Temporal networks with node-specific memory: Unbiased inference of transition probabilities, relaxation times, and structural breaks

Giulio Virginio Clemente,[1,*] Claudio J. Tessone [2] and Diego Garlaschelli [1,3,4]

[1]*IMT School for Advanced Studies, Piazza San Francesco 19, 55100 Lucca, Italy*

[2]*Blockchain & Distributed Ledger Technologies, UZH Blockchain Center, University of Zurich, Andreasstrasse 15, 8050 Zurich, Switzerland*

[3]*Lorentz Institute for Theoretical Physics, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands*

[4]*INdAM-GNAMPA Istituto Nazionale di Alta Matematica, 00185 Rome, Italy*

One of the main challenges in the study of time-varying networks is the interplay of memory effects with structural heterogeneity. In particular, different nodes and dyads can have very different statistical properties in terms of both link formation and link persistence, leading to a superposition of typical timescales, suboptimal parametrizations, and substantial estimation biases. Here we develop an unbiased maximum-entropy framework to study empirical network trajectories by controlling for the observed structural heterogeneity and local link persistence simultaneously. An exact mapping to a heterogeneous version of the one-dimensional Ising model leads to an analytic solution that rigorously disentangles the hidden variables that jointly determine both static and temporal properties. Additionally, model selection via likelihood maximization identifies the most parsimonious structural level (either global, node specific, or dyadic) accounting for memory effects. As we illustrate on real-world social networks, this method enables an improved estimation of dyadic transition probabilities, relaxation times, and structural breaks between dynamical regimes. In the resulting picture, the graph follows a generalized configuration model with given degrees and given time-persisting degrees, undergoing transitions between empirically identifiable stationary regimes.

I. INTRODUCTION

Time-varying graphs have been studied intensively due to their relevance for many systems and processes. The temporal variability of networks is critical in several diverse fields for multiple reasons. Examples include (a) the study of communication in highly dynamic networks, e.g., broadcasting and routing in delay-tolerant networks [1]; (b) the exploitation of passive mobility, e.g., the opportunistic use of transportation networks [2]; and (c) the analysis of complex social systems, e.g., the characterization of the dynamic interaction patterns emerging in a social network [3–5].

Different ways to represent the data in such dynamic settings are proposed [6]. Among these, the snapshots' representation deserves special mention as the primary approach used in this work. This method aggregates temporal information into nonoverlapping fixed time windows, thereby creating an ordered sequence of static networks. Even if this representation can sometimes be misleading [7], mainly for reasons linked to the time windows' choice, its usage can bring

different advantages [8]. Snapshot representation is easy to interpret and gives a natural setup to readapt some models widely used for static cases.

Often, the transition of models from static to dynamic cases involves incorporating "memory" terms, typically of a Markovian nature [9]. For instance, exponential random graph models (ERGMs) [10] are adapted to dynamic contexts by adding mechanisms that account for temporal evolution. Originating with the first descriptions of dynamic networks using a Markovian model [11], Snijders [12] proposed an evolutionary mechanism viewed as a continuous process, where ministeps represent changes between two consecutive snapshots. Alternatively, ERGMs can be adapted using a discrete mechanism, as in the temporal exponential random graph models (TERGMs) [13], which do not presume any specific sequence for the creation or disruption of links between snapshots. Instead, links are generated simultaneously in each snapshot based solely on the temporal order of the network, aiming to replicate the model's imposed constraints. It is within this latter framework that we position our contribution.

Initially, we define a quantity that captures the temporal correlation between different snapshots: the persisting connectivity, an indicator of the stability of connections between nodes. Due to its intuitive interpretation, link stability is a commonly used measure in the literature. For instance, the concept is implemented in "edge-Markovian evolving graphs" [14], where the stability of links is observed to study memory effects. Attention to this mechanism is also underscored in

*Contact author: giulio.clemente@imtlucca.it

[15], which proposes a methodology to describe link persistence and the dynamics of latent variables associated with each node that contribute to link formation. Latent variables are also employed by Zhang *et al.* [16], where the connection between nodes is modeled as a continuous-time Markov process, with a characteristic rate influenced by the nodes' latent variables. Drawing on characteristics observed in the literature, we propose our models. We introduce various maximum-entropy models, each sharing the same functional form but designed with different levels of heterogeneity in the constraints used. These models are tailored for stationary time networks. We begin with three memoryless models, each employing different constraints: the first uses a time-averaged constraint for each link, the second employs a time-averaged constraint for each node's degree, and the third implements a constraint for the overall average degree, also averaged over time. We enhance these models by adding memory-based constraints to one of the memoryless frameworks, which previously only constrained average degrees. These new specifications include the time-averaged persisting connectivity, the time-averaged persisting degree, defined as the amount of a degree conserved from one snapshot to the next, and the average persisting degree across nodes, also averaged over time. After a suitable reparametrization, the resulting models show an analogy with the one-dimensional Ising model. Exploiting this analogy, we analytically solved the models, enabling the use of the maximum likelihood approach to estimate the models' parameters and to precisely compute the expected values of several key quantities of interest.

The proposed models have the property of being applicable only to study temporal networks that present a stationary evolution. We overcome this limitation by proposing a structural break detection method that leverages these models to identify stationary periods. Thanks to this solution, we can also capture changing points, often corresponding to exogenous events.

We tested the proposed approaches using two proximity data sets: the MIT Proximity Network (2004–2005) [17], which includes a documented list of external events, and the Primary School Proximity Network [18]. For the latter, we assumed the presence of external events based on the school's scheduled activities. To assess the degree of heterogeneity in the models, we employed the Akaike information criterion (AIC) [19]. Initially, we evaluated the memoryless models and found that the model with constraints on degrees performed best according to the AIC. We then examined the memory model and discovered that the model incorporating constraints on both the degree and persisting degree outperformed the others. This finding indicates that the system possesses memory, which can be explained by latent variables associated with each node. Additionally, our structural break approach effectively identified many real-world events that, according to the algorithm, induced structural changes in the network's evolution. Ultimately, we characterized each node by leveraging the Markovian process involved in link formation.

II. MAXIMUM-ENTROPY PROBABILITY FOR COMPLEX NETWORKS

Before addressing temporal networks, we must revisit some foundational theories underpinning our models for time-varying graphs. We introduce the maximum-entropy approach to complex networks as a powerful tool for modeling complex systems [20]. Given a graph G with a number of nodes N and specific properties $C(G)$, the maximum-entropy approach provides a method to derive the most unbiased probability distribution $P(G)$ [21]. This distribution encompasses all possible graphs of the same type, ensuring that $C(G)$ is accurately reproduced on average.

To obtain this probability, we go through a constrained entropy maximization [22], which involves finding the maximum of the Shannon-Gibbs entropy:

$$S[P] = - \sum_G P(G) \ln P(G), \qquad (1)$$

granting the constraints that the probability distribution has to reproduce on average

$$\langle C \rangle = \sum_G C(G) P(G) = C^*, \qquad (2)$$

where G is a generic network in the ensemble, C^* is the observed value of the constraint, and $\langle \cdot \rangle$ indicates the expected value over the ensemble of networks. The constrained maximization problem is solved by introducing a Lagrange multiplier (θ) for each constraint. The solution results in a probability distribution that turns out to have the same functional form of the exponential random graphs models (ERGMs) [23]:

$$P(G|\vec{\theta}) = \frac{e^{-H(G|\vec{\theta})}}{Z(\vec{\theta})}, \qquad (3)$$

where $H(G|\vec{\theta}) = \sum_i \theta_i C_i$ is the graph Hamiltonian, and $Z(\vec{\theta}) = \sum_G e^{-H(G|\vec{\theta})}$ is the normalization constant, also called partition function.

A critical assumption that is often made when we work with these models is the independence of link formation, such that we are allowed to write the probability for the formation of the entire graph as

$$P(G) \equiv \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}}, \qquad (4)$$

where a_{ij} is the i, j entry of the adjacency matrix of the graph G, and p_{ij} represents the probability that $a_{ij} = 1$ (nodes i and j are connected).

A relevant example for this paper is given by the binary configuration model (CM) [23], where the resulting probability distribution is obtained after constraining the degree sequence, such that it admits a Hamiltonian representation

$$H(G|\vec{\theta}) \equiv \sum_{i=1}^{N} \theta_i k_i, \qquad (5)$$

where k_i is the degree of node i. For this problem, the p_{ij} is [20,24,25]

$$p_{ij} \equiv \frac{x_i x_j}{1 + x_i x_j}, \qquad (6)$$

where $x_i \equiv e^{-\theta_i}$ for each node i [20,25]. x_i is often called the "fitness" of node i and can be interpreted as the propensity of node i to link with other nodes.

A. Maximum likelihood principle

When applying the model to a real-world graph G^*, it is necessary to introduce a criterion to select the values of the parameters $\vec{x}$. According to the maximum likelihood principle, these parameters should be set to values that maximize the likelihood of the observed data under the given model [25]. Specifically, we seek the parameter values $\vec{x}^*$ that maximize the likelihood function $P(G^*|\vec{x})$. For the CM, $\vec{x}^*$ can be determined by solving N coupled equations [20,25]:

$$\langle k_i \rangle = \sum_{j \neq i} \frac{x_i x_j}{1 + x_i x_j} = k_i^* \quad \forall \, i, \tag{7}$$

where $k_i^* \equiv k_i(G^*)$ represents the empirical degree of node i in G^*. In other words, the maximum likelihood principle shows that one needs to set $\vec{x}$ to the particular value $\vec{x}^*$ (or, equivalently, to set $\vec{\theta}$ to the particular value $\vec{\theta}^*$) ensuring that each expected degree $\langle k_i \rangle = \sum_{j \neq i} p_{ij}$ matches the corresponding observed degree k_i^*.

Interestingly, this model has been shown to effectively reproduce the empirical topology of various real-world networks [20]. In the case of the world trade web (WTW), which represents the network of international trade among countries, the fitness parameter x_i^* has been related with an empirical property, specifically the gross domestic product (GDP) of each country i [24,25]. Given its theoretical and practical advantages, we will use the CM Eq. (6) as a starting point for the extension of the formalism to the temporal case.

