbf{z} = \mathbf{x} \circ \mathbf{y}$, which by construction is a common coarsening of $\mathbf{x}$ and $\mathbf{y}$. The conditional independence of $\mathbf{x}$ and $\mathbf{y}$ given this $\mathbf{z}$ can now be shown from its definition via a direct calculation:

$$\begin{aligned}P_{\mathbf{y}|\mathbf{x}\circ\mathbf{y}}(\mathbf{y}^\dagger|\mathbf{x}\circ\mathbf{y}(G^*)) &= \sum_{G'\in\mathcal{G}}\delta_{\mathbf{y}^\dagger,\mathbf{y}'}P_{\mathbf{x}\circ\mathbf{y}}(G'|\mathbf{x}\circ\mathbf{y}(G^*)) \\ &= \sum_{G'\in\mathcal{G}}\sum_{G''\in\mathcal{G}}\frac{\delta_{\mathbf{x}'',\mathbf{x}^*}\delta_{\mathbf{y}^\dagger,\mathbf{y}'}\delta_{\mathbf{y}',\mathbf{y}''}}{\Omega_{\mathbf{y}''}\Omega_{\mathbf{x}^*}} \\ &= \sum_{G''\in\mathcal{G}}\frac{\delta_{\mathbf{x}'',\mathbf{x}^*}\delta_{\mathbf{y}^\dagger,\mathbf{y}''}}{\Omega_{\mathbf{y}''}\Omega_{\mathbf{x}^*}}\sum_{G'\in\mathcal{G}}\delta_{\mathbf{y}',\mathbf{y}''} \\ &= \sum_{G''\in\mathcal{G}}\frac{\delta_{\mathbf{x}'',\mathbf{x}^*}\delta_{\mathbf{y}^\dagger,\mathbf{y}''}}{\Omega_{\mathbf{x}^*}} \\ &= \sum_{G''\in\mathcal{G}}\delta_{\mathbf{y}^\dagger,\mathbf{y}''}P_{\mathbf{x}}(G''|\mathbf{x}^*) \\ &= P_{\mathbf{y}|\mathbf{x}}(\mathbf{y}^\dagger|\mathbf{x}^*) \ . \end{aligned} \tag{A13}$$

In the third equality we again used the property of Kronecker delta functions $\delta_{a,b}\delta_{b,c} = \delta_{a,c}\delta_{b,c}$, and in the fourth equality we used the definition of the partition function $\Omega_{\mathbf{y}''}$. □

Proof of Theorem 2: Consider two compatible MRRMs P[$\mathbf{x}$] and P[$\mathbf{y}$], and any adapted refinements of these, P[$\mathbf{y}, \mathbf{f}(\mathbf{x})$] and P[$\mathbf{x}, \mathbf{g}(\mathbf{y})$]. Then P[$\mathbf{y}, \mathbf{f}(\mathbf{x})$] and P[$\mathbf{x}, \mathbf{g}(\mathbf{y})$] are compatible, and their composition is given by P[$\mathbf{x} \circ \mathbf{y}, \mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{y})$].

Proof. To prove Theorem 2, it is sufficient to prove that $(\mathbf{y}, \mathbf{f}(\mathbf{x}))$ and $\mathbf{x}$ are independent conditioned on their common coarsening $(\mathbf{y} \circ \mathbf{x}, \mathbf{f}(\mathbf{x}))$. From this it then immediately follows that $(\mathbf{y}, \mathbf{f}(\mathbf{x}))$ and $(\mathbf{x}, \mathbf{g}(\mathbf{y}))$ are independent conditioned on their common coarsening $(\mathbf{y} \circ \mathbf{x}, \mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{y}))$ and thus that P[$\mathbf{y}$,$\mathbf{f}(\mathbf{x})$] and P[$\mathbf{x}$,$\mathbf{g}(\mathbf{y})$] are compatible with P[$(\mathbf{y}, \mathbf{f}(\mathbf{x})) \circ (\mathbf{x}, \mathbf{g}(\mathbf{y}))$] = P[$\mathbf{y}\circ\mathbf{x}, \mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{y})$]. To develop the proof, we consider the conditional probability of $\mathbf{x}$ given $(\mathbf{y}, \mathbf{f}(\mathbf{x}))$:

$$P_{\mathbf{x}|\mathbf{y},\mathbf{f}(\mathbf{x})}(\mathbf{x}^\dagger|\mathbf{y}^*,\mathbf{f}(\mathbf{x}^*)) = \frac{\Omega_{(\mathbf{x}^\dagger,\mathbf{y}^*,\mathbf{f}(\mathbf{x}^*))}}{\Omega_{(\mathbf{y}^*,\mathbf{f}(\mathbf{x}^*))}} \ . \tag{A14}$$

Since $\mathbf{x} \leq \mathbf{f}(\mathbf{x})$, we have $\Omega_{(\mathbf{x}^\dagger,\mathbf{y}^*,\mathbf{f}(\mathbf{x}^*))} = \Omega_{(\mathbf{x}^\dagger,\mathbf{y}^*)}$ whenever $\mathbf{f}(\mathbf{x}^\dagger) = \mathbf{f}(\mathbf{x}^*)$. We use this to rewrite the equation above, multiplying by the factor $\Omega_{\mathbf{y}^*}/\Omega_{\mathbf{y}^*} = 1$, along the way,

$$\begin{aligned} P_{\mathbf{x}|\mathbf{y},\mathbf{f}(\mathbf{x})}(\mathbf{x}^\dagger|\mathbf{y}^*,\mathbf{f}(\mathbf{x}^*)) &= \frac{\Omega_{(\mathbf{x}^\dagger,\mathbf{y}^*)}\delta_{\mathbf{f}(\mathbf{x}^\dagger),\mathbf{f}(\mathbf{x}^*)}}{\Omega_{\mathbf{y}^*}} \frac{\Omega_{\mathbf{y}^*}}{\Omega_{(\mathbf{y}^*,\mathbf{f}(\mathbf{x}^*))}} \\ &= \frac{P_{\mathbf{x}|\mathbf{y}}(\mathbf{x}^\dagger|\mathbf{y}^*)\delta_{\mathbf{f}(\mathbf{x}^\dagger),\mathbf{f}(\mathbf{x}^*)}}{P_{\mathbf{f}(\mathbf{x})|\mathbf{y}}(\mathbf{f}(\mathbf{x}^*)|\mathbf{y}^*)} \quad . \text{(A15)} \end{aligned}$$

Now, since $\mathbf{x}$ and $\mathbf{y}$ are compatible, we have that $P_{\mathbf{x}|\mathbf{y}}(\mathbf{x}^\dagger|\mathbf{y}^*) = P_{\mathbf{x}|\mathbf{z}}(\mathbf{x}^\dagger|\mathbf{z}^*)$, with $\mathbf{z} = \mathbf{y} \circ \mathbf{x}$. Furthermore, it also means that $P_{\mathbf{f}(\mathbf{x})|\mathbf{y}}(\mathbf{f}(\mathbf{x}^*)|\mathbf{y}^*) = P_{\mathbf{f}(\mathbf{x})|\mathbf{z}}(\mathbf{f}(\mathbf{x}^*)|\mathbf{z}^*)$. The following calculation shows this:

$$\begin{aligned} P_{\mathbf{f}(\mathbf{x})|\mathbf{y}}(\mathbf{f}(\mathbf{x}^*)|\mathbf{y}^\dagger) &= \sum_{\mathbf{x}^\ddagger} P_{\mathbf{f}(\mathbf{x})|\mathbf{x}}(\mathbf{f}(\mathbf{x}^*)|\mathbf{x}^\ddagger) P_{\mathbf{x}|\mathbf{y}}(\mathbf{x}^\ddagger|\mathbf{y}^\dagger) \\ &= \sum_{\mathbf{x}^\ddagger} P_{\mathbf{f}(\mathbf{x})|\mathbf{x}}(\mathbf{f}(\mathbf{x}^*)|\mathbf{x}^\ddagger) P_{\mathbf{x}|\mathbf{z}}(\mathbf{x}^\ddagger|\mathbf{z}^\dagger) \\ &= P_{\mathbf{f}(\mathbf{x})|\mathbf{z}}(\mathbf{f}(\mathbf{x}^*)|\mathbf{z}^\dagger) \quad . \end{aligned} \tag{A16}$$

(Note that we do not necessarily have $\mathbf{f}(\mathbf{x}) \leq \mathbf{z}$, though.) Plugging these two identities into Eq. (A15) gives:

$$\begin{aligned} P_{\mathbf{x}|\mathbf{y},\mathbf{f}(\mathbf{x})}(\mathbf{x}^\dagger|\mathbf{y}^*,\mathbf{f}(\mathbf{x}^*)) &= \frac{P_{\mathbf{x}|\mathbf{z}}(\mathbf{x}^\dagger|\mathbf{z}^*)\delta_{\mathbf{f}(\mathbf{x}^\dagger),\mathbf{f}(\mathbf{x}^*)}}{P_{\mathbf{f}(\mathbf{x})|\mathbf{z}}(\mathbf{f}(\mathbf{x}^*)|\mathbf{z}^*)} \\ &= \frac{\Omega_{(\mathbf{x}^\dagger,\mathbf{z}^*)}\delta_{\mathbf{f}(\mathbf{x}^\dagger),\mathbf{f}(\mathbf{x}^*)}}{\Omega_{\mathbf{z}^*}} \frac{\Omega_{\mathbf{z}^*}}{\Omega_{(\mathbf{z}^*,\mathbf{f}(\mathbf{x}^*))}} \\ &= \frac{\Omega_{(\mathbf{x}^\dagger,\mathbf{z}^*,\mathbf{f}(\mathbf{x}^*))}}{\Omega_{(\mathbf{z}^*,\mathbf{f}(\mathbf{x}^*))}} \\ &= P_{\mathbf{x}|\mathbf{z},\mathbf{f}(\mathbf{x})}(\mathbf{x}^\dagger|\mathbf{z}^*,\mathbf{f}(\mathbf{x}^*)) \quad . \end{aligned} \tag{A17}$$

Since both $\mathbf{z}^* \geq \mathbf{x}$ and $\mathbf{f}(\mathbf{x}^*) \geq \mathbf{x}$, we have $(\mathbf{z},\mathbf{f}(\mathbf{x})) \geq \mathbf{x}$, which together with Eq. (A17) shows that $\mathbf{x}$ is independent of $(\mathbf{y},\mathbf{f}(\mathbf{x}))$ conditionally on their common coarsening $(\mathbf{z},\mathbf{f}(\mathbf{x}))$, thus completing the proof. □

Proof of Proposition V.5: Any link shuffling $\mathrm{P}[\mathbf{f}(\mathcal{L}),\boldsymbol{\Theta}]$ and timeline shuffling $\mathrm{P}[\mathcal{L},\mathbf{g}(\boldsymbol{\Theta})]$ are compatible and their composition is given by $\mathrm{P}[L,\mathbf{f}(\mathcal{L}),\mathbf{g}(\boldsymbol{\Theta})]$.