Generally, including the CM, if we require that a network model generates a desired set $\vec{C} \equiv \{C_1, \ldots, C_K\}$ of topological properties, the maximization of (1) leads to Eq. (3) where $H(G) = \sum_{k=1}^{K} \theta_k C_k$ [20,23]. If we apply the maximum likelihood principle, then we find that the value $\vec{\theta}^*$ maximizing the likelihood of the model [25] is also the value ensuring that the expected value $\langle C_k \rangle$ of each property C_k matches the corresponding empirical value $C_k^* \equiv C_k(G^*)$. Thus, $\vec{\theta}^*$ can be found as the solution of the K coupled equations $\langle C_k \rangle = C_k^*$ for all k.

III. FROM STATIC TO TIME-VARYING GRAPHS

There are different ways to move from a model for static graphs to one for time-varying graphs, mainly depending on the representation and on the assumption made on the system [6]. Our purpose is to create a model that generates a sequence of graphs with fixed number of nodes and with properties that can fluctuate in time but are on average constant.

A. Extensions to time-varying graphs

In this section, we discuss possible ways to create a model capable of generating a sequence of T temporal snapshots of a network, which we refer to as the "graph trajectory":

$$\mathcal{G} \equiv \{G_1, \ldots, G_T\}. \tag{8}$$

We assume the number of nodes N remains constant throughout the graph's evolution. Additionally, each node is assigned a quantity x_i, analogous to the "fitness" in Eq. (7).

The simplest option is that of regarding the observed graph trajectory $\mathcal{G}$ as a sequence of independent realizations of the same process, the latter being still specified by the static probabilities p_{ij} where $\vec{x}$ is time independent. This assumption implies that $\vec{x}$ remains fixed, and the probability of the entire graph trajectory factorizes over all time steps as

$$\mathcal{P}(\mathcal{G}|\vec{x}) = \prod_{t=1}^{T} P(G_t|\vec{x}) = \prod_{t=1}^{T} \prod_{i<j} p_{ij}^{a_{ij}(t)} (1 - p_{ij})^{1-a_{ij}(t)}. \tag{9}$$

This means that we are not introducing any explicit statistical dependence among different temporal snapshots of the network. So, while the dependence of all snapshots on the same fitness vector $\vec{x}$ implies that the graph trajectory will have a certain degree of stationarity (especially for the pairs of nodes with p_{ij} close to either 0 or 1), the conditional independence (given $\vec{x}$) of different snapshots implies that there are no temporal correlations among such snapshots.

B. Dynamic fitness, no temporal dependencies

Another approach to incorporating temporal dynamics involves making $\vec{x}$ time dependent, i.e., replacing x_i with $x_i(t)$, and accordingly, p_{ij} with $p_{ij}(t)$, while retaining the same functional form for the probability. This modification allows each snapshot of the system to be described by a distinct probability, yet they remain statistically independent in principle. The probability of the entire graph trajectory $\mathcal{G}$ still factorizes as follows:

$$\mathcal{P}(\mathcal{G}|\vec{x}) = \prod_{t=1}^{T} P(G_t|\vec{x}(t))$$
$$= \prod_{t=1}^{T} \prod_{i<j} p_{ij}(t)^{a_{ij}(t)} [1 - p_{ij}(t)]^{1-a_{ij}(t)}, \tag{10}$$

where $\vec{x}$ is represented as a $T \times N$ matrix, with the generic entry $x_i(t)$ located in the tth row and ith column. Assuming that the fitness $x_i(t)$ is a dynamic variable suggests that node fitness and network topology evolve over comparable timescales. Conversely, treating x_i as fixed implies that node fitness changes more slowly than the network topology. A significant contribution to this approach is by Sarkar and Moore [26], who introduced a latent distance model with Markovian properties.

C. Static fitness, temporal dependencies

A third possibility is to assume that different snapshots are statistically dependent, representing a fundamental change to the model as it no longer allows different snapshots to be generated by a single function p_{ij}. In the simplest case, we might assume that $\vec{x}$ is time independent, as in Eq. (9). However, the probability of the graph trajectory in this context does not factorize over time steps:

$$\mathcal{P}(\mathcal{G}|\vec{x}) \neq \prod_{t=1}^{T} P(G_t|\vec{x}). \tag{11}$$

Significant contributions in this area include the works of Hanneke (TERGMs) [13] and Zhang *et al.* [16], wherein both the structural dynamics are assumed to have a Markovian nature while the parameters remain constant. Alternatively,

we may consider $\vec{x}$ to be time dependent as in Eq. (10), as proposed by Mazzarisi *et al.* [15], though the probability still does not factorize.

However, for the purpose of this paper, we will focus on scenarios where $\vec{x}$ is fixed. For the class of models described in Eq. (11), the temporal dependency structure among different snapshots must be explicitly specified. Generally, this implies that the connection probability at time t depends not only on t but also on previous times $t' < t$. Despite this complexity, we aim to preserve the core feature of the fitness model, where the connection between nodes i and j depends solely on their individual properties. Thus, while $\mathcal{P}(\mathcal{G}|\vec{x})$ does not factorize over time steps, it still factorizes over pairs of vertices.

Our primary goal is to develop and solve a model under these conditions. As a last remark, we underline that if combined with suitable techniques, as discussed in Sec. VI, this class of methods can effectively handle graph trajectories generated with time-varying parameters that present periods of stationarity.

D. Measures of time correlation: Persisting connectivity

Before we provide an explicit model with time dependence, we introduce some useful definitions to capture time correlations. In models with no memory, such as those defined by Eqs. (9) and (10), the conditional (given $\vec{x}$ or **x**) independence of different snapshots implies

$$\langle a_{ij}(t)a_{ij}(t+\tau)\rangle_\nu = \langle a_{ij}(t)\rangle_\nu \langle a_{ij}(t+\tau)\rangle_\nu \quad \forall\ t, \tau > 0, \tag{12}$$

where the subscript ν stands for "no memory."

The above observation suggests that a useful quantity to introduce at the level of nodes is the *persisting degree* $h_i(t, \tau)$, that we define as the part of the degree $k_i(t)$ that also persists at time $t + \tau$m [or sum of links that persists $h_{ij}(t, \tau)$]:

$$h_i(t, \tau) \equiv \sum_{j\neq i} a_{ij}(t)a_{ij}(t+\tau) = \sum_{j\neq i} h_{ij}(t, \tau). \tag{13}$$

The persisting degree is computed by summing the quantity $a_{ij}(t)a_{ij}(t+\tau)$, to which we will refer as *persisting link*, a concept also employed in studies such as [14–16].

Note that the value of $a_{ij}(t')$ at intermediate times ($t < t' < \tau$) has no effect on the above definition. From Eq. (12), the expected value of $h_i(t, \tau)$ under a model with no memory is always

$$\langle h_i(t, \tau)\rangle_\nu = \sum_{j\neq i} \langle a_{ij}(t)\rangle_\nu \langle a_{ij}(t+\tau)\rangle_\nu \quad \forall\ t, \tau > 0. \tag{14}$$

Comparing the observed and expected (under a model with no memory) value of $h_i(t, \tau)$ is a useful criterion to measure time correlations in an observed graph trajectory. If we average over time, the quantity

$$A_i(\tau) \equiv \frac{1}{T}\sum_{t=1}^{T}[h_i(t, \tau) - \langle h_i(t, \tau)\rangle_\nu] \tag{15}$$

is a node-specific autocovariance function, which allows us to measure the timescale of the decay of memory for each node. If the measured autocovariance is zero for all nodes, then there is no need to introduce a model with memory since

a model described by Eqs. (9) or (10) will be enough. By contrast, nonzero autocovariance can only be modeled using time dependencies of some form, thus leading to Eq. (11).

For static cases defined by Eq. (9), the expressions become

$$\langle a_{ij}(t)a_{ij}(t+\tau)\rangle_\nu = p_{ij}^2 \quad \forall\ t,\ \tau > 0, \tag{16}$$

$$\langle h_i(t, \tau)\rangle_\nu = \sum_{j\neq i} p_{ij}^2 \quad \forall\ t,\ \tau > 0. \tag{17}$$

These expressions will be useful in the following.

IV. A SOLVABLE MODEL WITH MEMORY

We now consider an explicit, solvable model that incorporates time correlations. As previously discussed, we assume that $\vec{x}$ is fixed. Our model employs the maximum-entropy formalism outlined earlier, which has been readapted to describe ensembles of network trajectories. We define the entropy for these network trajectories as follows:

$$S[\mathcal{P}] \equiv -\sum_{\mathcal{G}} \mathcal{P}(\mathcal{G})\ln\mathcal{P}(\mathcal{G}). \tag{18}$$

By maximizing this entropy [Eq. (18)] subject to the constraint that certain properties $C_i(\mathcal{G})$ observed in the trajectory are replicated on average, we obtain the solution

$$P(\mathcal{G}|\vec{\theta}) = \frac{e^{-\sum_i \theta_i C_i(\mathcal{G})}}{Z(\vec{\theta})}, \tag{19}$$

where the partition function is given by

$$Z(\vec{\theta}) = \sum_{\mathcal{G}} e^{-\sum_i \theta_i C_i(\mathcal{G})}. \tag{20}$$

Using the concept of persisting connectivity, we define a specific type of maximum-entropy model for temporal networks that serves as a natural generalization of the configuration model (CM). Interestingly, TERGMs can be exactly derived following the procedure described above. See Appendix H for a detailed derivation.

A. Preliminary extension: Memoryless model

As a preliminary step, we first consider three memoryless models, each defined by using a different type of constraint aimed at reproducing varying degrees of heterogeneity among nodes.

The first model precisely replicates the average of each link. The corresponding Hamiltonian for this model is given by

$$\mathcal{H}_{1,\nu}(\mathcal{G}) \equiv \frac{1}{T}\sum_{t=1}^{T} H_1(G_t)$$

$$= \frac{1}{T}\sum_{t=1}^{T}\sum_{j<i} \alpha_{ij}a_{ij}(t) = \sum_{j<i} \alpha_{ij}\overline{a_{ij}}. \tag{21}$$

This specification leads to a nontrivial model because, unlike in the static case, here we constrain the time-averaged links. This model is nondegenerative, as the Lagrange multipliers can assume values other than 0 or ∞, which typically occur in the static binary case.

The second kind of constraint is similar to the one for the static CM, with the distinction that degrees are averaged over time. This results in the following Hamiltonian:

$$\mathcal{H}_{2,\nu}(\mathcal{G}) \equiv \frac{1}{T} \sum_{t=1}^{T} H_2(G_t)$$

$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \alpha_i k_i(t) = \sum_{i=1}^{N} \alpha_i \bar{k}_i. \qquad (22)$$

The last kind of memoryless model is built taking as constraint the average over time and over nodes of the degree. This turns out to have a structure analogous to the Erdős-Rényi model [27], having the following Hamiltonian:

$$\mathcal{H}_{3,\nu}(\mathcal{G}) \equiv \frac{1}{T} \sum_{t=1}^{T} H_3(G_t)$$

$$= \frac{1}{TN} \sum_{t=1}^{T} \sum_{i=1}^{N} \alpha k_i(t) = \frac{\alpha}{N} \sum_{i=1}^{N} \bar{k}_i. \qquad (23)$$

By using expressions of each Hamiltonian above mentioned in Eq. (19), these models maximize the entropy of the entire ensemble of graph trajectories, as defined in Eq. (18), under the constraints specified.

When applying the maximum likelihood principle to each of these models, we search for the specific values of $\vec{\alpha}^*$ that maximize the likelihood $\mathcal{P}(\mathcal{G}^*|\vec{\alpha})$ of an observed graph trajectory $\mathcal{G}^*$. For all three specifications, we find that the probability of a link existing between two nodes in each snapshot shares the same form, as given by the following equation:

$$p_{ij} = \frac{x_{ij}}{1 + x_{ij}}, \qquad (24)$$

where

$$x_{ij} = \begin{cases} x_{ij} = e^{-\alpha_{ij}/T} & \text{if } \mathcal{H}_{1,\nu}(\mathcal{G}), \\ x_i x_j = e^{-\alpha_i/T} e^{-\alpha_j/T} & \text{if } \mathcal{H}_{2,\nu}(\mathcal{G}), \\ x = e^{-\alpha/TN} & \text{if } \mathcal{H}_{3,\nu}(\mathcal{G}). \end{cases} \qquad (25)$$

In Appendix A, we detail the method for determining $\vec{\alpha}^*$ for the three specifications.

While these kind of models can replicate any of the three observed time-averaged quantities we constrain, they predict no time correlations. Using Eq. (24), we find that for the three specifications, the expected value of the persisting degree is

$$\langle h_i(t, \tau) \rangle = \sum_{j \neq i} \left(\frac{x_{ij}}{1 + x_{ij}} \right)^2 \quad \forall t, \tau > 0. \qquad (26)$$

Comparing the predicted values from Eq. (26) with the observed values of $h_i(t, \tau)$ provides a straightforward test for the adequacy of our models. Specifically, the expected value of the node-specific autocovariance $A_i(\tau)$, as defined in Eq. (15), for the memoryless cases is

$$\langle A_i(\tau) \rangle = \frac{1}{T} \sum_{t=1}^{T} [\langle h_i(t, \tau) \rangle - \langle h_i(t, \tau) \rangle_\nu] = 0. \qquad (27)$$

In Fig. 1, we show the normalized autocovariance function for the degree for various nodes for one of the real-world data sets [17] used in the empirical application of the model.
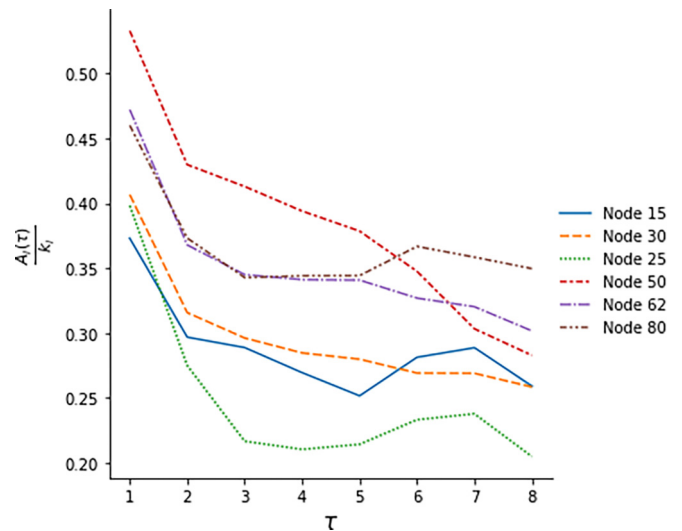

FIG. 1. The plot displays the normalized empirical autocovariance [Eq. (15)] as a function of τ for various nodes of the MIT proximity data set: the outcome is a nonzero decreasing function.

From the plot in Fig. 1, we observe that the node-specific autocovariance varies for each node and is distinctly nonzero. This variation provides initial evidence of heterogeneity in the memory characteristics associated with each node, underscoring the need for a method that can effectively capture this diversity. These empirical observations allow us to seize the opportunity and introduce our main contribution.

B. Full model

Now, we introduce temporal correlations at varying degrees of heterogeneity into our models. In these models, the expected persisting degree is nontrivial, and the autocovariance is nonzero. As we did with the memoryless cases, we define three sets of constraints integrated with the Hamiltonian from Eq. (22), resulting in three distinct models:

(i) *One-lagged persisting link*. The most constrained model involves constraining the one-lagged persisting link, given by $\overline{a_{ij}(t)a_{ij}(t+1)}$.

(ii) *One-lagged persisting degree*. The second model imposes a constraint on the one-lagged persisting degree $\overline{h_i(t, 1)}$.

(iii) *Average one-lagged persisting degree*. The third possibility is to constrain the average of the one-lagged persisting degrees across all nodes.

As a result, the empirical value of $h_i(t, 1)$, as well as that of the one-lagged autocovariance $A_i(1)$, will be replicated exactly for the first two models, while on average for the third one.

In this way, the three Hamiltonians to be added to Eq. (22) are, in order, as follows:

$$\mathcal{H}_\mu(\mathcal{G}) = \begin{cases} \mathcal{H}_{\mu,1}(\mathcal{G}) = \frac{1}{T} \sum_{j<i} \sum_t^T \beta_{ij} a_{ij}(t) a_{ij}(t+1), \\ \mathcal{H}_{\mu,2}(\mathcal{G}) = \frac{1}{T} \sum_i^N \sum_t^T \beta_i h_i(1, t), \\ \mathcal{H}_{\mu,3}(\mathcal{G}) = \frac{1}{TN} \sum_i^N \sum_t^T \beta h_i(1, t), \end{cases}$$

$$\qquad (28)$$

where the subscription μ stands for memory.

To analytically solve these models and derive tractable expressions for certain quantities of interest, we simplify by assuming periodic boundary conditions. Specifically, we introduce a fictitious time step $T + 1$ such that

$$a_{ij}(T + 1) \equiv a_{ij}(1). \tag{29}$$

As we will clarify, aside from the stationarity discussed in Sec. VI, this assumption is the only significant approximation made to adapt this model to real-world scenarios. Furthermore, the impact of this artifact diminishes as T increases.

We define our models through the following extension of Eq. (22):

$$\mathcal{H}(\mathcal{G}) \equiv \mathcal{H}_{2,\nu}(\mathcal{G}) + \mathcal{H}_{\mu}(\mathcal{G}). \tag{30}$$

This expression varies depending on the chosen constraints. As demonstrated in Appendices B, C, D, and E, after appropriate reparametrization, these models can be mapped exactly to a superposition of noninteracting one-dimensional Ising models. They can then be analytically solved, and their solutions expressed in terms of a constant connection probability:

$$p_{ij} = \langle \overline{a_{ij}(t)} \rangle, \tag{31}$$

and a "memory" function, which captures the persisting interactions over time:

$$q_{ij}(\tau) = \langle \overline{a_{ij}(t)a_{ij}(t + \tau)} \rangle. \tag{32}$$

The probability p_{ij} mirrors the concept of spin up and down in a one-dimensional Ising model, representing here the likelihood of a connection's presence. At the same time, q_{ij} can be seen as the probability of observing two spins at a distance τ in the same direction.

Explicitly,

$$p_{ij} = \left(\frac{x_i x_j y_{ij} - 1}{2\sqrt{4x_i x_j + (x_i x_j y_{ij} - 1)^2}} \right) \left(\frac{(\lambda_{ij}^+)^T - (\lambda_{ij}^-)^T}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T} \right) + \frac{1}{2}, \tag{33}$$

$$q_{ij}(\tau) = p_{ij}^2 + p_{ij}(1 - p_{ij}) \left(\frac{(\lambda_{ij}^-)^\tau(\lambda_{ij}^+)^{T-\tau} + (\lambda_{ij}^+)^\tau(\lambda_{ij}^-)^{T-\tau}}{(\lambda_{ij}^+)^\tau + (\lambda_{ij}^-)^\tau} \right), \tag{34}$$

with

$$\lambda_{ij}^{\pm} = e^{J_{ij}} \cosh B_{ij} \pm \sqrt{e^{2J_{ij}} \sinh^2 B_{ij} + e^{-2J_{ij}}}, \tag{35}$$

and

$$B_{ij} \equiv \tfrac{1}{2}[\ln(x_i) + \ln(x_j) + \ln(y_{ij})], \tag{36}$$

$$J_{ij} \equiv \tfrac{1}{4}[\ln(y_{ij})]. \tag{37}$$

Here, $x_i = e^{-\alpha_i/T}$, while y_{ij} assumes different expression depending on the constraints. Like x_{ij} for the different memoryless models, we have

$$y_{ij} = \begin{cases} y_{ij} = e^{-\beta_{ij}/T}, & \mathcal{H}_{\mu,1} \\ y_i y_j = e^{-\beta_i/T} e^{-\beta_j/T}, & \mathcal{H}_{\mu,2} \\ y = e^{-\beta/TN}, & \mathcal{H}_{\mu,3}. \end{cases} \tag{38}$$

The above expressions imply that the conditional probability that nodes i and j are connected at time $t + \tau$, given that they

were connected at time t, is

$$P[a_{ij}(t + \tau) = 1 | a_{ij}(t) = 1] = \frac{q_{ij}(\tau)}{p_{ij}}$$

$$= p_{ij} + (1 - p_{ij}) \left(\frac{\lambda_-^\tau \lambda_+^{T-\tau} + \lambda_+^\tau \lambda_-^{T-\tau}}{\lambda_+^T + \lambda_-^T} \right). \tag{39}$$

This shows how the models can generate nontrivial temporal dependencies. The expected node-specific autocovariance function is

$$\langle A_i(\tau) \rangle = \sum_{j \neq i} p_{ij}(1 - p_{ij})$$

$$\times \left(\frac{(\lambda_{ij}^-)^\tau(\lambda_{ij}^+)^{T-\tau} + (\lambda_{ij}^+)^\tau(\lambda_{ij}^-)^{T-\tau}}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T} \right). \tag{40}$$

V. TRANSITION MATRICES AND CORRELATION LENGTH

The introduced model can be utilized to generate graph trajectories with the prescribed properties. Unlike the static and memoryless cases, where p_{ij} alone suffices to generate the network trajectory, this model requires a combination of p_{ij} and q_{ij} to create a coherent time-varying graph.