Proof. It is clear that the content of the individual timelines $\Theta_{(i,j)} \in p_\mathcal{L}(\boldsymbol{\Theta})$ does not in any way constrain what values $\mathcal{L}$ may take, only their number L does. Furthermore, the number of ways that we can distribute the L timelines on the links is independent of the particular configuration of $\mathcal{L}$, so $\Omega_{\mathcal{L}',p_\mathcal{L}(\boldsymbol{\Theta}^*)} = \Omega_{\mathcal{L}'',p_\mathcal{L}(\boldsymbol{\Theta}^*)}$ for all $\mathcal{L}'$, $\mathcal{L}''$. Similarly, the way we can distribute the E instantaneous events on the timelines depends only on $\mathcal{L}$ through L, so also $\Omega_{\mathcal{L}',L^*} = \Omega_{\mathcal{L}'',L^*}$ for all $\mathcal{L}'$, $\mathcal{L}''$. This means that $\Omega_{\mathcal{L}',L^*} \propto \Omega_{\mathcal{L}',p_\mathcal{L}(\boldsymbol{\Theta}^*)}$ for all L', and since the conditional probabilities must be normed, that $P_{\mathcal{L}|p_\mathcal{L}(\boldsymbol{\Theta})}(\mathcal{L}^\dagger|p_\mathcal{L}(\boldsymbol{\Theta}^*)) = P_{\mathcal{L}|L}(\mathcal{L}^\dagger|L^*)$, i.e. that $\mathcal{L}$ and $p_\mathcal{L}(\boldsymbol{\Theta})$ are independent conditioned on L. Since $(L,E) \geq (\mathcal{L},E)$ and $(L,E) \geq p_\mathcal{L}(\boldsymbol{\Theta})$, it then follows from Theorem 1 that $(\mathcal{L},E)$ and $p_\mathcal{L}(\boldsymbol{\Theta})$ are compatible. This shows that the coarsest link shuffling, $\mathrm{P}[\boldsymbol{\Theta}]$ (equivalent to $\mathrm{P}[p_\mathcal{L}(\boldsymbol{\Theta})]$), and the coarsest timeline shuffling, $\mathrm{P}[\mathcal{L},E]$, are compatible. We next note that any link and timeline shufflings are adapted refinements of $\mathrm{P}[\boldsymbol{\Theta}]$ and $\mathrm{P}[\mathcal{L},E]$, respectively (compare Defs. A.4 and A.5 with Def. V.5). So applying Theorem 2 gives that any link shuffling $\mathrm{P}[\mathbf{f}(\mathcal{L}),\boldsymbol{\Theta}]$ and any timeline shuffling $\mathrm{P}[\mathcal{L},\mathbf{g}(\boldsymbol{\Theta})]$ are compatible and that their composition is $\mathrm{P}[L,\mathbf{f}(\mathcal{L}),\mathbf{g}(\boldsymbol{\Theta})]$. □

Proof of Proposition V.6: Any sequence shuffling $\mathrm{P}[\mathbf{f}(\mathbf{t}),p_\mathcal{T}(\boldsymbol{\Gamma})]$ and snapshot shuffling $\mathrm{P}[\mathbf{t},\mathbf{g}(p_\mathcal{T}(\boldsymbol{\Gamma}))]$ are compatible and their composition is given by $\mathrm{P}[p(\mathbf{A}),\mathbf{f}(\mathbf{t}),\mathbf{g}(p_\mathcal{T}(\boldsymbol{\Gamma}))]$.

Proof. Following the same reasoning as in the proof of Proposition V.5, we note that $\Omega_{\mathbf{t}',p_\mathcal{T}(\boldsymbol{\Gamma}^*)} = \Omega_{\mathbf{t}',p(\mathbf{A}^*)}$ for all $\mathbf{t}'$ and that $p(\mathbf{A})$ satisfies $p(\mathbf{A}) \geq \mathbf{t}$ and $p(\mathbf{A}) \geq p_\mathcal{T}(\boldsymbol{\Gamma})$. Thus, $\mathbf{t}$ is independent of $p_\mathcal{T}(\boldsymbol{\Gamma})$ conditioned on $p(\mathbf{A})$. So the coarsest sequence and snapshot shufflings, $\mathrm{P}[p_\mathcal{T}(\boldsymbol{\Gamma})]$ and $\mathrm{P}[\mathbf{t}]$ are compatible by Theorem 1. Consequently, since all sequence shufflings $\mathrm{P}[\mathbf{f}(\mathbf{t}),p_\mathcal{T}(\boldsymbol{\Gamma})]$ and snapshot shufflings $\mathrm{P}[\mathbf{A},\mathbf{f}(p_\mathcal{T}(\boldsymbol{\Gamma}))]$ are adapted refinements of $\mathrm{P}[p_\mathcal{T}(\boldsymbol{\Gamma})]$ or $\mathrm{P}[\mathbf{t}]$, respectively (Def. V.5), Theorem 2 tells us that they are compatible and that their composition is $\mathrm{P}[p(\mathbf{A}),\mathbf{f}(\mathbf{t}),\mathbf{g}(p_\mathcal{T}(\boldsymbol{\Gamma}))]$. □

[1] Erdos, P. & Rényi, A. On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5**, 17–61 (1960).

[2] Newman, M. E. The structure and function of complex networks. *SIAM Review* **45**, 167–256 (2003).

[3] Fosdick, B. K., Larremore, D. B., Nishimura, J. & Ugander, J. Configuring random graph models with fixed degree sequences. *SIAM Review* **60**, 315–355 (2018).

[4] Kovanen, L., Karsai, M., Kaski, K., Kertész, J. & Saramäki, J. Temporal motifs in time-dependent networks. *J. Stat. Mech. Theory Exp* **2011**, P11005 (2011).

[5] Karsai, M., Kaski, K., Barabási, A.-L. & Kertész, J. Universal features of correlated bursty behaviour. *Sci. Rep.* **2**, 397 (2012).

[6] Karsai, M., Kaski, K. & Kertész, J. Correlated dynamics in egocentric communication networks. *PLoS One* **7**, e40612 (2012).

[7] Kovanen, L., Kaski, K., Kertész, J. & Saramäki, J. Temporal motifs reveal homophily, gender-specific patterns, and group talk in call sequences. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18070–5 (2013).

[8] Orsini, C. *et al.* Quantifying randomness in real networks. *Nat. Commun.* **6**, 8627 (2015).

[9] Karsai, M. *et al.* Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E* **83**, 025102(R) (2011).

[10] Rocha, L. E. C., Liljeros, F. & Holme, P. Simulated Epidemics in an Empirical Spatiotemporal Network of

50,185 Sexual Contacts. *PLoS Comp. Biol.* **7**, e1001109 (2011).

[11] Miritello, G., Moro, E. & Lara, R. Dynamical strength of social ties in information spreading. *Phys. Rev. E* **83**, 045102(R) (2011).

[12] Starnini, M., Baronchelli, A., Barrat, A. & Pastor-Satorras, R. Random walks on temporal networks. *Phys. Rev. E* **85**, 056115 (2012).

[13] Kivelä, M. *et al.* Multiscale analysis of spreading in a large communication network. *J. Stat. Mech. Theory Exp* **2012**, P03005 (2012).

[14] Gauvin, L., Panisson, A., Cattuto, C. & Barrat, A. Activity clocks: spreading dynamics on temporal networks of human contact. *Sci. Rep.* **3**, 3099 (2013).

[15] Karimi, F. & Holme, P. Threshold model of cascades in empirical temporal networks. *Physica A* **392**, 3476–3483 (2013).

[16] Takaguchi, T., Masuda, N. & Holme, P. Bursty Communication Patterns Facilitate Spreading in a Threshold-Based Epidemic Dynamics. *PLoS One* **8** (2013).

[17] Holme, P. & Liljeros, F. Birth and death of links control disease spreading in empirical contact networks. *Sci. Rep.* **4**, 4999 (2014).

[18] Karsai, M., Perra, N. & Vespignani, A. Time varying networks and the weakness of strong ties. *Sci. Rep.* **4**, 4001 (2014).

[19] Backlund, V.-P., Saramäki, J. & Pan, R. K. Effects of temporal correlations on cascades: Threshold models on temporal networks. *Phys. Rev. E* **89**, 062815 (2014).

[20] Cardillo, A. *et al.* Evolutionary dynamics of time-resolved social interactions. *Phys. Rev. E* **90**, 052825 (2014).

[21] Thomas, B., Jurdak, R., Zhao, K. & Atkinson, I. Diffusion in colocation contact networks: the impact of nodal spatiotemporal dynamics. *PLoS One* **11**, e0152624 (2015).

[22] Delvenne, J.-C., Lambiotte, R. & Rocha, L. E. C. Diffusion on networked systems is a question of time or structure. *Nat. Commun.* **6**, 7366 (2015).

[23] Saramäki, J. & Holme, P. Exploring temporal networks with greedy walks. *Eur. Phys. J. B* **88**, 1–8 (2015).

[24] Génois, M., Vestergaard, C. L., Cattuto, C. & Barrat, A. Compensating for population sampling in simulations of epidemic spread on temporal contact networks. *Nat. Commun.* **6**, 8860 (2015).

[25] Valdano, E., Poletto, C. & Colizza, V. Infection propagator approach to compute epidemic thresholds on temporal networks: impact of immunity and of limited temporal resolution. *Eur. Phys. J. B* **88**, 341 (2015).

[26] Holme, P. Temporal network structures controlling disease spreading. *Phys. Rev. E* **94**, 022305 (2016).

[27] Barrat, A., Barthélemy, M. & Vespignani, A. *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).

[28] Gleeson, J. P. High-accuracy approximation of binary-state dynamics on networks. *Phys. Rev. Lett.* **107**, 068701 (2011).

[29] Pastor-Satorras, R. & Vespignani, A. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200 (2001).

[30] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A. L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).

[31] Solé, R. V. & Montoya, J. M. Complexity and fragility in ecological networks. *Proc. R. Soc. Lond. B* **268**, 2039–2045 (2001).

[32] Holme, P., Kim, B. J., Yoon, C. N. & Han, S. K. Attack vulnerability of complex networks. *Phys. Rev. E* **65**, 056109 (2002).

[33] Bollobás, B. & Riordan, O. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics* **1**, 1–35 (2004).

[34] Albert, R., Jeong, H. & Barabási, A.-L. Error and attack tolerance of complex networks. *Nature* **406**, 378 (2000).

[35] Cohen, R., Erez, K., Ben-Avraham, D. & Havlin, S. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.* **85**, 4626 (2000).

[36] Newman, M. E. J. & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).

[37] Karrer, B. & Newman, M. E. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **83**, 016107 (2011).

[38] Maslov, S. & Sneppen, K. Specificity and Stability in Topology of Protein Networks. *Science* **296**, 910–913 (2002).

[39] Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).

[40] Watts, D. J. & Strogatz, S. H. Collective dynamics of "small-world" networks. *Nature* **393**, 440–442 (1998).

[41] Onnela, J.-P. *et al.* Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332–7336 (2007).

[42] Bajardi, P., Barrat, A., Natale, F., Savini, L. & Colizza, V. Dynamical patterns of cattle trade movements. *PLoS One* **6**, e19869 (2011).

[43] Squartini, T., Fagiolo, G. & Garlaschelli, D. Randomizing world trade. i. a binary network analysis. *Phys. Rev. E* **84**, 046117 (2011).

[44] Squartini, T., Fagiolo, G. & Garlaschelli, D. Randomizing world trade. ii. a weighted network analysis. *Phys. Rev. E* **84**, 046118 (2011).

[45] Holme, P. & Saramäki, J. Temporal networks. *Phys. Rep.* **519**, 97–125 (2012). 1108.1780.

[46] Jurgens, D. & Lu, T.-C. Temporal Motifs Reveal the Dynamics of Editor Interactions in Wikipedia. In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, 1, 162–169 (2012).

[47] Holme, P. Modern temporal network theory: a colloquium. *Eur. Phys. J. B* **88**, 1–30 (2015).

[48] Saracco, F., Clemente, R. D., Gabrielli, A. & Squartini, T. Detecting early signs of the 2007-2008 crisis in the world trade. *Sci. Rep.* **30286** (2016).

[49] Zhang, Y., Garas, A. & Scholtes, I. Controllability of temporal networks: An analysis using higher-order networks. *arXiv preprint arXiv:1701.06331* (2017).

[50] Pósfai, M. & Hövel, P. Structural controllability of temporal networks. *New J. Phys.* **16**, 123055 (2014).

[51] Takaguchi, T., Sato, N., Yano, K. & Masuda, N. Importance of individual events in temporal networks. *New J. Phys.* **14**, 093003 (2012).

[52] Holme, P. Network reachability of real-world contact sequences. *Phys. Rev. E* **71**, 046119 (2005).

[53] Takaguchi, T., Sato, N., Yano, K. & Masuda, N. Inferring directed static networks of influence from undirected temporal networks. In *Proceedings of IEEE 37th Annual Computer Software and Applications Confer-*

ence, 155–156 (Ieee, Kyoto, 2013).

[54] Takaguchi, T., Yano, Y. & Yoshida, Y. Coverage centralities for temporal networks. *Eur. Phys. J. B* **89**, 35 (2016).

[55] Masuda, N. & Lambiotte, R. *A Guide to Temporal Networks* (World Scientific, 2016).

[56] Li, A., Cornelius, S. P., Liu, Y.-Y., Wang, L. & Barabási, A.-L. The fundamental advantages of temporal networks. *Science* **358**, 1042–1046 (2017).

[57] Tang, J., Scellato, S., Musolesi, M., Mascolo, C. & Latora, V. Small-world behavior in time-varying graphs. *Phys. Rev. E* **81**, 055101 (2010).

[58] Valencia, M., Martinerie, J., Dupont, S. & Chavez, M. Dynamic small-world behavior in functional brain networks unveiled by an event-related networks approach. *Phys. Rev. E* **77**, 050905 (2008).

[59] Sun, Y., Collinson, S. L., Suckling, J. & Sim, K. Dynamic reorganization of functional connectivity reveals abnormal temporal efficiency in schizophrenia. *Schizophrenia Bulletin* **45**, 659–669 (2019).

[60] Pan, R. K. & Saramäki, J. Path lengths, correlations, and centrality in temporal networks. *Phys. Rev. E* **81**, 016105 (2011).

[61] Alessandretti, L., Sapiezynski, P., Lehmann, S. & Baronchelli, A. Evidence for a conserved quantity in human mobility. *Nat. Hum. Behav.* **2**, 485–491 (2018).

[62] Redmond, U. & Cunningham, P. Identifying over-represented temporal processes in complex networks. In *Proceedings of the 2nd Workshop on Dynamic Networks and Knowledge Discovery co-located with ECML PKDD*, vol. 1229, 61–72 (2014).

[63] Sun, K., Baronchelli, A. & Perra, N. Contrasting effects of strong ties on SIR and SIS processes in temporal networks. *Eur. Phys. J. B* **88**, 326 (2015).

[64] Holme, P. & Saramäki, J. A map of approaches to temporal networks. In *Temporal Network Theory*, 1–24 (Springer, 2019).

[65] Jaynes, E. T. Information theory and statistical mechanics. *Physical Review* **106**, 620–630 (1957).

[66] Presse, S., Ghosh, K., Lee, J. & Dill, K. A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **85**, 1115–1141 (2013).

[67] Good, P. *Permutation, Parametric and Bootstrap Tests of Hypotheses (3rd ed.),* (Springer, 2005).

[68] Katz, L. & Powell, J. H. Probability distributions of random variables associated with a structure of the sample space of sociometric investigations. *Ann. Math. Stat.* **28**, 442 (1957).

[69] Snijders, T. Enumeration and simulation methods for 0–1 matrices with given marginals. *Psychometrika* **56**, 397–417 (1991).

[70] Latapy, M., Viard, T. & Magnien, C. Stream graphs and link streams for the modeling of interactions over time. *arXiv:1710.04073* (2017).

[71] Fournet, J. & Barrat, A. Contact patterns among high school students. *PLoS One* **9**, e107878 (2014).

[72] Stopczynski, A. *et al.* Measuring large-scale social networks with high resolution. *PLoS One* **9**, e95978 (2014).

[73] Hrbacek, K. & Jech, T. *Introduction to Set Theory, Revised and Expanded* (Crc Press, 1999).

[74] Pastor-Satorras, R., Castellano, C., Mieghem, P. V. & Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925 (2015).

[75] Vestergaard, C. L. & Génois, M. Temporal gillespie algorithm: Fast simulation of contagion processes on time-varying networks. *PLoS Comp. Biol.* **11**, e1004579 (2015).

[76] Alon, U. Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* **8**, 450–461 (2007).

[77] Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of escherichia coli. *Nat. Genet.* **31**, 64 (2002).

[78] Beber, M. E. *et al.* Artefacts in statistical analyses of network motifs: general framework and application to metabolic networks. *Journal of The Royal Society Interface* **9**, 3426–3435 (2012).

[79] Granovetter, M. S. The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).

[80] Perotti, J. I., Jo, H.-H., Holme, P. & Saramäki, J. Temporal network sparsity and the slowing down of spreading. *arXiv:1411.5553* (2014).

[81] Watts, D. J. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5766–5771 (2002).

[82] Goh, K. I. & Barabási, A.-L. Burstiness and memory in complex systems. *EPL* **81**, 48002 (2008).

[83] Nowak, M. A. *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, 2006).

[84] Perra, N., Gonçalves, B., Pastor-Satorras, R. & Vespignani, A. Activity driven modeling of time varying networks. *Sci. Rep.* **2**, 469 (2012).

[85] Kikas, R., Dumas, M. & Karsai, M. Bursty egocentric network evolution in Skype. *SNAM* **3**, 1393–1401 (2013).

[86] Karsai, M., Iñiguez, G., Kikas, R., Kaski, K. & Kertész, J. Local cascades induced global contagion: How heterogeneous thresholds, exogenous effects, and unconcerned behaviour govern online adoption spreading. *Sci. Rep.* **6** (2016).

[87] Kirk, D. E. *Optimal Control Theory: An Introduction* (Prentice-Hall, Inc, 1971).

[88] Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).

[89] Stanley, R. P. Enumerative combinatorics (Vol. 1, 2nd ed.). *Cambridge studies in advanced mathematics* (2011).

[90] Berend, D. & Tassa, T. Improved bounds on bell numbers and on moments of sums of random variables. *Probability and Mathematical Statistics* **30**, 185–205 (2010).

[91] Galimberti, E., Barrat, A., Bonchi, F., Cattuto, C. & Gullo, F. Mining (maximal) span-cores from temporal networks. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 107–116 (2018).

[92] Holland, P. W., Laskey, K. B. & Leinhardt, S. Stochastic blockmodels: First steps. *Soc. networks* **5**, 109–137 (1983).

[93] Zhang, X., Moore, C. & Newman, M. E. Random graph models for dynamic networks. *Eur. Phys. J. B* **90**, 200 (2017).

[94] Squartini, T., Mastrandrea, R. & Garlaschelli, D. Unbiased sampling of network ensembles. *New J. Phys.* **17**, 023052 (2015).

[95] Squartini, T. & Garlaschelli, D. Analytical maximum-likelihood method to detect patterns in real networks. *New J. Phys.* **13**, 083001 (2011).

[96] Pfitzner, R., Scholtes, I., Garas, A., Tessone, C. J. & Schweitzer, F. Betweenness preference: Quantify-

ing correlations in the topological dynamics of temporal networks. *Phys. Rev. Lett.* **110**, 198701 (2013). 1208.0588.

[97] Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D. & Lambiotte, R. Memory in network flows and its effects on spreading dynamics and community detection. *Nat. Commun.* **5**, 4630 (2014).

[98] Scholtes, I. *et al.* Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nat. Commun.* **5**, 5024 (2014).

[99] Peixoto, T. P. Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Phys. Rev. E* **92**, 042807 (2015).

[100] Peixoto, T. P. & Rosvall, M. Modeling sequences and temporal networks with dynamic community structures. *Nat. Commun.* **8**, 582 (2017).

[101] Peixoto, T. P. & Gauvin, L. Change points, memory and epidemic spreading in temporal networks. *Sci. Rep* **8**, 15511 (2018).

[102] Casiraghi, G., Nanumyan, V., Scholtes, I. & Schweitzer, F. Generalized hypergeometric ensembles: Statistical hypothesis testing in complex networks. *arXiv:1607.02441* (2016).

[103] Miritello, G., Lara, R., Cebrian, M. & Moro, E. Limited communication capacity unveils strategies for human interaction. *Sci. Rep.* **3**, 1950 (2013).

[104] Wu, Y., Zhou, C., Xiao, J., Kurths, J. & Schellnhuber, H. J. Evidence for a bimodal distribution in human communication. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18803–18808 (2010).

[105] Kivelä, M. *et al.* Multilayer networks. *J. Complex Netw.* **2**, 203–271 (2014).

[106] Jerrum, M. R., Valiant, L. G. & Vazirani, V. V. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.* **43**, 169–188 (1986).

[107] Stehlé, J. *et al.* High-resolution measurements of face-to-face contact patterns in a primary school. *PloS One* **6**, e23176 (2011).

[108] Gemmetto, V., Barrat, A. & Cattuto, C. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect. Dis.* **14**, 695 (2014).

[109] Lin, J. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **37**, 145–151 (1991).

Supplementary material

Supplementary TABLE S1: **Sequences, distributions, and moments of features.** Below, $(\cdot)$ denotes an ordered sequence and $[\cdot]$ denotes a multiset, equivalent to the empirical distribution.

Symbol	*Meaning of symbol*	*Definition*
$\mathbf{x}$	One-level sequence of link features.	$\mathbf{x} = (x_{(i,j)})_{(i,j)\in\mathcal{L}}$
	One-level sequence of node features.	$\mathbf{x} = (x_i)_{i\in\mathcal{V}}$
	One-level sequence of snapshot features.	$\mathbf{x} = (x^t)_{t\in\mathcal{T}}$
	Two-level sequence of link features.	$\mathbf{x} = (\mathbf{x}_{(i,j)})_{(i,j)\in\mathcal{L}}$ [a]
	Two-level sequence of node features.	$\mathbf{x} = (\mathbf{x}_i)_{i\in\mathcal{V}}$ [b]
$\boldsymbol{\pi}_{\mathcal{L}}$	Sequence of local distributions on links.	$\boldsymbol{\pi}_{\mathcal{L}}(\mathbf{x}) = (\pi_{(i,j)}(\mathbf{x}_{(i,j)}))_{(i,j)\in\mathcal{L}}$ [d]
$\boldsymbol{\pi}_{\mathcal{V}}$	Sequence of local distributions on nodes.	$\boldsymbol{\pi}_{\mathcal{V}}(\mathbf{x}) = (\pi_i(\mathbf{x}_i))_{i\in\mathcal{V}}$ [e]
$\boldsymbol{\pi}_{\mathcal{T}}$	Sequence of local distributions in snapshots.	$\boldsymbol{\pi}_{\mathcal{T}}(\mathbf{x}) = (\pi^t(\mathbf{x}^t))_{t\in\mathcal{T}}$ [f]
$p_{\mathcal{L}}$	Distribution of local sequences on links.	$p_{\mathcal{L}}(\mathbf{x}) = [\mathbf{x}_{(i,j)}]_{(i,j)\in\mathcal{L}}$ [a]
$p_{\mathcal{V}}$	Distribution of local sequences on nodes.	$p_{\mathcal{V}}(\mathbf{x}) = [\mathbf{x}_i]_{i\in\mathcal{V}}$ [b]
$p_{\mathcal{T}}$	Distribution of local sequences in snapshots.	$p_{\mathcal{T}}(\mathbf{x}) = [\mathbf{x}^t]_{t\in\mathcal{T}}$ [c]
p	Distribution of one-level link features.	$p(\mathbf{x}) = [x_{(i,j)}]_{(i,j)\in\mathcal{L}}$
	Distribution of one-level node features	$p(\mathbf{x}) = [x_i]_{i\in\mathcal{V}}$
	Distribution of one-level snapshot features	$p(\mathbf{x}) = [x^t]_{t\in\mathcal{T}}$
	Global distribution of two-level link features.	$p(\mathbf{x}) = [x^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)},\,(i,j)\in\mathcal{L}}$
	Global distribution of two-level node features.	$p(\mathbf{x}) = [x^m_i]_{m\in\mathcal{M}_i,\,i\in\mathcal{V}}$
$\boldsymbol{\mu}_{\mathcal{L}}$	Sequence of local means on links.	$\boldsymbol{\mu}_{\mathcal{L}}(\mathbf{x}) = (\mu_{(i,j)}(\mathbf{x}_{(i,j)}))_{(i,j)\in\mathcal{L}}$ [g]
$\boldsymbol{\mu}_{\mathcal{V}}$	Sequence of local means on nodes.	$\boldsymbol{\mu}_{\mathcal{V}}(\mathbf{x}) = (\mu_i(\mathbf{x}_i))_{i\in\mathcal{V}}$ [h]
$\boldsymbol{\mu}_{\mathcal{T}}$	Sequence of local means in snapshots.	$\boldsymbol{\mu}_{\mathcal{T}}(\mathbf{x}) = (\mu^t(\mathbf{x}^t))_{t\in\mathcal{T}}$ [i]
μ	Mean of one-level link features.	$\mu(\mathbf{x}) = \sum_{(i,j)\in\mathcal{L}} x_{(i,j)}/L$
	Mean of one-level node features.	$\mu(\mathbf{x}) = \sum_{i\in\mathcal{V}} x_i/N$
	Mean of one-level snapshot features.	$\mu(\mathbf{x}) = \sum_{t\in\mathcal{T}} x^t/T$
	Global mean of two-level link features.	$\mu(\mathbf{x}) = \sum_{(i,j)\in\mathcal{L}} \sum_{m\in\mathcal{M}_{(i,j)}} x^m_{(i,j)}/(\sum_{(i,j)\in\mathcal{L}} M_{(i,j)})$
	Global mean of two-level node features.	$\mu(\mathbf{x}) = \sum_{i\in\mathcal{V}} \sum_{m\in\mathcal{M}_i} x^m_i/(\sum_{i\in\mathcal{V}} M_i)$
–	Feature is not conserved.	