Specifically, given the assumption of independent link formation, we need $\frac{N(N-1)}{2}$ transition matrices. Starting from an initial configuration, these matrices enable the generation of the entire graph trajectory. As detailed in Appendix G, in constructing these matrices, whose functional form is shown in the following equation, we utilize quantities estimated from our models:

$$P_{ij} = \begin{bmatrix} \frac{q_{ij}}{p_{ij}} & \frac{p_{ij} - q_{ij}}{p_{ij}} \\ \frac{p_{ij} - q_{ij}}{1 - p_{ij}} & \frac{1 - 2p_{ij} + q_{ij}}{1 - p_{ij}} \end{bmatrix}. \tag{41}$$

This matrix, as described in Eq. (41), retains the same form regardless of the model specification chosen; the variations are only in the expressions for p_{ij} and q_{ij}. Upon deriving the transition matrices, we employ them to characterize the stationary distribution between pairs of nodes and to calculate the average time required to reach this state. By construction, the largest eigenvalue of the stochastic matrix is one, and its associated eigenvector is $(p_{ij}, 1 - p_{ij})$, as detailed in Appendix G. This configuration ensures that the stationary connection probability precisely matches the marginal connection probability p_{ij}. Furthermore, the second largest eigenvalue μ_{ij} of the stochastic matrix equals

$$\mu_{ij} = \frac{q_{ij} - p_{ij}^2}{p_{ij} - p_{ij}^2}, \tag{42}$$

and it is indicative of the rapidity with which the system converges to the stationary distribution. Indeed, due to the mapping of the system to the one-dimensional Ising model, μ_{ij} turns out to be related to the correlation length τ_{ij}^c, which in this context acts as a correlation time. The relationship is expressed as follows:

$$\mu_{ij} = \left(\frac{\lambda_- \lambda_+^{T-1} + \lambda_+ \lambda_-^{T-1}}{\lambda_+^T + \lambda_-^T} \right) = e^{-\frac{1}{\tau_{ij}^c}}. \tag{43}$$

This implies

$$\tau_{ij}^c = \frac{1}{\ln\left(\frac{1}{\mu_{ij}}\right)}. \qquad (44)$$

The correlation time τ_{ij}^c tells us how strong the memory is, i.e., how fast the correlation vanishes as a function of the time lag between two snapshots. Clearly, different pairs of nodes have different values of τ_{ij}^c, so some pairs require more time to relax to their stationary connection probability p_{ij}. One may also analyze the neighborhood of each node i by considering the largest second eigenvalue $\mu_i \equiv \max_{j\neq i}\{\mu_{ij}\}$ among a maximum of $N-1$ pairs involving that node. This eigenvalue serves as a node-specific feature, defining the leading time $\tau_i^c \equiv \max_{j\neq i}\{\tau_{ij}^c\} = -1/\ln(\mu_i)$ required for the entire neighborhood of node i to reach the stationary distribution. Thus, even in systems that are already at equilibrium, understanding the correlation times can offer insights into how long the system is expected to take to return to equilibrium after a perturbation, assuming the model accurately reflects the system dynamics.

VI. STRUCTURAL BREAK DETECTION

To apply these models to real-world data sets, we must address a critical issue: real-world temporal networks often exhibit change points, indicating shifts in the statistical properties of the system. Such shifts render our earlier models, designed for trajectories with static parameters, inappropriate. Nonetheless, by carefully selecting a reasonable timescale and window, it is possible to identify segments where the system's statistical properties remain stationary, and so do the parameters used to describe them.

The challenge of detecting changing points or structural breaks has been addressed using various techniques, from model-based approaches [28–30] to model-free methods [31]. In this section, we introduce a model-dependent methodology to detect these structural breaks, adaptable to the different models we have discussed earlier.

Our approach utilizes a widely recognized technique in the field of detecting structural breaks known as binary segmentation [32]. This methodology relies on an appropriate cost function that measures the level of uniformity among observations. The selection of the correct cost function is crucial as it reflects the underlying mechanism we believe governs the temporal evolution and characterizes the structural breaks. In our case, we use a cost function based on the AIC [19]. By incorporating AIC, we consider not only the likelihood value but also the number of parameters, enhancing our model's robustness. The AIC is defined as follows:

$$\text{AIC}_{\text{model}} = 2K_{\text{model}} - 2\ln(L_{\text{model}}), \qquad (45)$$

where K_{model} corresponds to the number of parameters of the model, and L_{model} to the likelihood, computed by fitting the data.

Breaks are defined by solving the following optimization problem:

$$t_{\text{break}} = \underset{t\in[1,\dots,T-1]}{\arg\min}\ \Delta\text{AIC}(t), \qquad (46)$$

where $\Delta\text{AIC}(t)$ is our cost function and, given a temporal network of length T, is defined as follows:

$$\Delta\text{AIC}(t) = \text{AIC}_{\text{tot}} - \text{AIC}_{\text{diff}}, \qquad (47)$$

where

$$\text{AIC}_{\text{tot}} = 2K_{\text{model}} - 2\ln(L_{\text{model}}) \qquad (48)$$

and

$$\text{AIC}_{\text{diff}}(t) = 4k_{\text{model}} - 2\{\ln[L(\mathcal{G}_{0,t})] + \ln[L(\mathcal{G}_{t,T})]\}. \qquad (49)$$

A break time t_{break} is accepted only if $\Delta\text{AIC}(t_{\text{break}}) < 0$; otherwise, we assume there are no breaks. The underlying idea is to compare a model fitted to the entire temporal network, assuming uniform parameters, against a model that considers potential parameter changes at a specific point (t_{break}). The latter model doubles the number of parameters to better capture any structural changes in the system.

Once a structural break is identified, the new segment $[0, t_{\text{break}}]$ is isolated, and the procedure is repeated on this updated temporal network. This process continues until further division yields no significant benefit. The final break is designated as $t_{\text{break final}}$. Subsequently, the segment $[0, t_{\text{break final}}]$ is removed, focusing on the remaining segment $[t_{\text{break final}}, T]$. This method results in a collection of segments that, according to the model, are considered to be generated by models with distinct parameters, which remain approximately constant within each segment.

For our types of models, we can specify a general expression for the loglikelihood, given by

$$\ln(L_{\text{model}}) = -H - \ln(Z), \qquad (50)$$

where H represents the Hamiltonian of the system, and Z denotes the partition function. The specific values of these components are determined by the formulation of the Hamiltonian and the characteristics of the partition function used in the model. Consequently, their selection depends on the relevance of the constraints underlying the model designed to interpret the available data. The extent to which the identified points correspond to actual events or genuine changes in the data-generating process largely hinges on how accurately the employed model captures the underlying dynamics of the data.

A crucial aspect to emphasize is that the methodology we introduced presupposes that the identified segments are stationary. However, the assumption of segment stationarity is not always empirically verified. Specifically, if one of the identified segments is only as long as the minimum chosen length (e.g., three snapshots in our application), it may not be reasonable to consider that segment as stationary. Conversely, segments exceeding this minimum length are generally assumed to be stationary relative to the model used for their identification. This assumption implies that the parameters describing the data generation process are relatively constant within the identified segment, justifying their use to characterize that interval.

Finally, to validate the effectiveness of this methodology, we conducted a series of synthetic experiments. The results, detailed in Appendix I, demonstrate that when external shocks, interpreted as structural breaks, are sufficiently significant, our methodology successfully identifies them.
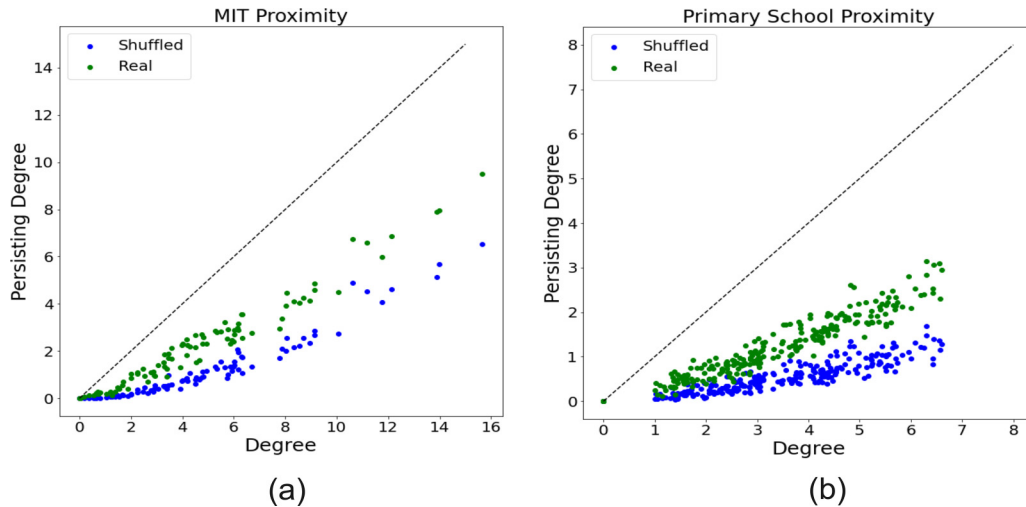
FIG. 2. Comparison of persisting degree at $\tau = 1$. This plot illustrates the persisting degree computed for $\tau = 1$ in two actual networks and compares these values with the average persisting degrees from 1000 shuffled instances of each network. The comparison highlights the differences between structured temporal interactions in real data sets and their randomized sequences, underscoring the impact of temporal order. (a) Shows results for the MIT Proximity Network, while (b) displays results for the Primary School Proximity Network.

VII. APPLICATION TO SOCIAL PROXIMITY NETWORKS

In this section, we demonstrate the capabilities of our approach by examining various aspects that our models can effectively characterize. Specifically, we aim to determine whether incorporating a one-step memory is sufficient to accurately reproduce the quantities $h_i(t, \tau)$ for different values of τ. Additionally, we seek to assess the level of heterogeneity in memory across different nodes. This analysis not only tests the validity of the one-step memory assumption but also enhances our understanding of memory dynamics within the network, offering insights into how individual nodes contribute to the overall temporal structure.