[a] $\mathbf{x}_{(i,j)}$: Local sequence on link, $\mathbf{x}_{(i,j)} = (x^m_{(i,j)})_{m\in\mathcal{M}_{(i,j)}}$, where $\mathcal{M}_{(i,j)}$ is a temporally ordered index set.
[b] $\mathbf{x}_i$: Local sequence on node, $\mathbf{x}_i = (x^m_i)_{m\in\mathcal{M}_i}$, where $\mathcal{M}_i$ is a temporally ordered index set.
[c] $\mathbf{x}^t$: Local sequence in snapshot, $\mathbf{x}^t = (x^t_i)_{i\in\mathcal{V}}$.
[d] $\pi_{(i,j)}(\mathbf{x}_{(i,j)})$: Local distribution on link, $\pi_{(i,j)}(\mathbf{x}_{(i,j)}) = [x^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}}$.
[e] $\pi_i(\mathbf{x}_i)$: Local distribution on node, $\pi_i(\mathbf{x}_i) = [x^m_i]_{m\in\mathcal{M}_i}$.
[f] $\pi^t(\mathbf{x}^t)$: Local distribution in snapshot, $\pi^t(\mathbf{x}^t) = [x^t_i]_{i\in\mathcal{V}}$.
[g] $\mu_{(i,j)}(\mathbf{x}_{(i,j)})$: Local mean on link, $\mu_{(i,j)}(\mathbf{x}_{(i,j)}) = \sum_{m\in\mathcal{M}_{(i,j)}} x^m_{(i,j)}/M_{(i,j)}$.
[h] $\mu_i(\mathbf{x}_i)$: Local mean on node, $\mu_i(\mathbf{x}_i) = \sum_{m\in\mathcal{M}_i} x^m_i/M_i$.
[i] $\mu^t(\mathbf{x}^t)$: Local mean in snapshot, $\mu^t(\mathbf{x}^t) = \sum_{i\in\mathcal{V}} x^t_i/N$.

Supplementary TABLE S2: **Additional features of directed temporal networks.** Several of the studies in the literature survey (Sec. III) considered MRRMs specifically defined for directed temporal networks, namely $\mathrm{P}[\mathbf{d}_{\rightarrow}]$ and $\mathrm{P}[\mathbf{s}_{\rightarrow}, p(\boldsymbol{\Delta\tau})]$. The definition of MRRMs that take the directionality of events into account is straightforward as it simply requires defining the appropriate directed features. For features of links, no generalizations are necessary since they all generalize automatically to directed networks by using the convention that (i, j) designates an interaction from i to j. However, since in directed networks a link from i to j does not imply the presence of the reciprocal link from j to i, the interpretation of link features may change. For each feature of nodes, three generalizations typically exist: an outgoing version, e.g., the out-strength $s_{i\rightarrow}$, an ingoing version, e.g., the in-strength $s_{i\leftarrow}$, and a combined version, e.g., the total strength $s_i = s_{i\rightarrow} + s_{i\leftarrow}$. We here list some generalizations of node features to directed temporal networks. Below, $(\cdot)$ denotes an ordered sequence, $\{\cdot\}$ denotes an unordered set, $|\cdot|$ denotes the cardinality of a set, : means "for which" or "such that".

Symbol	*Meaning of symbol*	*Definition*
Topological-temporal features		
$\mathcal{V}_{i\rightarrow}$	Outgoing neighborhood of node.	$\mathcal{V}_{i\rightarrow} = \{j : (i,j) \in \mathcal{L}\}$
$\mathcal{V}_{i\leftarrow}$	Incoming neighborhood of node.	$\mathcal{V}_{i\leftarrow} = \{j : (j,i) \in \mathcal{L}\}$
$\mathcal{V}_i$	Neighborhood of node.	$\mathcal{V}_i = \{j : (i,j) \in \mathcal{L} \text{ or } (j,i) \in \mathcal{L}\}$
Topological-temporal features		
$d^t_{i\rightarrow}$	Instantaneous out-degree.	$d^t_{i\rightarrow} = \lvert\{j : (i,j) \in \mathcal{E}^t\}\rvert$
$d^t_{i\leftarrow}$	Instantaneous in-degree.	$d^t_{i\leftarrow} = \lvert\{j : (j,i) \in \mathcal{E}^t\}\rvert$
d^t_i	Instantaneous (total) degree	$d^t_i = \lvert\{j : (i,j) \in \mathcal{E}^t \text{ or } (j,i) \in \mathcal{E}^t\}\rvert$
$\Phi_{i\rightarrow}$	Node activity timeline.	$\Phi_{i\rightarrow} = \left((v^1_{i\rightarrow}, \alpha^1_{i\rightarrow}), (v^2_{i\rightarrow}, \alpha^2_{i\rightarrow}), \ldots, (v^{a_{i\rightarrow}}_{i\rightarrow}, \alpha^{a_{i\rightarrow}}_{i\rightarrow})\right)$
$\alpha^m_{i\rightarrow}$	Activity duration.	Consecutive interval during which i has at least one outgoing contact.
$\Delta\alpha^m_{i\rightarrow}$	Inactivity duration.	$\Delta\alpha^m_{i\rightarrow} = v^{m+1}_{i\rightarrow} - (v^m_{i\rightarrow} + \alpha^m_{i\rightarrow})$
Aggregated features		
$a_{i\rightarrow}$	Outgoing node activity.	$a_{i\rightarrow} = \sum_{j\in\mathcal{V}_{i\rightarrow}} n_{(i,j)}$
$a_{i\leftarrow}$	Ingoing node activity.	$a_{i\leftarrow} = \sum_{j\in\mathcal{V}_{i\leftarrow}} n_{(i,j)}$
a_i	(Total) node activity.	$a_i = \sum_{(i,j)\in\mathcal{L}_i} n_{(i,j)}$
$s_{i\rightarrow}$	Node out-strength.	$s_{i\rightarrow} = \sum_{(i,j)\in\mathcal{L}_{i\rightarrow}} w_{(i,j)}$
$s_{i\leftarrow}$	Node in-strength.	$s_{i\leftarrow} = \sum_{(i,j)\in\mathcal{L}_{i\leftarrow}} w_{(i,j)}$
s_i	Node (total) strength.	$s_i = \sum_{(i,j)\in\mathcal{L}_i} w_{(i,j)}$
$k_{i\rightarrow}$	Node out-degree.	$k_{i\rightarrow} = \lvert\mathcal{V}_{i\rightarrow}\rvert$
$k_{i\leftarrow}$	Node in-degree.	$k_{i\leftarrow} = \lvert\mathcal{V}_{i\leftarrow}\rvert$
k_i	Node (total) degree.	$k_i = \lvert\mathcal{V}_i\rvert$

Supplementary Note 1: Applying instant-event shufflings to temporal networks with event durations

Event shufflings by definition conserve the events' durations in a temporal network, but we may randomize the durations events in a temporal network by first representing the network as an instant-event network and then applying an instant-event shuffling to it using the following procedure:

i) Choose an appropriate time-resolution for discretization. A natural choice may be the time-resolution of recordings, but for high resolution measurements a lower time-resolution may be more practical.

ii) Construct the corresponding instant-event temporal network by defining an instantaneous event between a pair of nodes at the start of each time-interval during which the nodes are in contact (in the temporal network shown Fig. II.2 this splits the four long events into two instantaneous events each and creates one instantaneous event for each of the shorter events).

iii) Randomize this instant-event network using an instant-event shuffling.

iv) Recreate a randomized version of the temporal network by concatenating consecutive instantaneous events between the same nodes into single events.

Table IV.2 show the effects of both event and instant-event shufflings on the features of a temporal network. To understand the effects of the same shufflings on an instant-event network one should simply ignore the features that are not defined for instant-event temporal networks, i.e. a_i, $n_{(i,j)}$, $\alpha^m_{(i,j)}$ and $\tau^m_{(i,j)}$.

From the above procedure, the number of instantaneous events on a link in the generated instant-event network is seen to correspond to the weight $w_{(i,j)}$ of the link in the original temporal network. Using $w_{(i,j)}$ to designate the number of instantaneous events on the links in the instant-event network thus makes it possible to name each event and instant-event shuffling based on the features they conserve in a consistent manner, no matter whether the shuffling is applied to a temporal network or an instant-event temporal network. It follows from this alternative definition of $w_{(i,j)}$ that s_i designates the activity of the node i in an instant-event network (see the section "Notation for features of instant-event networks" in Supplementary Note 2 for further discussion).

Supplementary Note 2: Features of temporal networks

In this supplementary note we provide detailed definitions for a selection of temporal network features that have been shown to play an important role in network dynamics, and which are sufficient to name the MRRMs found in our literature survey presented in Section IV B. Table IV.1 lists basic features of temporal networks, often describing single elements of a network such as nodes or events, and Table S1 lists different general ways to construct features describing the whole temporal network by using the basic features as building blocks.

A feature of a temporal network is any function that takes a network as an input (Def. II.4). Clearly, there is a very large number of such functions which could be defined[7], and as we will see later, a multitude of such functions have been (often implicitly) used in the literature. Here we attempt to organize this set of functions in a way that it is compatible with the different temporal network representations introduced in Section II A and the concept of order of features developed in Section IV A.

For definiteness we will here consider features of temporal networks with event durations. In order to make our description of MRRMs consistent for both networks with and without event durations, we shall need to modify the definitions of some of these features for instant-event networks. We do this at the end of this note.

Many features are ones returning a sequence of lower-dimensional features (Def. IV.3), e.g., the degree sequence of a static graph, $\mathbf{k}$, is given by the sequence of the individual node degrees k_i. Temporal network features are often given by a nested sequence where individual features in the sequence themselves are a sequence of scalar features. Sequences and nested sequences can further be turned into distributions and average values in multiple ways.

The basic building blocks of many features constructed as sequences and used in MRRMs are scalar features describing single elements of the network (Table IV.1). We mainly consider two types of sequences: *one-level* sequences of scalar features and *two-level* sequences of sequences of scalar features.

Definition S2.1. *One-level and two-level sequences.*

1. *One-level sequence.* We refer to a non-nested sequence of scalar features, $\mathbf{x} = (x_q)_{q\in\mathcal{Q}}$, as a *one-level sequence*.
2. *Two-level sequence.* We refer to a nested sequence of features that are themselves one-level sequences, $\mathbf{x} = (\mathbf{x}_q)_{q\in\mathcal{Q}}$ with $\mathbf{x}_q = (x_q^r)_{r\in\mathcal{R}_q}$, as a *two-level sequence*. We refer to the $\mathbf{x}_q$ as *local sequences*.

Detailed definitions of particular types of one- and two-level sequences are given in Table S1 (symbol: $\mathbf{x}$).

One-level sequences are typically used to represent features that are aggregated over the temporal or topological dimension of the temporal network, while two-level sequence are composed of features that depend both on topology and time. The following examples illustrate this.

Example S2.1. *One-level sequence of static degrees.* A well known example of an aggregated graph feature is the node degree, k_i, giving the number of nodes in G^{stat} that are connected to the node i.

Example S2.2. *Two-level sequence of instantaneous degrees.* A generalization to temporal networks of the static degree k_i of a node is the *instantaneous degree* d_i^t. It is given by the number of nodes that the node i is in contact with at time t. The (two-level) sequence of instantaneous degrees is $\mathbf{d} = (\mathbf{d}^t)_{t\in\mathcal{T}} = ((d_i^t)_{i\in\mathcal{V}})_{t\in\mathcal{T}}$, or alternatively $\mathbf{d} = (\mathbf{d}_i)_{i\in\mathcal{V}} = ((d_i^t)_{t\in\mathcal{T}})_{i\in\mathcal{V}}$ since the order of the indices i and t does not matter here.