For our real-world application, we examine two proximity temporal networks derived from two different data sets. The first data set considers an experiment conducted by the Massachusetts Institute of Technology (MIT). Specifically, during the academic year 2004–2005, the MIT Media Laboratory conducted a reality mining experiment [17]. The experiment involved tracking 92 participants through their smart phones. Utilizing Bluetooth data, they obtained proximity measurements, which we interpret as connections between the individuals. By organizing the linkages on a daily basis and recording the time stamps of these proximity events, we constructed a daily empirical temporal network consisting of 244 undirected snapshots.

The second data set originates from a study conducted in a primary school, where proximity detectors worn by students and teachers captured interactions over a period of two days, specifically from October 1 to 2, 2009 [18]. For this analysis, we focus on data from the first day, where the linkages are collected on a 10-min time-window basis, resulting in snapshots representing all interactions within each interval. For this second temporal network, we have a total of 52 undirected snapshots, each of 242 nodes. By construction, both temporal networks have snapshots that are undirected and unweighted.

Before applying the models directly to the data sets, we conduct a preliminary analysis to evaluate the impact of the

temporal order of the snapshots on the network dynamics. To investigate this aspect, we generated 1000 instances of each temporal network by completely randomizing the order of the snapshots in every instance. For each randomized network, we then calculated the persisting degree for various values of τ.

Figure 2 demonstrates the significance of temporal order by comparing the persisting degree measured in the original network with its average from 1000 reshuffled instances. The persisting degree from the reshuffled networks consistently falls below that of the original, suggesting a nontrivial temporal structure in the data.

To quantify the significance of these observations, we compute the Z-score [33] for each node across three different values of τ. This Z-score measures how many standard deviations the original network's persisting degree is from the reshuffled networks' mean. Our analysis reveals that a significant majority of nodes exhibit persisting degrees markedly higher than expected under random temporal order.

Specifically, for $\tau = 1$, 2, and 3, approximately 97% of the total nodes display a Z-score greater than 1.96, for the MIT data set. In contrast, for the Primary School network, the proportion of nodes with significant Z-scores decreases from 97% at $\tau = 1$ to 57% at $\tau = 3$, suggesting a decay in temporal correlations as τ increases. This decay indicates a transition toward memoryless dynamics.

While the School Proximity data set exhibits behavior indicative of Markovian dynamics, making our models particularly applicable, the MIT Proximity data set presents a different scenario. Here, as τ increases, the number of nodes showing a significant positive divergence in their persisting degree, compared to values from randomization, remains constant. This preliminary analysis indicates significant deviations from Markovian dynamics, suggesting that the models introduced do not fully capture the MIT temporal network. It is crucial to recognize that this analysis does not account for possible structural breaks, which, if present, may conceal a behavior more closely aligned with our models. To explore this possibility, we employ the structural break detec-

TABLE I. The table shows the performance metrics for the three memoryless models applied to each data set. The model demonstrating the best performance for each data set is highlighted in bold.

Model	AIC$_{\text{total}}$	
	MIT	Primary School
$\mathcal{H}_{1,v}$	8624.4	850 304 908.75
$\mathcal{H}_{2,v}$	**1657.11**	4652.06
$\mathcal{H}_{3,v}$	1716.39	**4362.65**

tion techniques previously described to identify and adjust for potential changes in the network dynamics of both the systems analyzed.

Initially, we evaluated the three memoryless models to assess the heterogeneity in the link formation mechanisms. Following this, we investigated the effects of memory. During the process of detecting structural breaks, our objective is to identify a vector of $\mathcal{N}$ changing points:

$$T_v = [t_{\text{break}}(0), \dots, t_{\text{break}}(\mathcal{N})]. \tag{51}$$

Once those breaks are defined, we can compute the value of the AIC for each bounded period.

Where, for each period, AIC depends on the likelihood and on the number of parameters each model uses. For the three memoryless cases, we have that

$$K_{\text{model}} = \begin{cases} N(N-1)/2 & \text{for } \mathcal{H}_{1,v}, \\ N & \text{for } \mathcal{H}_{2,v}, \\ 1 & \text{for } \mathcal{H}_{3,v} \end{cases} \tag{52}$$

and the likelihood measures the ability to reproduce the average over time of the observed links.

To compare which of the models explain better the data, we compute the total AIC for the three models, defined as the summation of the AIC computed for each segment individuate:

$$\text{AIC}_{\text{total}} = \sum_{t_b \in T_v} \text{AIC}\big(\mathcal{G}_{t_{b-1}, t_b}\big), \tag{53}$$

where when $t_b = t_{\text{break}}(0)$, $t_{b-1} = 0$.

In the practical implementation of our algorithm for detecting structural breaks, we initiate the analysis with segments comprising a maximum of 50 data points. We then progressively move the window forward if no break point is identified within these segments. This approach is chosen to mitigate potential numerical issues that can arise from analyzing larger segments, such as computational inefficiencies.

From these three memoryless models, see Table I.

Analysis of the performance metrics in Table I reveals differing optimal models for the MIT and Primary School data sets. For the MIT data set, the model with node-specific constraints on degrees emerges as the winner, suggesting that connection patterns are better explained by latent variables unique to each node, rather than a uniform parameter for each link. This finding underscores the importance of considering individual node characteristics, which the AIC scores support by indicating that models using a single parameter oversimplify the network dynamics. This implies that while link formations

TABLE II. The table shows the performance metrics for the memoryless models with local constraints and the three models with memory. The model demonstrating the best performance for each data set is highlighted in bold.

Model	AIC$_{\text{total}}$	
	MIT	Primary School
$\mathcal{H}_{2,v}$	1 398 430.29	1 956 993.64
$\mathcal{H}_1$	336 021.35	625 824.76
$\mathcal{H}_2$	**268 522.79**	**169 403.17**
$\mathcal{H}_3$	272 456.55	170 927.21

are independent, links sharing common nodes might still reflect interconnected behaviors due to their shared origins.

Conversely, the Primary School data set shows a preference for the model with global-level constraints, suggesting the influence of overarching structures, possibly driven by classroom dynamics, which dominate the characteristics of individual and dyadic interactions.

Given these insights, we proceed to integrate memory effects into our models to establish their impact on explaining network dynamics. We extend the analysis using three different specifications, each augmenting the time-averaged degree sequence with memory-based constraints. Our objective in this next phase is to quantify the significance of memory and its heterogeneity across the data sets. We continue to use the AIC to rank the performance of these "full" models, similar to our approach in the memoryless scenario. For each specification, we have that

$$K_{\text{model}} = \begin{cases} N + N(N-1)/2 & \text{for } \mathcal{H}_1, \\ N + N & \text{for } \mathcal{H}_2, \\ N + 1 & \text{for } \mathcal{H}_3. \end{cases} \tag{54}$$

We compare these three models, as well as a baseline model that only constrains the degree sequence, using the AIC$_{\text{total}}$. It is important to emphasize that the likelihood expressions employed in this comparison are distinct from those discussed previously. In this analysis, we evaluate each model's ability not only to reproduce the presence of links accurately, but also to capture their persistence over time effectively. Based on the results presented in Table II, the preferred model for both data sets is the one that incorporates constraints on the persisting degree. This outcome suggests that describing the persistence of links is more effectively achieved using variables associated with individual nodes rather than through global system attributes or link-specific properties. Thus, memory in these networks is better explained by node-level characteristics.

Given that we have data on external events for these empirical temporal networks, we further evaluate the ability of our best-performing model to capture real-world changes. We apply the winning model and compare the detected structural breaks with documented external events to assess its predictive accuracy.

In assessing the model's performance, it is important to note that our approach requires at least two snapshots, setting a resolution limit. We consider real events detected if they fall within $t_{\text{break}} \pm 2$. As shown in Fig. 3, for the MIT data set [Fig. 3(a)], our model predicted a total of 20 events. Within
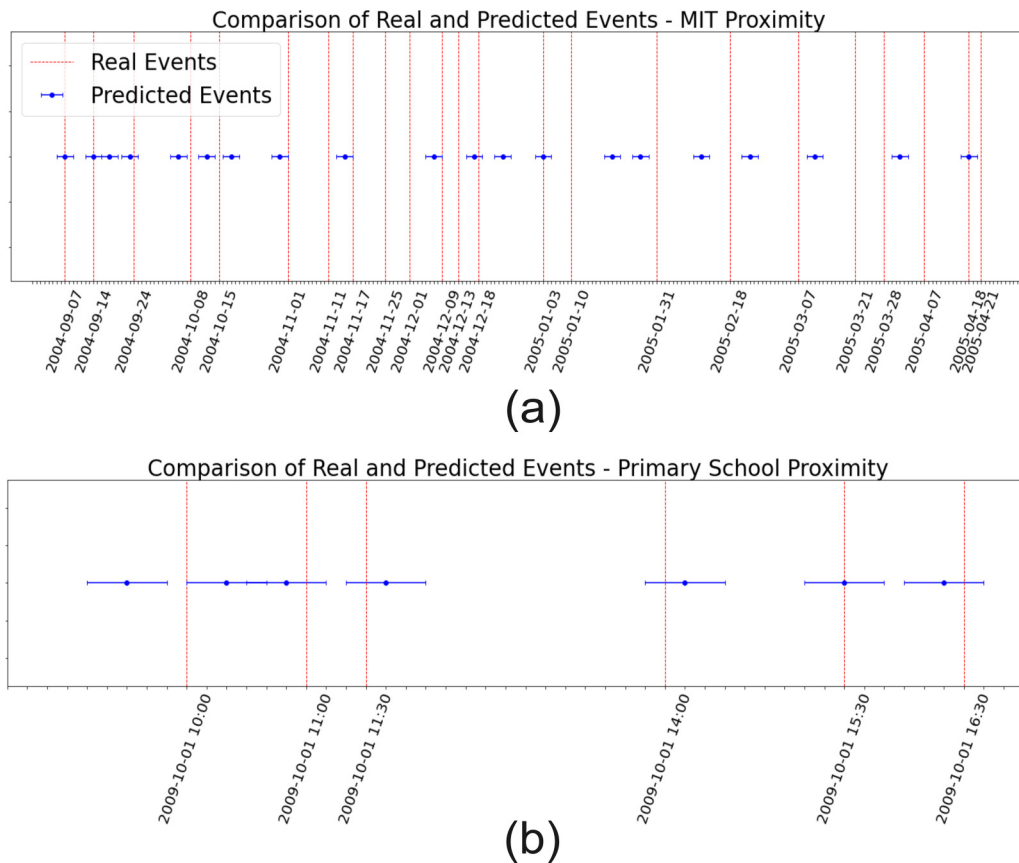
FIG. 3. Figure illustrates the detection of structural breaks: red lines represent known external events, while blue points denote predicted events. Horizontal bars show the uncertainty associated with each predicted event. (b) Corresponds to the Primary School data set, with events such as classroom changes and playground times, while (a) relates to the MIT data set, featuring identified external events.

a range of two days (snapshots), we accurately detected 9 out of 23 real events. In the case of the Primary School data set [Fig. 3(b)], our model predicted 7 events and successfully captured all 6 of the assumed real events within a two-snapshot range. Although the model in the case of the MIT data set only partially aligned with real events, many predicted points closely approximate actual events, underscoring our memory model's capacity to reflect the impact of external influences on the system. Identified changes are not only linked to external events but also indicate substantial shifts in the model's parameters that define interaction dynamics among network participants. These findings suggest distinct dynamic segments within the network, each characterized by unique interaction patterns.