Example S2.3. *Two-level sequence of inter-event durations.* A feature of temporal networks that has been shown to have a profound impact on dynamic processes is the durations between consecutive events in the timelines, termed the *inter-event durations* and defined by $\Delta\tau_{(i,j)}^m = t_{(i,j)}^{m+1} - (t_{(i,j)}^m + \tau_{(i,j)}^m)$. Their (two-level) sequence is $\boldsymbol{\Delta\tau} = (\boldsymbol{\Delta\tau}_{(i,j)})_{(i,j)\in\mathcal{L}}$, where $\boldsymbol{\Delta\tau}_{(i,j)} = (\Delta\tau_{(i,j)}^m)_{m\in\mathcal{M}_{(i,j)}}$. Here $\mathcal{M}_{(i,j)} = \{1, 2, \ldots, n_{(i,j)} - 1\}$ indexes the inter-event durations in the timeline $\Theta_{(i,j)}$ by temporal order, with $n_{(i,j)}$ the number of events in the timeline. Note that due to the temporal extent of the inter-event durations, we cannot inverse the order of the indices m and (i,j) as we could for the instantaneous degrees presented in the previous example.

Example S2.4. *Sequence of snapshot graphs.* A notable example of a sequence of features that are neither scalar nor sequences of scalars is the sequence of snapshot graphs $\boldsymbol{\Gamma} = (\Gamma^t)_{t\in\mathcal{T}}$ (Def. II.7).

Instead of constraining an ordered sequence itself, many MRRMs constrain marginal distributions (Def. IV.4) or moments of a sequence. Before we define these marginals and moments in detail for temporal networks, we consider as a simpler example the degree sequence of a static graph.

Example S2.5. From the sequence of degrees in a static graph, $\mathbf{k}$, we may calculate their marginal distribution

[7] The number of functions leading to different MRRMs is equal to the the number of possible partitions of the state space of temporal networks of a given size. In practice, the number of possible temporal networks is very large and the number of partitions of the state space is a super-exponential function of this number (see Section II B).

$p(\mathbf{k})$ (equivalent to the multiset of their values) as well as their mean $\mu(\mathbf{k})$. This leads to three different features, each corresponding to a different MRRM: one that constrains the complete sequence of degrees, P[$\mathbf{k}$], one that constrains their distribution, P[$p(\mathbf{k})$], and one that constrains their mean P[$\mu(\mathbf{k})$] (which is equivalent to P[L] if the number of nodes N is kept constant). (P[$\mathbf{k}$] and P[$p(\mathbf{k})$] are both often referred to as the configuration model or the Maslov-Sneppen model, and P[$\mu(\mathbf{k})$] = P[L] is the Erdős-Rényi model with a fixed number of links.) The three features (and corresponding MRRMs) satisfy a linear order: $\mathbf{k} \leq p(\mathbf{k}) \leq \mu(\mathbf{k})$.

Since we have both a topological and temporal dimension in temporal networks, a much larger number of different ways to marginalize the sequence of features is possible than for a simple static graph. We define here those needed to characterize the MRRMs surveyed in this article, but many more could be defined. The different ways to marginalize a sequence of features give rise to the following different three types of distributions.

Definition S2.2. *Distributions over a sequence.* We here list different ways to construct distributions of feature values by marginalizing over a sequence $\mathbf{x}$. The different types of distributions are defined in more detail in Table S1 (symbols: p, $p_{\mathcal{L}}$, $p_{\mathcal{V}}$, $p_{\mathcal{T}}$, $\boldsymbol{\pi}_{\mathcal{L}}$, $\boldsymbol{\pi}_{\mathcal{V}}$, and $\boldsymbol{\pi}_{\mathcal{T}}$).

1. *Global distribution* $p(\mathbf{x})$. The global distribution $p(\mathbf{x})$ returns the number of times each possible scalar value appears in a measured sequence $\mathbf{x}^* = \mathbf{x}(G^*)$. For a one-level sequence $\mathbf{x}^* = (x_q^*)_{q\in\mathcal{Q}}$, it is obtained by marginalizing over the sole index set $\mathcal{Q}$: $p(\mathbf{x}^*) = [x_q^*]_{q\in\mathcal{Q}}$. For a two-level sequence $\mathbf{x}^* = (((x_q^r)^*)_{r\in\mathcal{R}_q})_{q\in\mathcal{Q}}$, it is obtained by marginalizing both over the inner and outer index sets, $\mathcal{R}_q$ and $\mathcal{Q}$: $p(\mathbf{x}^*) = [(x_q^r)^*]_{r\in\mathcal{R}_q, q\in\mathcal{Q}}$.

2. *Distribution of local features* $p_{\mathcal{Q}}(\mathbf{x})$. For a sequence of non-scalar features, $\mathbf{x} = (\mathbf{x}_q)_{q\in\mathcal{Q}}$, the distribution of local features $p_{\mathcal{Q}}(x^*) = [\mathbf{x}_q^*]_{q\in\mathcal{Q}}$ reports the number of times each possible value of the local features $\mathbf{x}_q$ appears in a measured sequence $\mathbf{x}^*$.

3. *Sequence of local distributions* $\boldsymbol{\pi}_{\mathcal{Q}}(\mathbf{x})$. For a two-level sequence of features, $\mathbf{x} = (\mathbf{x}_q)_{q\in\mathcal{Q}}$, the sequence of local distributions $\boldsymbol{\pi}_{\mathcal{Q}}(\mathbf{x})$ is given by the ordered tuple $\boldsymbol{\pi}_{\mathcal{Q}}(\mathbf{x}^*) = (\pi_q(\mathbf{x}_q^*))_{q\in\mathcal{Q}}$, where each local distribution $\pi_q(\mathbf{x}_q^*) = [(x_q^r)^*]_{r\in\mathcal{R}_q}$ is the distribution of the scalar features in the local sequence $\mathbf{x}_q^*$.

The following examples illustrate the different types of distributions.

Example S2.6. *Global distribution.* The global distribution of the static degrees is given by $p(\mathbf{k}^*) = [k_i^*]_{i\in\mathcal{V}}$. The global distributions of the two-level instantaneous node degrees and the inter-event durations on links are given by $p(\mathbf{d}^*) = [(d_i^t)^*]_{t\in\mathcal{T}, i\in\mathcal{V}}$ and $p(\boldsymbol{\Delta\tau}^*) = [(\Delta\tau_{(i,j)}^m)^*]_{m\in\mathcal{M}_{(i,j)}, (i,j)\in\mathcal{L}}$, respectively.

Example S2.7. *Distribution of local features.* Two different types of distributions of local sequences of instantaneous degrees can be constructed from the sequence of instantaneous degrees: the distribution of local sequences of the instantaneous degrees of each node, given by $p_{\mathcal{V}}(\mathbf{d}^*) = [\mathbf{d}_i^*]_{i\in\mathcal{V}}$, and the distribution of local sequences of instantaneous degrees of nodes in each snapshot, given by $p_{\mathcal{T}}(\mathbf{d}^*) = [(\mathbf{d}^t)^*]_{t\in\mathcal{T}}$. From the sequence of inter-event durations, we can construct the distribution of local sequences of inter-event durations on the links, $p_{\mathcal{L}}(\boldsymbol{\Delta\tau}^*) = [\boldsymbol{\Delta\tau}_{(i,j)}^*]_{(i,j)\in\mathcal{L}}$. From the snapshot-graph sequence, we can construct the distribution of snapshot graphs $p_{\mathcal{T}}(\boldsymbol{\Gamma}^*) = [(\Gamma^t)^*]_{t\in\mathcal{T}}$.

Example S2.8. *Sequence of local distributions.* We can also construct two different sequences of local distributions from the sequence of instantaneous degrees: the sequence of local distributions of the instantaneous degrees of each node, $\boldsymbol{\pi}_{\mathcal{V}}(\mathbf{d}^*) = ([(d_i^t)^*]_{t\in\mathcal{T}})_{i\in\mathcal{V}}$. and the sequence of local distributions of instantaneous degrees in each snapshot, $\boldsymbol{\pi}_{\mathcal{T}}(\mathbf{d}^*) = ([(d_i^t)^*]_{i\in\mathcal{V}})_{t\in\mathcal{T}}$. The sequence of local distributions of inter-event durations is given by $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}^*) = ([(\Delta\tau_{(i,j)}^m)^*]_{m\in\mathcal{M}_{(i,j)}})_{(i,j)\in\mathcal{L}}$. We cannot construct a sequence of local distributions from the sequence of snapshot graphs since they are not sequences.

After the above definitions of different local and global marginalizations of a sequence of features, we now consider ways to define its moments. We shall here consider only first-order moments, i.e. means, but note that one may generally consider also higher order moments such as the variance. As for the distributions, it is natural to define the mean of a one- or two-level sequence simply as the average over the values of their scalar elements. For a two-level sequence, we may additionally construct a sequence of means of the local sequences.

Definition S2.3. *Means of a one- or two-level sequence of features.* We shall consider two different ways to average over a one- or two-level sequence. Detailed definitions for specific kinds of features are given in Table S1 (symbols: μ, $\boldsymbol{\mu}_{\mathcal{L}}$, $\boldsymbol{\mu}_{\mathcal{V}}$, and $\boldsymbol{\mu}_{\mathcal{T}}$).

1. *Global mean* $\mu(\mathbf{x})$. The global mean $\mu(\mathbf{x})$ of a sequence of features is defined as the average over all individual scalar features in $\mathbf{x}$. For a one-level sequence, it is given by $\mu(\mathbf{x}^*) = \sum_{q\in\mathcal{Q}} x_q^*/Q$, where Q is the number of elements in $\mathbf{x}^*$. For a two-level sequence, it is $\mu(\mathbf{x}^*) = \sum_{q\in\mathcal{Q}}\sum_{r\in\mathcal{R}_q}(x_q^r)^*/(\sum_{q\in\mathcal{Q}} R_q)$, where R_q is the number of elements in $\mathbf{x}_q^*$.

2. *Sequence of local means* $\boldsymbol{\mu}_{\mathcal{Q}}(\mathbf{x})$. For a two-level sequence $\mathbf{x} = (\mathbf{x}_q)_{q\in\mathcal{Q}}$, the sequence of local means $\boldsymbol{\mu}_{\mathcal{Q}}(\mathbf{x})$ is defined as the sequence of the means of each local sequence $\mathbf{x}_q$. Each of these *local* means is given by $\mu_q(\mathbf{x}_q^*) = \sum_{r\in\mathcal{R}_q}(x_q^r)^*/R_q$.

The different types of means obtained are illustrated in the following examples.

Example S2.9. *Global mean.* The mean of the one-level sequence of static degrees is given by $\mu(\mathbf{k}^*) = \sum_{i\in\mathcal{V}} k_i^*/N$. The global mean of the two-levels instantaneous degrees is given by $\mu(\mathbf{d}^*) = \sum_{i\in\mathcal{V}}\sum_{t\in\mathcal{T}}(d_i^t)^*/(NT)$, and the global mean of the inter-event durations is $\mu(\boldsymbol{\Delta\tau}^*) = \sum_{(i,j)\in\mathcal{L}}\sum_{m\in\mathcal{M}_{(i,j)}}(\Delta\tau_{(i,j)}^m)^*/\sum_{(i,j)\in\mathcal{L}} M_{(i,j)}$.

Example S2.10. *Sequence of local means.* Sequences of local means of the instantaneous degrees can be constructed in two ways: as the sequence of local means of the instantaneous degrees of each node, $\boldsymbol{\mu}_{\mathcal{V}}(\mathbf{d}^*) = (\sum_{t\in\mathcal{T}}(d_i^t)^*/T)_{i\in\mathcal{V}}$, and as the sequence of local means of the instantaneous degrees in each snapshot, $\boldsymbol{\mu}_{\mathcal{T}}(\mathbf{d}^*) = (\sum_{i\in\mathcal{V}}(d_i^t)^*/N)_{t\in\mathcal{T}}$. For the inter-event durations, the sequence of local means on each link is $\boldsymbol{\mu}_{\mathcal{L}}(\boldsymbol{\Delta\tau}^*) = (\sum_{m\in\mathcal{M}_{(i,j)}}(\Delta\tau_{(i,j)}^m)^*/M_{(i,j)})_{(i,j)\in\mathcal{L}}$.

The distributions and means defined above are functions of the sequence of features, so they are all coarser than the sequence. Many of them are also comparable (though not all of them), so we can establish a hierarchy between them using Definition IV.1.Table S1 lists all the distributions and moments for features of links, nodes, and snapshots, and we establish their hierarchies in Fig. S1.