At this point, we replicate the initial tests performed on the entire MIT data set, now focusing on the segments identified through structural break detection. We selected three of the longest periods for testing: $T_1^{\text{MIT}} = [2004/10/05 \rightarrow 2004/10/12]$, $T_2^{\text{MIT}} = [2004/11/15 \rightarrow 2004/12/07]$, and $T_3^{\text{MIT}} = [2005/03/11 \rightarrow 2005/04/01]$. For each segment, we compare the persisting degree for different values of τ computed on the shuffled network with the actual data. Our observations reveal a gradual decrease in memory effect across all three periods as τ increases. Specifically, the percentage of nodes with a Z-score greater than 1.96, indicating a statistically significant difference, decreases from 33% for $\tau = 1\%$ to 0% for $\tau = 3$ in period T_1^{MIT}, from 87% for $\tau =$

1% to 3% for $\tau = 3$ in period T_2^{MIT}, and from 73% to 0% in period T_3^{MIT}. These results suggest that the system progressively loses memory of its previous state with increasing τ, highlighting behavior that approaches a Markovian model. However, this trend is observed only within specific segments delineated by the structural break detection algorithm, emphasizing the importance of localized analysis within broader temporal networks.

Here, we move to the last part of our real-world applications by assessing the predictive performance of the four models. We compute the expected values for $h_i(\tau)$ at $\tau = 2$ and 3, and compare them with the real values using the mean squared error (MSE) [see Eq. (55)]. We select specific periods for this comparison to demonstrate the models' effectiveness across different data sets. For the MIT data set, we focus on the period T_1^{MIT}. For the Primary School data set, we examine two distinct periods, $T_1^{\text{School}} = [8:30 \rightarrow 9:30]$ and $T_2^{\text{School}} = [11:40 \rightarrow 14:00]$:

$$\text{MSE}(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} (y_i - \hat{y}_i)^2, \qquad (55)$$

where $\hat{y}$ is the predicted value while y is the true one.

The results in Table III indicate that the models do not precisely reproduce values of $h_i(\tau)$ for $\tau \neq 1$. Notably, for periods T_1^{MIT} and T_2^{School}, the model with constraints on each node outperforms other models in predicting actual autocorrelation values across different τ, aligning with our earlier

TABLE III. In the table, we present the mean squared error (MSE) computed between the predicted values of $h_i(\tau)$ and the measured values, using different models. Each model's parameters, estimated from the nearest segment identified by the structural break detection algorithm of the respective model, were used for evaluation. Bold text highlights the minimum MSE for each segment and each value of τ, indicating the best-performing model for those cases.

	T_1^{MIT}				T_1^{School}				T_2^{School}			
τ	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\mathcal{H}_{2,\nu}$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\mathcal{H}_{2,\nu}$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\mathcal{H}_{2,\nu}$
2	3.99	**0.95**	1.13	8.63	0.59	0.31	**0.28**	0.57	1.68	**0.4**	0.58	2.59
3	3.73	**1.25**	1.46	7.45	0.62	0.44	**0.41**	0.62	1.19	**0.62**	0.75	1.84

findings. The only exception occurs during the period T_1^{School}, which corresponds to a time when students are in class. Here, a global driver appears to dominate local factors, leading to memory effects more effectively explained by a single global parameter. AIC calculations for this period further support this observation, with $\text{AIC}(\mathcal{H}_3) = 13\,983.2$ and $\text{AIC}(\mathcal{H}_2) = 14\,115.64$, indicating a stronger fit for the model with global constraints ($\mathcal{H}_3$). Despite the better performance of H_3 in the T_1^{School} segment, the overall evidence suggests that the model with local constraints (based on $\text{AIC}_{\text{total}}$) is generally preferable for this data set, indicating a broader efficacy of local constraints across the entire network.

It is worth noting how the memoryless model exhibits consistent or improved performance as τ increases. This trend suggests a gradual memory loss in certain periods as we move further from the system's current state. This effect can be clarified by considering the following equations for the expected behavior of the autocorrelation function in models with and without memory:

$$\langle a_{ij}(t)a_{ij}(t+\tau)\rangle_\mu = \langle a_{ij}(t)\rangle_\mu^2 + O(e^{-\frac{\tau}{\tau_c^i}}), \qquad (56)$$

$$\langle a_{ij}(t)a_{ij}(t+\tau)\rangle_\nu = \langle a_{ij}(t)\rangle_\nu^2. \qquad (57)$$

Here, Eq. (56) introduces a decay term that attenuates as τ increases, reflecting the memory model's capacity to gradually lose the memory of initial conditions. This decay is governed by the relaxation time τ_c, indicative of the number of iterations required for the system to reach a stationary distribution. The relaxation time, related to the second eigenvalue by Eq. (43), characterizes each link's tendency towards equilibrium. In contrast, Eq. (57) for the memoryless model suggests constant autocorrelation, independent of τ. The comparative analysis of these models, as shown in Table III, highlights a nontrivial Markovian nature of the data, reinforcing the suitability of our modeling approach for these systems. The implications of τ_c extend beyond theoretical interest; they provide practical insights into the dynamic behavior of nodes and the identified segments within the network.

VIII. CONCLUSIONS

Working with exponential random graph models often presents challenges in parameter estimation, particularly due to the complex analytical summations required to compute the partition function. These challenges are further relevant in temporal exponential random graph models, which introduce additional degrees of freedom, complicating their tractability.

In this paper, we tackled these challenges by mapping a specific class of TERGMs, defined by the concept of persist-

ing degree, to a combination of one-dimensional Ising models. This innovative approach allowed us to derive an analytic solution and calculate the partition function exactly, thereby providing precise expressions for the probability of a link between two nodes and the persistence of this link over time.

Our results emphasize the importance of models that capture memory effects, such as the persisting degree, for maintaining crucial network dynamics. While simplifications like averaging can be useful under certain conditions, they generally fall short of capturing the true dynamics. Our framework introduces a methodology to assess when such simplifications are warranted, distinguishing between, for example, global effects driven by external factors and local effects intrinsic to individual nodes.

While our proposed methodology initially applies only to stationary temporal networks, we address this limitation by introducing an approach that leverages our unique class of models to identify segments where the stationarity hypothesis holds. This involves exploring the presence of structural breaks within temporal networks. By utilizing techniques for detecting changing points, we successfully identified significant shifts in network behavior and dynamics. These shifts are indicative of underlying changes in the processes governing network evolution.

We applied our methodologies to two real-world datasets, demonstrating the model's ability to characterize each node by different memory effects and to leverage these changes to infer external events. Additionally, we introduced the concept of relaxation time for each node, which can be precisely computed in our framework. This measure represents the number of snapshots required for a network to reach a stationary distribution from a random configuration and is particularly useful for predicting the system's response to external perturbations.

Beyond describing network dynamics, the models we developed may serve as generative models, offering new directions for leveraging these approaches to address the lack of canonical generative models for temporal networks. Given the nature of maximum entropy, the null models developed in this framework can be broadly applied to various scenarios, providing an unbiased basis for comparison and uncovering hidden patterns and new insights in real systems.

and by the European Union - NextGenerationEU - National Recovery and Resilience Plan (NRRP) - Mission 4 Component 2 Investment 1.3 through the project RECON-NET: Reconstruction, Resilience and Recovery of Socio-Economic Networks CUP D63C22001310006 financed through the "Bando A Cascata" of the University of Pisa under the FAIR PE0000013 project; by D.G. acknowledges support from the Dutch Econophysics Foundation (Stichting Econophysics, Leiden, the Netherlands) and from the project NAEFR: Network Analysis of Economic and Financial Resilience (PRO3 Scuole), CUP D67G22000130001.

APPENDIX A: FINDING $\vec{\alpha}$ FOR THE THREE MEMORYLESS MODELS

Here, we display how to derive the functional form of p_{ij} for the three specifications of the memoryless model and how to estimate the values of $\vec{\alpha}$ to ensure the constraints are accurately reproduced. We focus particularly on the results for the most constraining Hamiltonian, which applies constraints at the dyadic level; similar reasoning applies to the other two cases.

Let us recall the Hamiltonian with constraints at the dyadic level:

$$\mathcal{H}_{1,v}(\mathcal{G}) = \frac{1}{T} \sum_{t} \sum_{j<i} \theta_{ij} a_{ij}(t) = \sum_{j<i} \theta_{ij} \overline{a_{ij}}. \quad (A1)$$

In this case, the expression for the Hamiltonian enables us to draw upon the results from the work of Newman [23]. By applying the same arguments used by Newman, we can derive the functional form of p_{ij},

$$p_{ij} = \langle \overline{a_{ij}} \rangle = \frac{e^{-\theta_{ij}}}{1 + e^{-\theta_{ij}}}. \quad (A2)$$

The models defined by Eqs. (22) and (23) are special cases of this general formulation, where in order, for the specification in Eq. (22), $\theta_{ij} = \theta_i + \theta_j$, while for Eq. (23), $\theta_{ij} = \theta/N \; \forall \; i, j$. The expression of x_{ij} is defined in the main text in Eq. (25).

To capture the right values of $\vec{\theta}$, such that the observed constraints are reproduced, we have to solve a series of equations that came from the maximum likelihood principle. For the three specifications, the equations are

$$\langle \overline{a_{ij}} \rangle = \frac{x_{ij}}{1 + x_{ij}} = \overline{a_{ij}^*} \quad \forall \; i, \; j \in N \quad (A3)$$

for the first specification in Eq. (21). Here we have to solve $\frac{N(N-1)}{2}$ equations, and $\overline{a_{ij}^*}$ is the average over time of the link between i and j.

$$\langle \overline{k_i} \rangle = \sum_{i \neq j} \frac{x_i x_j}{1 + x_i x_j} = \overline{k_i^*} \quad \forall \; i \in N. \quad (A4)$$

For the second expression of the general model in Eq. (22). In this case we have to solve N equations, and $\overline{k_i^*}$ is the average over time of the degree of node i:

$$\frac{x}{1+x} = \frac{1}{N(N-1)} \sum_i \overline{k_i^*} = \frac{\overline{k^*}}{(N-1)}. \quad (A5)$$

In this last case, the equation to solve is only one, and $\overline{k^*}$ is the average overtime of the average degree.

APPENDIX B: MAPPING TO THE ONE-DIMENSIONAL ISING MODEL

Here we provide a detailed proof of the main results presented in the text. To begin, we explicitly rewrite the Hamiltonian given in Eq. (30), considering constraints at the level of links ($\mathcal{H}_{1,m}$):

$$\mathcal{H}_1(\mathcal{G}) \equiv \mathcal{H}_{2,v} + \mathcal{H}_{1,m}$$

$$= \sum_{i=1}^{N} (\alpha_i \bar{k}_i) + \sum_{j<i} \beta_{ij} \overline{a_{ij}(t) a_{ij}(t+1)}$$

$$= \frac{1}{T} \sum_{t=1}^{T} \left[\sum_{i=1}^{N} \alpha_i k_i(t) + \sum_{i<j} \beta_{ij} a_{ij}(t) a_{ij}(t+1) \right]$$

$$= \frac{1}{T} \sum_{t=1}^{T} \left[\sum_{i \neq j} \alpha_i a_{ij}(t) + \sum_{j<i} \beta_{ij} a_{ij}(t) a_{ij}(t+1) \right]$$

$$= \frac{1}{T} \sum_{t=1}^{T} \sum_{j<i} [(\alpha_i + \alpha_j) a_{ij}(t) + \beta_{ij} a_{ij}(t) a_{ij}(t+1)]. \quad (B1)$$

The other two cases, as for the memoryless model, are just a particular specification of this, where the two Hamiltonian reads as

$$\mathcal{H}_2(\mathcal{G}) \equiv \mathcal{H}_{2,v} + \mathcal{H}_{2,m}$$

$$= \sum_{i=1}^{N} [\alpha_i \bar{k}_i + \beta_i \bar{h}_i(1)]. \quad (B2)$$

Using $\mathcal{H}_{2,m}$, and

$$\mathcal{H}_3(\mathcal{G}) \equiv \mathcal{H}_{2,v} + \mathcal{H}_{3,m}$$

$$= \left(\sum_{i=1}^{N} \alpha_i \bar{k}_i \right) + \beta \bar{h}, \quad (B3)$$

for $\mathcal{H}_{3,m}$. From now on, we will refer to $\mathcal{H}_1(\mathcal{G})$ as $\mathcal{H}(\mathcal{G})$.

We now focus on Eq. (B1). Let us introduce the "spin" variables $\sigma_{ij}(t) = \pm 1$ defined through

$$a_{ij}(t) \equiv \frac{\sigma_{ij}(t) + 1}{2}. \quad (B4)$$

In Eq. (B1), each graph $G(t)$ in the trajectory $\mathcal{G}$ has been parametrized by the adjacency matrix $\mathbf{A}(t)$ with entries $a_{ij}(t)$. Equivalently, we can parametrize $G(t)$ by the "spin matrix" $\boldsymbol{\Sigma}(t)$ with entries $\sigma_{ij}(t)$. If we insert Eq. (B4) into (B1), after some algebra we arrive at

$$\mathcal{H}(\mathcal{G}) = \mathcal{H}_0 - \sum_{t=1}^{T} \sum_{j<i} [B_{ij} \sigma_{ij}(t) + J_{ij} \sigma_{ij}(t) \sigma_{ij}(t+1)],$$

where we have defined

$$B_{ij} \equiv -\frac{1}{2T} (\alpha_i + \alpha_j + \beta_{ij}), \quad (B5)$$

$$J_{ij} \equiv -\frac{1}{4T} (\beta_{ij}), \quad (B6)$$

$$\mathcal{H}_0 \equiv \sum_{j<i} \left(\frac{\alpha_i + \alpha_j}{2} + \frac{\beta_{ij}}{4} \right). \quad (B7)$$

Note that $\mathcal{H}_0$ is a constant term that does not depend on the specific graph trajectory $\mathcal{G}$. As it has no effect on any observable of the model, we can omit it from further considerations. Consequently, the Hamiltonian simplifies to

$$\mathcal{H}(\mathcal{G}) = \sum_{j<i} H(\mathcal{G}_{ij}), \qquad (B8)$$

where, if we still parametrize $G(t)$ using the matrix $\boldsymbol{\Sigma}(t)$, $\mathcal{G}_{ij} = \{\sigma_{ij}(1), \dots, \sigma_{ij}(T)\}$ is the restriction of $\mathcal{G}$ to the pair of nodes i, j, i.e., the temporal sequence of (connection and disconnection) events for such pair, and

$$H(\mathcal{G}_{ij}) \equiv -\sum_{t=1}^{T} [B_{ij}\sigma_{ij}(t) + J_{ij}\sigma_{ij}(t)\sigma_{ij}(t+1)]. \qquad (B9)$$

A crucial observation is that Eq. (B9) corresponds to the Hamiltonian of a one-dimensional Ising model with first-neighbor interactions, where each "spin" $\sigma_{ij}(t)$ is located at lattice site t and interacts with the spin at site $t+1$. Here, the parameters B_{ij} and J_{ij} function as the external field and coupling constant, respectively. This model is well known for its solvability in an analytical framework [34]. Before presenting the solution, it is important to note that Eq. (B8) implies that

$$Z(B_{ij}, J_{ij}) \equiv \sum_{\mathcal{G}_{ij}} e^{-H(\mathcal{G}_{ij})} \qquad (B10)$$

is the partition function for the dyad-specific Hamiltonian (B9). If we assume independence among dyads, the partition function for the graph-wide Hamiltonian (B1) can be expressed as the product of the partition functions of all dyads:

$$\mathcal{Z}(\mathbf{B}, \mathbf{J}) \equiv \sum_{\mathcal{G}} e^{-\mathcal{H}(\mathcal{G})} = \prod_{j<i} Z(B_{ij}, J_{ij}). \qquad (B11)$$

Similarly, if

$$P(\mathcal{G}_{ij}|B_{ij}, J_{ij}) = \frac{e^{-H(\mathcal{G}_{ij})}}{Z(B_{ij}, J_{ij})} \qquad (B12)$$

is the probability of the dyadic trajectory $\mathcal{G}_{ij}$ in the model defined by Eq. (B9), then the probability $\mathcal{P}(\mathcal{G}|\mathbf{B}, \mathbf{J})$ of an entire graph trajectory $\mathcal{G}$ in the model specified by Eq. (B1) is

$$\mathcal{P}(\mathcal{G}|\mathbf{B}, \mathbf{J}) = \frac{e^{-\mathcal{H}(\mathcal{G})}}{\mathcal{Z}(\mathbf{B}, \mathbf{J})} = \prod_{j<i} P(\mathcal{G}_{ij}|B_{ij}, J_{ij}). \qquad (B13)$$

APPENDIX C: PARTITION FUNCTION

We now present the solution of the dyadic model as defined by Eq. (B9), by adapting it from the well-known solution of the one-dimensional Ising model [34]. The imposition of the periodicity condition, as specified in Eq. (29), ensures that all sites (time steps) are statistically equivalent, implying that each site contributes equally to the overall statistical properties of the model:

$$\langle \sigma_{ij}(1) \rangle = \langle \sigma_{ij}(2) \rangle = \cdots = \langle \sigma_{ij}(T) \rangle. \qquad (C1)$$

The system is therefore translationally (here, temporally) invariant. The partition function (B10) is

$$Z(B_{ij}, J_{ij}) = \sum_{\mathcal{G}_{ij}} \exp[B_{ij}\sigma_{ij}(t) + J_{ij}\sigma_{ij}(t)\sigma_{ij}(t+1)], \quad (C2)$$

and can be rewritten as a product of terms involving only two successive time steps:

$$Z(B_{ij}, J_{ij}) = \prod_{t=1}^{T} \sum_{\sigma_{ij}} V_{ij}(\sigma_{ij}(t), \sigma_{ij}(t+1)), \qquad (C3)$$

where the function $V_{ij}(x, y)$ is defined as

$$V_{ij}(x, y) \equiv \exp\left(\frac{x+y}{2}B_{ij} + xyJ_{ij}\right). \qquad (C4)$$

Given that both x and y can take the values ± 1, we can define the four possible states of V_{ij} as V_{ij}^{++}, V_{ij}^{+-}, V_{ij}^{-+}, and V_{ij}^{--}, corresponding to the combinations of x and y being $(+1, +1)$, $(+1, -1)$, $(-1, +1)$, and $(-1, -1)$, respectively. These can be arranged to form a 2×2 matrix called as the *transfer matrix* $\mathbf{V}_{ij}$, crucial in solving the model [34]

$$\mathbf{V}_{ij} \equiv \begin{pmatrix} V_{ij}^{++} & V_{ij}^{+-} \\ V_{ij}^{-+} & V_{ij}^{--} \end{pmatrix} = \begin{pmatrix} e^{J_{ij}+B_{ij}} & e^{-J_{ij}} \\ e^{-J_{ij}} & e^{J_{ij}-B_{ij}} \end{pmatrix}. \qquad (C5)$$

This allows us to rewrite Eq. (C3) as

$$Z(B_{ij}, J_{ij}) = \mathrm{Tr}(\mathbf{V}_{ij}^T) \qquad (C6)$$

(where $\mathbf{V}_{ij}^T$ denotes the Tth matrix power of $\mathbf{V}_{ij}$, and *not* the transpose of the latter).