By combining Tables IV.1 and S1, as shown in the examples above, we may describe most features constrained by MRRMs found in the literature.

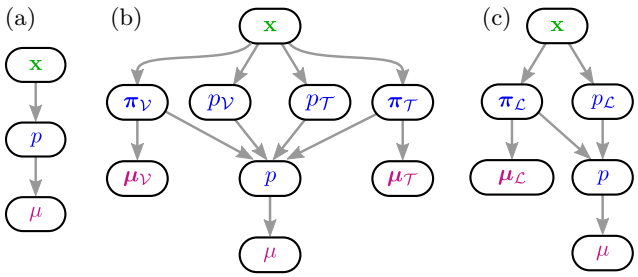
Some of the different basic features listed in Table IV.1 are also pairwise comparable. This enables us to construct a hierarchy of the different features listed in Table IV.1 together with their marginals and moments (Table S1). Figure S2 shows such a hierarchy. It may be used to derive which features are conserved by a MRRM that constrains a given feature: the MRRM conserves all features that are below the constrained feature in the hierarchy.

Note that if two features are not comparable, it does not imply that they are independent. So one cannot conclude from the absence of a link between two features in Fig. S2 that one does not influence the other, only that it does not constrain it completely; the features may be correlated. The correlations between features that are neither comparable nor independent depend on the input temporal network that is considered. Thus, they can only be investigated on a case-by-case basis.

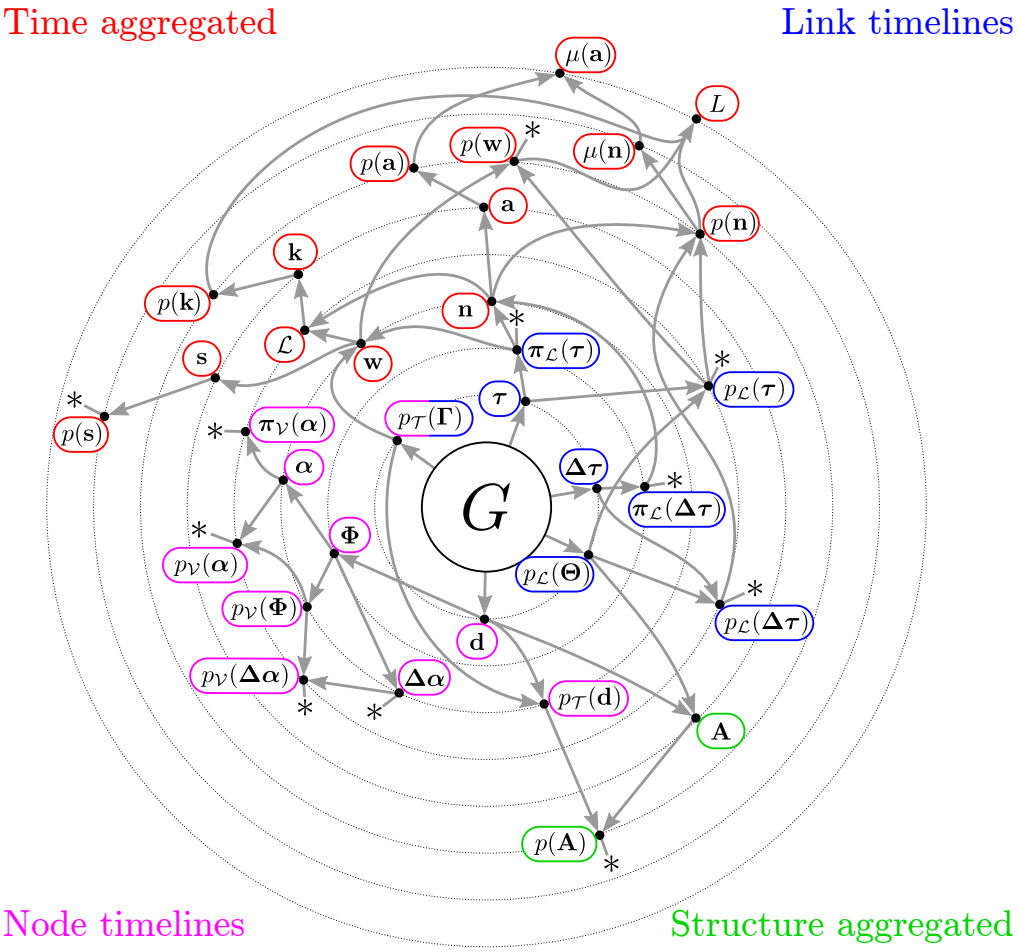
Notation for features of instant-event networks

Whenever possible we use the same symbols and names for features of instant-event temporal networks as for networks with event durations. However, we have adopted a different notation for the number of instantaneous events on a link, $w_{(i,j)}$, than for the number of events on a link, $n_{(i,j)}$ (Table IV.1). This is needed to make our description of MRRMs consistent when they are applied to both temporal networks and instant-event networks (see discussion in Supplementary Note 1). This furthermore means that s_i denotes the total number of instantaneous events that a node partakes in for instant-event networks, and that a_i, and $n_{(i,j)}$ are not defined for these networks. Similarly, the event durations $\tau_{(i,j)}^m$ and activity durations $\alpha_{(i,j)}^m$ are not defined for instant-event networks.

Several other definitions are changed slightly to accommodate the fact that instantaneous events do not have durations. This is namely the case for $\mathcal{L}$, $\Theta_{(i,j)}$, $\Delta\tau_{(i,j)}^m$, $t_{(i,j)}^w$, $\Delta\alpha_{(i,j)}^m$, and Φ_i (see Table IV.1). Conversely, the snapshot-graph sequence is only defined for instant-event networks, and thus so are also the following associated features: $\mathcal{T}$, Γ^t, and $\mathcal{E}^t$.



Supplementary FIG. S1: **Hierarchies of the marginals and moments of a sequence of features.** An arrow from a higher node to a lower one indicates that the former feature is finer than the latter. Thus, a MRRM that conserves the former feature necessarily conserves all downstream features. Conversely, a MRRM that randomizes a given feature also randomizes any features upstream of it as well. (a) For a sequences of scalar features, e.g. the static degrees $\mathbf{k} = (k_i)_{i\in\mathcal{V}}$. (b) For two-level sequences of features of nodes , namely $\boldsymbol{\alpha}$, $\boldsymbol{\Delta\alpha}$ and $\mathbf{d}$. (c) For two-level sequences of features of link timelines, namely $\boldsymbol{\tau}$ and $\boldsymbol{\Delta\tau}$.



Supplementary FIG. S2: **Feature hierarchy diagram.** An arrow from a higher ranking (more central) to a lower ranking feature (vertex in the diagram) indicates that the former feature is finer than the latter. Thus, a MRRM that constrains a given feature also constrains all downstream features. Conversely, a MRRM that randomizes (i.e. does not constrain) a given feature does not constrain any of the upstream features either. See Tables IV.1 and S1 for definitions of the features. A star (∗) emanating from a node indicates that lower hierarchical levels follow as shown in Fig. S1. The color coding shows a division of features into different types: time-aggregated features (i.e. topological and weighted), link-timeline features, node-timeline features, and structure-aggregated features (i.e. purely temporal).

Supplementary Note 3: Analyzing a face-to-face interaction network using MRRMs

In this supplementary note we apply a selection of MRRMs to build a statistical portrait of a temporal network of face-to-face interactions recorded in a primary school [107, 108]. The dataset, which is freely available at `www.sociopatterns.org/datasets` was recorded with a time-resolution of 20 s and forms a temporal network consisting of 242 nodes and 77 521 events of varying duration.

We apply a series of MRRMs to gradually randomize different network features, allowing us to unravel the features' effects on the network's topology and temporal structure. Supplementary Table S3 lists which temporal network features the different models conserve (a more detailed description is found in Table IV.2). The MRRMs are described in detail in Section IV B.

Supplementary Figure S3 is our statistical portrait of the temporal network. It quantifies how far the values a selection of features of the empirical network are from the networks generated by each MRRM. The differences are quantified by the Jensen-Shannon divergence (JSD) [109] between the null distributions and their distribution in the empirical network (for the activity timeline $\mathbf{A}$, the difference is quantified by the L1 distance). The values of the features in the empirical network and for the networks generated by each MRRM are shown in Supplementary Figs. S4 and S5 below. We next explore Supplementary Fig. S3 panel by panel and discuss what can be learned about the network from it.

We first study the activity timeline $\mathbf{A} = (|\mathcal{E}^t|)_{t\in\mathcal{T}}$. It is by construction completely constrained by all link shufflings and snapshot shufflings, while at the opposite end it is essentially completely randomized by $\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})]$, $\mathrm{P}[\mathcal{L}, p(\boldsymbol{\tau})]$, and $\mathrm{P}[p(\boldsymbol{\tau})]$ (see Supplementary Fig. S5). This shows that $\mathbf{A}$ is not constrained by the static graph of the network. Comparison between $\mathrm{P}[p(\boldsymbol{\tau})]$ and $\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})]$ shows that the distribution of inter-event durations does affect $\mathbf{A}$, but not to a large extent. Comparing this with $\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \mathbf{t}^1, \mathbf{t}^w]$ shows that the timing of the first and last events on each link does on the other hand have a significant effect on $\mathbf{A}$ in the network. Furthermore, comparison with $\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}), \mathbf{t}^1]$ shows that constraining both $\mathbf{t}^1$ and $\mathbf{t}^w$ together with $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$ imposes an even stronger constraint on the activity timeline (see also Supplementary Fig. S5).

We next consider time-aggregated features of the temporal network, starting with the distribution of node degrees, $p(\mathbf{k}) = [k_i]_{i\in\mathcal{V}}$. This feature is constrained by most of the MRRMs applied in this example, with the exception of $\mathrm{P}[p_{\mathcal{L}}(\boldsymbol{\Theta})]$ (which draws G^{stat} from an Erdős-Rényi model), and $\mathrm{P}[p(\mathbf{t}, \boldsymbol{\tau})]$ and $\mathrm{P}[p(\boldsymbol{\tau})]$ (which do not conserve the number of links in G^{stat}). The high divergence seen for $\mathrm{P}[p_{\mathcal{L}}(\boldsymbol{\Theta})]$ shows that the empirical network's degree distribution is significantly nonrandom (even if it does not seem to follow a broad-tailed distribution, see Supplementary Fig. S4).

Most of the MRRMs also conserve the distributions of link weights and event frequencies, $p(\mathbf{w}) = [w_{(i,j)}]_{(i,j)\in\mathcal{L}}$, with $w_{(i,j)} = \sum_{m\in\mathcal{M}_{(i,j)}} \tau^m_{(i,j)}$ and $p(\mathbf{n}) = [n_{(i,j)}]_{(i,j)\in\mathcal{L}}$, with $n_{(i,j)} = |\Theta_{(i,j)}|$, respectively. The exceptions are $\mathrm{P}[\mathcal{L}, p(\boldsymbol{\tau})]$ (which conserves the static structure, but not the heterogeneity in the number and durations of events in timelines), and $\mathrm{P}[p(\mathbf{t}, \boldsymbol{\tau})]$ and $\mathrm{P}[p(\boldsymbol{\tau})]$ (which do not conserve the number of links in G^{stat}). The effects of these shufflings on $p(\mathbf{w})$ and $p(\mathbf{n})$ are very similar, highlighting the fact that $\mathbf{w}$ and $\mathbf{n}$ are highly correlated features. Note that the smaller divergences seen for the more random $\mathrm{P}[p(\boldsymbol{\tau})]$ than for $\mathrm{P}[\mathcal{L}, p(\boldsymbol{\tau})]$ are due to the JSD putting most weight on low values of $w_{(i,j)}$ and $n_{(i,j)}$ since these are most probable. Since $\mathrm{P}[p(\boldsymbol{\tau})]$ produces a much larger number of links than in the original network but conserves the number of events, it also produces a large fraction of links with low $w_{(i,j)}$ and $n_{(i,j)}$, similarly to the original network. Conversely, $\mathrm{P}[\mathcal{L}, p(\boldsymbol{\tau})]$ conserves the number of links and homogenizes $w_{(i,j)}$ and $n_{(i,j)}$, leading to fewer low values of these.

The majority of the shufflings do not constrain the distributions of node strengths and activities, $p(\mathbf{s}) = [\sum_{j\in\mathcal{V}} w_{(i,j)}]_{i\in\mathcal{V}}$ and $p(\mathbf{a}) = [\sum_{j\in\mathcal{V}} n_{(i,j)}]_{i\in\mathcal{V}}$, respectively. As for $p(\mathbf{w})$ and $p(\mathbf{n})$, their effects on the two features are very similar. Due to this we take $p(\mathbf{a})$ as example and note that the results are qualitatively the same for $p(\mathbf{s})$. The distribution of a_i in the empirical network is indistinguishable from networks generated by $\mathrm{P}[\mathbf{k}, p_{\mathcal{L}}(\boldsymbol{\Theta})]$. This shows that $p(\mathbf{a})$ is simply determined by the convolution of the individual distributions of $\mathbf{k}$ and $\mathbf{n}$ and that correlations between the two are unimportant. Comparison with $\mathrm{P}[\mathcal{L}, p(\boldsymbol{\tau})]$ and $\mathrm{P}[p_{\mathcal{L}}(\boldsymbol{\Theta})]$ shows that both $p(\mathbf{n})$ (randomized by $\mathrm{P}[\mathcal{L}, p(\boldsymbol{\tau})]$) and $p(\mathbf{k})$ (randomized by $\mathrm{P}[p_{\mathcal{L}}(\boldsymbol{\Theta})]$) are needed to reproduce the nonrandom shape of $p(\mathbf{a})$ though.

We finally investigate temporal-structural features of nodes and links. We note first that the distribution of event durations, $p(\boldsymbol{\tau}) = [\tau]_{(i,j,t,\tau)\in\mathcal{C}}$ is conserved by all the MRRMs employed by construction since these are all event shufflings (Def. A.2).

The distribution of inter-event durations on the links, $p(\boldsymbol{\Delta\tau}) = \cup_{(i,j)\in\mathcal{L}}[\Delta\tau^m_{(i,j)}]_{m\in\mathcal{M}_{(i,j)}}$, is constrained by all link shufflings, but not by most of the other shufflings. Comparison of the effects of $\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \mathbf{t}^1, \mathbf{t}^w]$ and $\mathrm{P}[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})]$ demonstrates that the timing of the first and last events in the timelines constrain the inter-event durations to some degree in the network (see also Supplementary Fig. S5). The much larger divergence found for $\mathrm{P}[\mathcal{L}, p(\boldsymbol{\tau})]$ highlights that the number of events $n_{(i,j)}$ on each link strongly influences the inter-event durations.

None of the MRRMs completely constrain the distri-

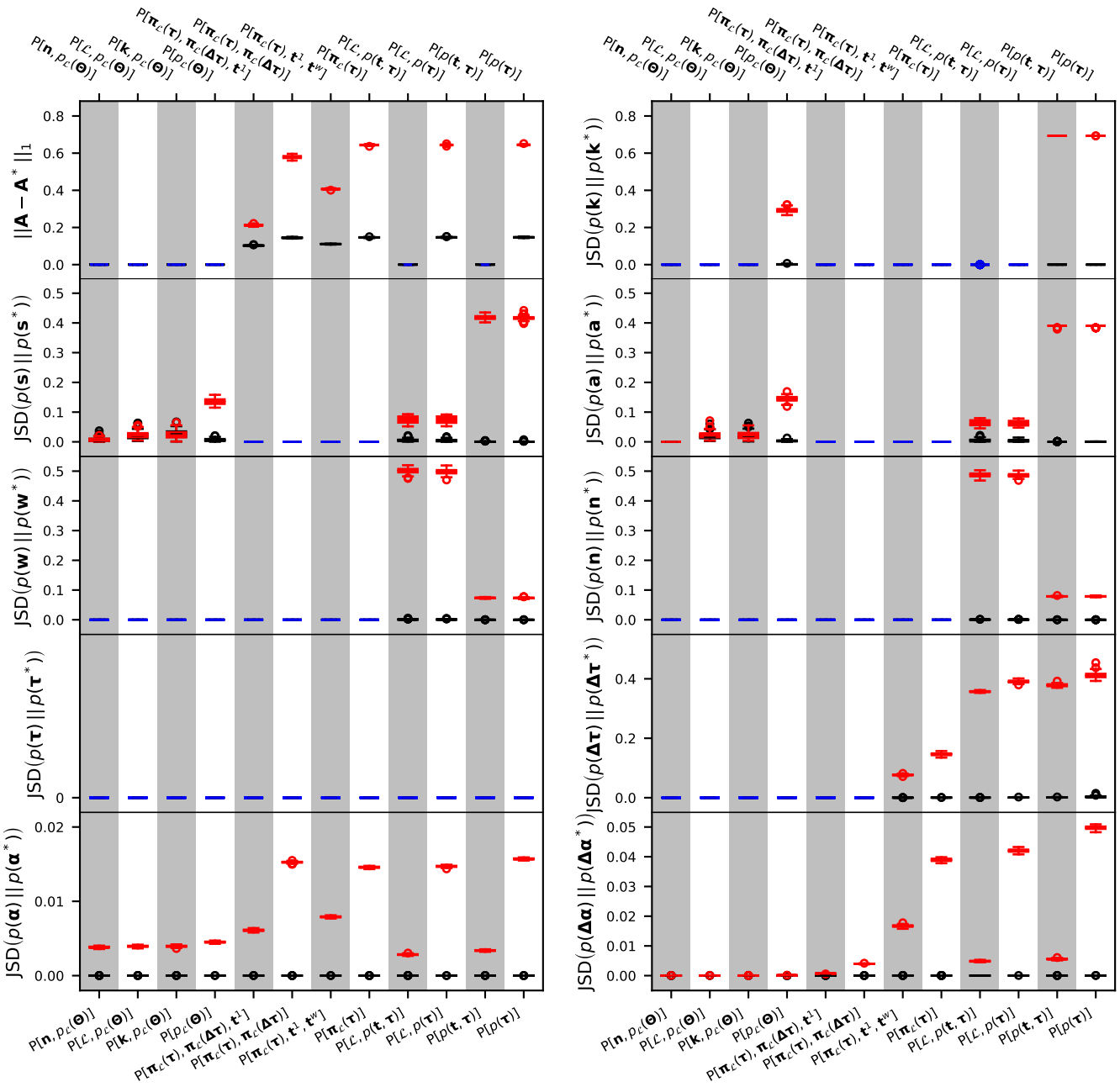
Supplementary TABLE S3: **Effects of selected MRRMs on temporal network features.** See Table IV.1 for definitions of features. Colored symbols show to what extent each feature is conserved. Informal definitions are found in the tablenotes (detailed definitions are found in Supplementary Table S1).

Model	Features													
	G^{stat}	$\mathbf{k}$	L	$\mathbf{a}$	$\mathbf{s}$	$\mathbf{n}$	$\mathbf{w}$	$\mathbf{A}$	$\boldsymbol{\alpha}$	$\boldsymbol{\Delta\alpha}$	$\boldsymbol{\tau}$	$\boldsymbol{\Delta\tau}$	$\mathbf{t}^1$	$\mathbf{t}^w$
P[$p(\boldsymbol{\tau})$]	✗	✗	✗	✗	μ	✗	✗	μ	✗	✗	p	✗	✗	✗
P[$p(\mathbf{t},\boldsymbol{\tau})$]	✗	✗	✗	✗	μ	✗	✗	✓	✗	✗	p	✗	✗	✗
P[$p_{\mathcal{L}}(\boldsymbol{\Theta})$]	✗	μ	✓	μ	μ	p	p	✓	✗	✗	p	p	p	p
P[$\mathbf{k},p_{\mathcal{L}}(\boldsymbol{\Theta})$]	✗	✓	✓	μ	μ	p	p	✓	✗	✗	p	p	p	p
P[$\mathcal{L}$,$p(\boldsymbol{\tau})$]	✓	✓	✓	μ	μ	μ	μ	μ	✗	✗	p	✗	✗	✗
P[$\mathcal{L}$,$p(\mathbf{t},\boldsymbol{\tau})$]	✓	✓	✓	μ	μ	μ	μ	✓	✗	✗	p	✗	✗	✗
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$]	✓	✓	✓	✓	✓	✓	✓	μ	✗	✗	p	✗	✗	✗
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$,$\mathbf{t}^1$,$\mathbf{t}^w$]	✓	✓	✓	✓	✓	✓	✓	μ	✗	✗	p	μ	✓	✓
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$,$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$]	✓	✓	✓	✓	✓	✓	✓	μ	✗	✗	p	p	✗	✗
P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$,$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$,$\mathbf{t}^1$]	✓	✓	✓	✓	✓	✓	✓	μ	✗	✗	p	p	✓	✓
P[$\mathcal{L}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$]	✓	✓	✓	μ	μ	p	p	✓	✗	✗	p	p	p	p
P[$\mathbf{n}$,$p_{\mathcal{L}}(\boldsymbol{\Theta})$]	✓	✓	✓	✓	μ	✓	p	✓	✗	✗	p	p	p	p

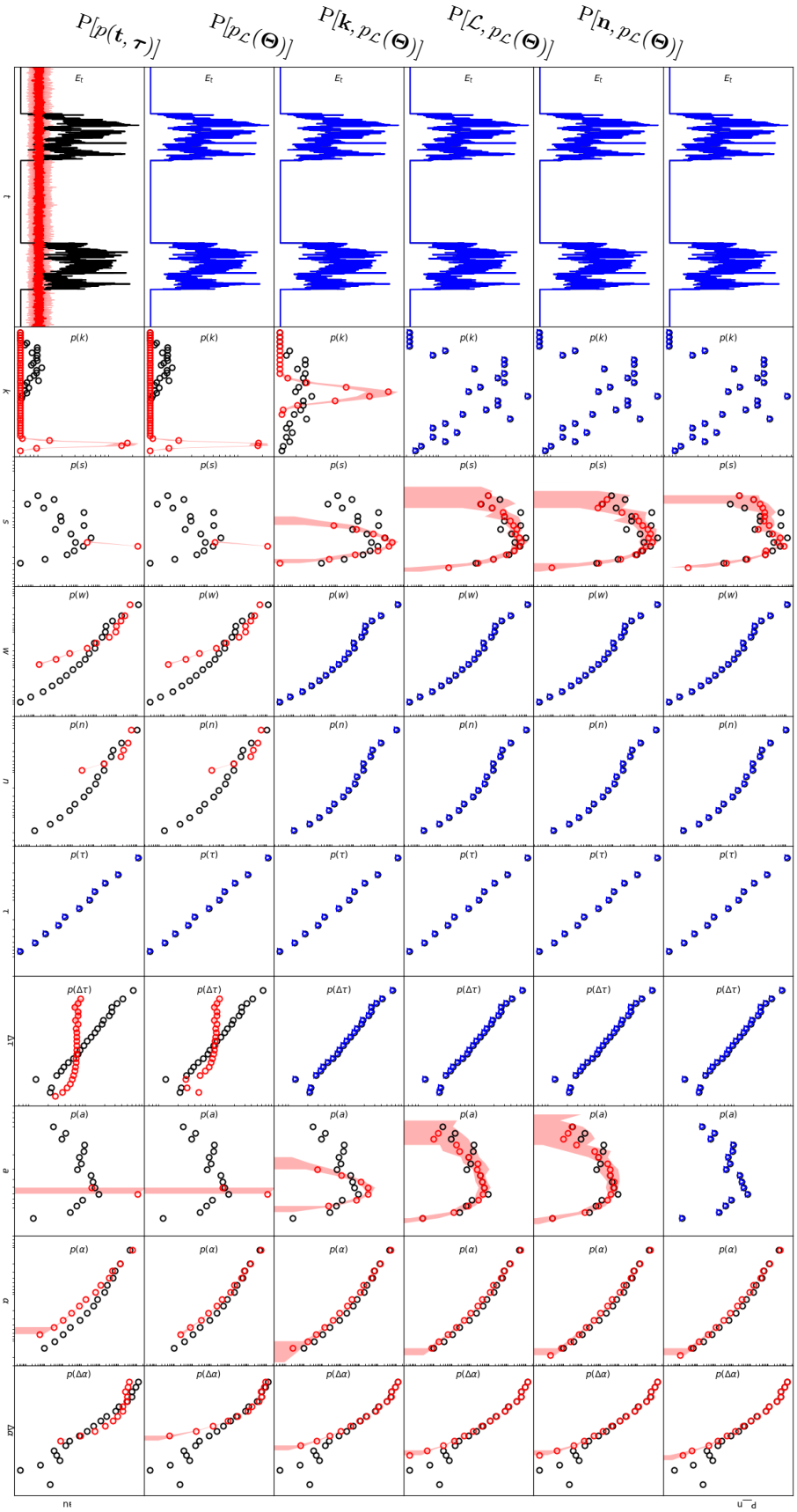
✓ Feature completely conserved.
^{p} Distribution (i.e. the multiset) of individual values in sequence conserved.
$^{\mu}$ Mean value of the individual features in sequence conserved.
✗ Feature not conserved.

butions of the nodes' activity and inactivity durations, $p(\boldsymbol{\alpha}) = \cup_{i\in\mathcal{V}}[\alpha_i^m]_{m\in\mathcal{M}_i}$ and $p(\boldsymbol{\Delta\alpha}) = \cup_{i\in\mathcal{V}}[\Delta\alpha_i^m]_{m\in\mathcal{M}_i}$, respectively. However, all link shufflings produce null distributions that are relatively close to the empirical ones, though they are still statistically significantly different. This indicates that the temporal correlations of the individual links' activity strongly constrain the nodes' activity. More surprisingly, the small divergence observed in $p(\boldsymbol{\alpha})$ for P[$\mathcal{L}, p(\mathbf{t},\boldsymbol{\tau})$] and P[$p(\mathbf{t},\boldsymbol{\tau})$] as compared to the other MRRMs points to the global timing of the events as the most important of the features in determining the node activity durations in the network. It is more important than the number of events and the distributions of inter-event durations on the links. Conversely, we see that the distribution of inter-event durations, $\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$, is the most important temporal feature in determining the nodes' inactivity durations, $p(\boldsymbol{\Delta\alpha})$, while the timing of the events is a close second (compare P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}), \mathbf{t}^1$] and P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})$] to P[$\mathcal{L}, p(\mathbf{t},\boldsymbol{\tau})$], and these three to P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \mathbf{t}^1, \mathbf{t}^w$] and P[$\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})$]).

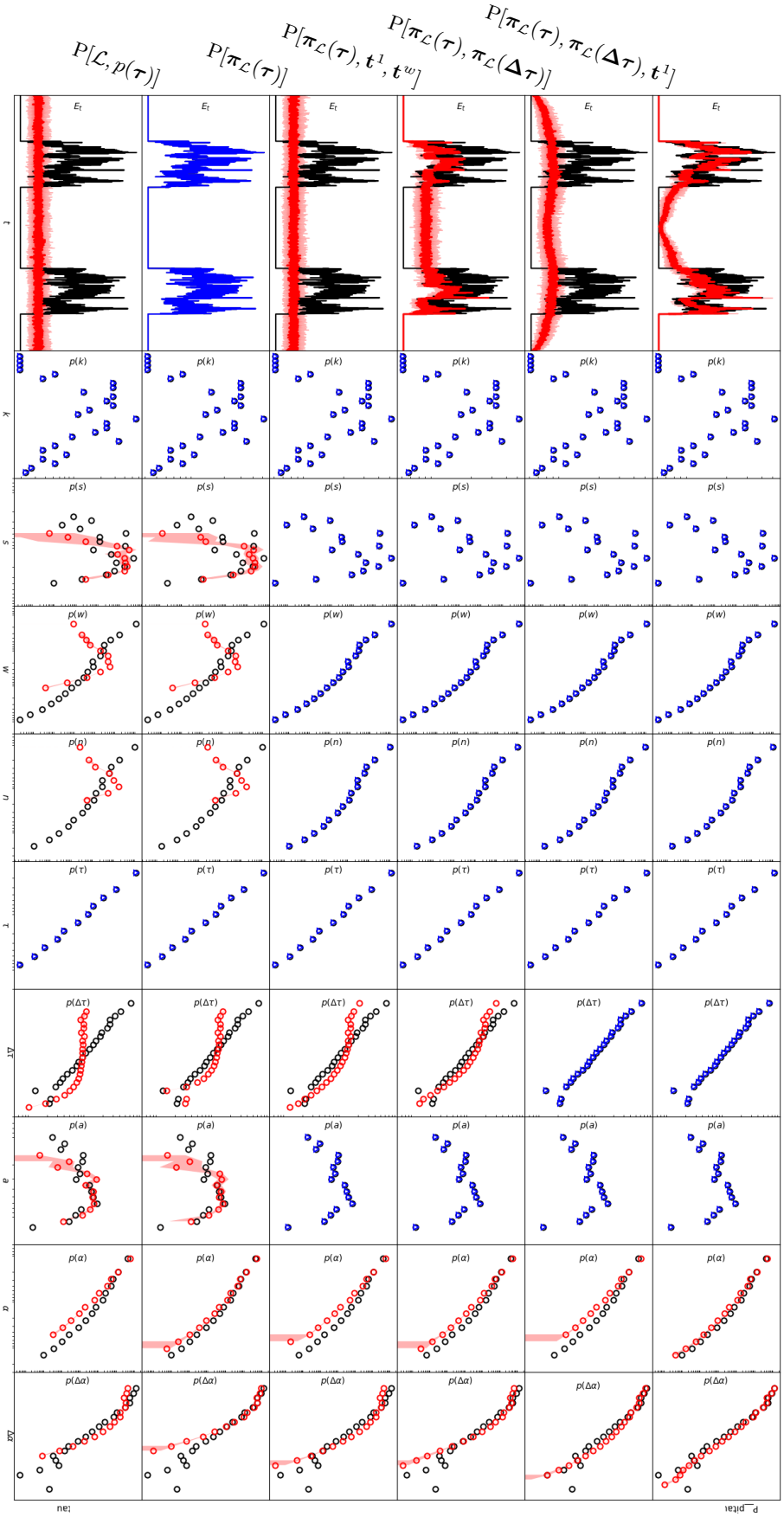
As seen in Supplementary Figs. S4 and S5, the distributions of the different features obtained from a single randomized network generally vary little around their median, even though the empirical network studied here is of relatively modest size.



Supplementary FIG. S3: **Effects of the MRRMs on different features of a temporal network of face-to-face interactions in a primary school.** Each panel shows the difference between the value of the feature in the empirical network and its null distribution under each model (red symbols for MRRMs that do not constrain the feature, and blue symbols for MRRMs that do) as well as the differences between its value in different randomized networks in each null model (black). The latter serves as a benchmark that shows the expected difference due to random fluctuations if a null model were true. For the activity timeline $\mathbf{A}$, the difference is quantified as the L1-distance between the activity at each time. For all other features, the difference is quantified as the Jensen-Shannon divergence (JSD) between the global distributions of the values of individual scalar features. Each box-and-whiskers summarizes the distribution of the differences over 100 randomized networks generated by the MRRM in question: boxes show the 1st and 3rd quartiles; whiskers extend to 1.5 times the interquartile range or to the minimum (maximum) value, whichever is smaller; values beyond the whiskers are marked by open circles.



Supplementary FIG. S4: **Effects of different link and event shufflings on temporal network features.** Values of a selection of features in the empirical face-to-face interaction network and in randomized networks generated from it. Original data is in black. Randomized data is in blue if constrained, in red if not. Red lines are medians over 100 randomizations, red areas show 90 % confidence intervals.



Supplementary FIG. S5: **Effects of different timeline shufflings on temporal network features.** Values of a selection of features in the empirical face-to-face interaction network and in randomized networks generated from it. Original data is in black. Randomized data is in blue if constrained, in red if not. Red lines are medians over 100 randomizations, red areas show 90 % confidence intervals.

Supplementary Note 4: Names of MRRM algorithms in the Python library

Instant-event shufflings

- P$[E]$: `P__1`

Timeline shufflings.

- P$[\mathcal{L}, E]$: `P__L`
- P$[\mathbf{w}]$: `P__w`
- P$[\mathbf{w}, \mathbf{t}^1\mathbf{t}^w]$: `P__w_t1_tw`
- P$[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})]$: `P__pidtau`
- P$[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}), \mathbf{t}^1\mathbf{t}^w]$: `P__pidtau_t1_tw`

Sequence shufflings.

- P$[p_{\mathcal{T}}(\boldsymbol{\Gamma})]$: `P__pGamma`
- P$[p_{\mathcal{T}}(\boldsymbol{\Gamma}), \chi_{\mathbb{N}^+}(\mathbf{A})]$: `P__pGamma_sgnA`

Snapshot shufflings.

- P$[\mathbf{t}]$: `P__t`
- P$[\mathbf{t}, \boldsymbol{\Phi}]$: `P__t_Phi`
- P$[\mathbf{d}]$: `P__d`
- P$[\mathbf{iso}(\boldsymbol{\Gamma})]$: `P__isoGamma`
- P$[\mathbf{iso}(\boldsymbol{\Gamma}), \boldsymbol{\Phi}]$: `P__isoGamma_Phi`

Intersections.

- P$[\mathcal{L}, \mathbf{t}]$: `P__L_t`
- P$[\mathbf{w}, \mathbf{t}]$: `P__w_t`

Compositions.

- P$[L]$: `P__pTheta` with `P__L_E`
- P$[\mathbf{k}, p(\mathbf{w}), \mathbf{t}]$: `P__pTheta` with `P__w_t`
- P$[\mathbf{k}, \chi_\lambda, p(\mathbf{w}), \mathbf{t}]$: with `P__w_t`

Event shufflings

- P$[p(\boldsymbol{\tau})]$: `P__ptau`

Link shufflings.

- P$[p_{\mathcal{L}}(\boldsymbol{\Theta})]$: `P__pTheta`
- P$[\chi_\lambda, p_{\mathcal{L}}(\boldsymbol{\Theta})]$: `P__I_pTheta`
- P$[\mathbf{k}, p_{\mathcal{L}}(\boldsymbol{\Theta})]$: `P__k_pTheta`
- P$[\mathbf{k}, \chi_\lambda, p_{\mathcal{L}}(\boldsymbol{\Theta})]$: `P__k_I_pTheta`

Timeline shufflings.

- P$[\mathcal{L}, p(\boldsymbol{\tau})]$: `P__L_ptau`
- P$[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau})]$: `P__pitau`
- P$[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \mathbf{t}^1\mathbf{t}^w]$: `P__pitau_t1_tw`
- P$[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau})]$: `P__pitau_pidtau`
- P$[\boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\tau}), \boldsymbol{\pi}_{\mathcal{L}}(\boldsymbol{\Delta\tau}), \mathbf{t}^1]$: `P__pitau_pidtau_t1`
- P$[\mathbf{per}(\boldsymbol{\Theta})]$: `P__perTheta`
- P$[\boldsymbol{\tau}, \boldsymbol{\Delta\tau}]$: `P__tau_dtau`

Snapshot shufflings.

- P$[p(\mathbf{t}, \boldsymbol{\tau})]$: `P__pttau`

Intersections.

- P$[\mathcal{L}, p(\mathbf{t}, \boldsymbol{\tau})]$: `P__L_pttau`
- P$[\mathbf{n}, p(\mathbf{t}, \boldsymbol{\tau})]$: `P__n_pttau`
- P$[\mathcal{L}, p_{\mathcal{L}}(\boldsymbol{\Theta})]$: `P__L_pTheta`
- P$[\mathbf{w}, p_{\mathcal{L}}(\boldsymbol{\Theta})]$: `P__w_pTheta`
- P$[\mathbf{n}, p_{\mathcal{L}}(\boldsymbol{\Theta})]$: `P__n_pTheta`

Metadata shufflings

Link shufflings.

- P$[p_{\mathcal{L}}(\boldsymbol{\Theta}), \boldsymbol{\sigma}, \boldsymbol{\Sigma}_{\mathcal{L}}]$: `P__pTheta_sigma_SigmaL`
- P$[\mathbf{k}, p_{\mathcal{L}}(\boldsymbol{\Theta}), \boldsymbol{\sigma}, \boldsymbol{\Sigma}_{\mathcal{L}}]$: `P__k_pTheta_sigma_SigmaL`
- P$[G, p(\boldsymbol{\sigma})]$: `P__G_psigma`

Compositions.

- P$[\mathbf{k}, p(\mathbf{w}), \mathbf{t}, \boldsymbol{\sigma}, \boldsymbol{\Sigma}_{\mathcal{L}}]$: `P__k_LCM` with `P__w_t`