Now let $\vec{v}_{ij}^{\pm}$ denote the two eigenvectors of $\mathbf{V}_{ij}$, and $\lambda_{ij}^{\pm}$ the corresponding eigenvalues, with $\lambda_{ij}^{+} \geqslant \lambda_{ij}^{-}$. The relationship between the matrix, its eigenvectors, and eigenvalues is given by

$$\mathbf{V}_{ij}\vec{v}_{ij}^{\pm} = \lambda_{ij}^{\pm}\vec{v}_{ij}^{\pm}. \qquad (C7)$$

The 2×2 matrix $\mathbf{Q}_{ij} \equiv (\vec{v}_{ij}^{+}, \vec{v}_{ij}^{-})$, having column vectors $\vec{v}_{ij}^{+}$ and $\vec{v}_{ij}^{-}$, diagonalizes $\mathbf{V}_{ij}$, i.e.,

$$\mathbf{V}_{ij} = \mathbf{Q}_{ij}\begin{pmatrix} \lambda_{ij}^{+} & 0 \\ 0 & \lambda_{ij}^{-} \end{pmatrix}\mathbf{Q}_{ij}^{-1}. \qquad (C8)$$

A direct calculation of the eigenvalues and eigenvectors yields

$$\lambda_{ij}^{\pm} = e^{J_{ij}}\cosh B_{ij} \pm \sqrt{e^{2J_{ij}}\sinh^2 B_{ij} + e^{-2J_{ij}}}. \qquad (C9)$$

It then follows, using the cyclic properties of the trace, that Eq. (C6) simply reduces to

$$Z(B_{ij}, J_{ij}) = \mathrm{Tr}\begin{pmatrix} \lambda_{ij}^{+} & 0 \\ 0 & \lambda_{ij}^{-} \end{pmatrix}^T = (\lambda_{ij}^{+})^T + (\lambda_{ij}^{-})^T, \quad (C10)$$

and the full partition function (B11) is

$$\mathcal{Z}(\mathbf{B}, \mathbf{J}) = \prod_{j<i} [(\lambda_{ij}^{+})^T + (\lambda_{ij}^{-})^T]. \qquad (C11)$$

Similarly, the probability of the dyadic trajectory $\mathcal{G}_{ij}$ is

$$P(\mathcal{G}_{ij}|B_{ij}, J_{ij}) = \frac{\prod_{t=1}^{T} V_{ij}(\sigma_{ij}(t), \sigma_{ij}(t+1))}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T} \qquad \text{(C12)}$$

and that of the entire graph trajectory $\mathcal{G}$ is

$$\mathcal{P}(\mathcal{G}|\mathbf{B}, \mathbf{J}) = \prod_{j<i} \frac{\prod_{t=1}^{T} V_{ij}(\sigma_{ij}(t), \sigma_{ij}(t+1))}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T}. \qquad \text{(C13)}$$

APPENDIX D: EXPECTED VALUES

We start the calculation of the expected value of $\sigma_{ij}(t)$ using two distinct approaches to facilitate a comprehensive analysis and subsequent computation of the autocorrelation function. In the first approach, we determine the expected value by utilizing the free energy per spin, which is defined by the following equation:

$$F_{ij} = -\frac{1}{T}\ln Z(B_{ij}, J_{ij}), \qquad \text{(D1)}$$

and differentiating F_{ij} with respect to B_{ij}, we obtain $\langle\sigma_{ij}(t)\rangle$:

$$\frac{\partial \ln Z(B_{ij}, J_{ij})}{\partial B_{ij}} = \frac{1}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T}$$
$$\times \left[T(\lambda_{ij}^+)^{T-1}\frac{\partial \lambda_{ij}^+}{\partial B_{ij}} + T(\lambda_{ij}^-)^{T-1}\frac{\partial \lambda_-}{\partial B_{ij}} \right], \qquad \text{(D2)}$$

where $\frac{\partial \lambda_{ij}^\pm}{\partial B_{ij}}$ has the following expression:

$$\frac{\partial \lambda_{ij}^\pm}{\partial B_{ij}} = e^{J_{ij}}\left[\sinh B_{ij} \pm \frac{\sinh B_{ij}\cosh B_{ij}}{\sqrt{\sinh^2(B_{ij}) + e^{-4J_{ij}}}} \right]. \qquad \text{(D3)}$$

Considering that

$$\langle\sigma_{ij}(t)\rangle = -\frac{\partial F_{ij}}{\partial B_{ij}}, \qquad \text{(D4)}$$

we can finally obtain the expected value of $\sigma_{ij}(t)$:

$$\langle\sigma_{ij}(t)\rangle = \frac{e^{J_{ij}}}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T}\left[(\lambda_{ij}^+)^{T-1}\left(\sinh B_{ij} + \frac{\sinh B_{ij}\ \cosh B_{ij}}{\sqrt{\sinh^2(B_{ij}) + e^{-4J_{ij}}}} \right) + (\lambda_{ij}^-)^{T-1}\left(\sinh B_{ij} - \frac{\sinh B_{ij}\ \cosh B_{ij}}{\sqrt{\sinh^2(B_{ij}) + e^{-4J_{ij}}}} \right) \right]. \qquad \text{(D5)}$$

Given this general result, we can move to the thermodynamic limit, obtaining a simplified expression:

$$\langle\sigma_{ij}(t)\rangle_{tl} = \frac{e^{J_{ij}}}{\lambda_{ij}^+}\left[\sinh B_{ij} + \frac{\sinh B_{ij}\ \cosh B_{ij}}{\sqrt{\sinh^2(B_{ij}) + e^{-4J_{ij}}}} \right] = \frac{\sinh B_{ij}}{\sqrt{\sinh^2(B_{ij}) + e^{-4J_{ij}}}}. \qquad \text{(D6)}$$

As anticipated, the result obtained in Eq. (D6) can also be derived using an alternative technique: the transfer matrix method [34]. To employ this method, we introduce **S**, a diagonal matrix whose diagonal elements represent all possible spin values. This matrix plays a crucial role in transforming the system states into a format suitable for the transfer matrix analysis. Specifically, the matrix **S** is defined as follows:

$$\mathbf{S} \equiv \begin{pmatrix} S^{++} & S^{+-} \\ S^{-+} & S^{--} \end{pmatrix} = \begin{pmatrix} +1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \text{(D7)}$$

having elements $S^{\pm\pm} \equiv S(\pm 1, \pm 1)$, with

$$S(x, y) \equiv x\delta(x, y). \qquad \text{(D8)}$$

We can rewrite $\langle\sigma_{ij}(t)\rangle$:

$$\langle\sigma_{ij}(t)\rangle \equiv \sum_{\mathcal{G}_{ij}} \sigma_{ij}(t)P(\mathcal{G}_{ij}|B_{ij}, J_{ij}) = \frac{\text{Tr}(\mathbf{S}\mathbf{V}_{ij}^T)}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T}, \qquad \text{(D9)}$$

and using the eigenvector of the transfer matrix, we have

$$\langle\sigma_{ij}(t)\rangle = \frac{1}{Z(B_{ij}, J_{ij})}[(\lambda_{ij}^+)^T \langle v_{ij}^+|S|v_{ij}^+\rangle \qquad \text{(D10)}$$
$$+ (\lambda_{ij}^-)^T \langle v_{ij}^-|S|v_{ij}^-\rangle], \qquad \text{(D11)}$$

that, using also the results in Eq. (D6), in the thermodynamic limit is

$$\langle\sigma_{ij}(t)\rangle_{tl} = \langle v_{ij}^+|S|v_{ij}^+\rangle \qquad \text{(D12)}$$
$$= \frac{e^{J_{ij}}}{\lambda_{ij}^+}\left[\sinh B_{ij} + \frac{\sinh B_{ij}\ \cosh B_{ij}}{\sqrt{\sinh^2(B_{ij}) + e^{-4J_{ij}}}} \right]. \qquad \text{(D13)}$$

Hence, by combining Eqs. (D6), (D11), and (D13), we obtain an expression for the quantity $\langle v_{ij}^-|S|v_{ij}^-\rangle$:

$$\langle v_{ij}^-|S|v_{ij}^-\rangle = \frac{e^{J_{ij}}}{\lambda_{ij}^-}\left[\sinh B_{ij} - \frac{\sinh B_{ij}\ \cosh B_{ij}}{\sqrt{\sinh^2(B_{ij}) + e^{-4J_{ij}}}} \right]$$
$$= -\langle v_{ij}^+|S|v_{ij}^+\rangle, \qquad \text{(D14)}$$

and, after some algebraic manipulations, we find that for a finite number of snapshots, the expected value of a spin is

$$\langle\sigma_{ij}(t)\rangle = \langle v_{ij}^+|S|v_{ij}^+\rangle\left(\frac{(\lambda_{ij}^+)^T - (\lambda_{ij}^-)^T}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T} \right). \qquad \text{(D15)}$$

Now we have all the elements necessary to compute the expected value of $\sigma_{ij}(t)\sigma_{ij}(t+\tau)$ for $0 < \tau < T$. Similar to the expected value of the spin function, this can also be expressed using the transfer matrix:

$$
\begin{aligned}
&\langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle \\
&= \frac{1}{Z(B_{ij},J_{ij})}\sum_{\{\sigma_i=\pm1\}}\sigma_t\langle\sigma_t|V|\sigma_{t+1}\rangle\dots \\
&\quad\times\langle\sigma_{t+\tau-1}|V|\sigma_{t+\tau}\rangle\sigma_{t+\tau}\langle\sigma_{t+\tau}|V|\sigma_{t+\tau+1}\rangle\dots \\
&= \frac{1}{Z(B_{ij},J_{ij})}\sum_{\sigma_t,\sigma_{t+\tau}}\sigma_t\langle\sigma_t|V^\tau|\sigma_{t+\tau}\rangle\sigma_{t+\tau}\langle\sigma_{t+\tau}|V^{T-\tau}|\sigma_1\rangle.
\end{aligned}
$$

$$\text{(D16)}$$

It is also possible to write matrix V in its eigenvector space $|v_i\rangle$:

$$
V = \sum_{l\in\{+,-\}}|v_l\rangle\lambda_l\langle v_l|, \tag{D17}
$$

where λ_i are the correspondent eigenvalues. In this way, given that the eigenvectors are orthonormal we have

$$
V^n = \sum_{l\in\{+,-\}}|v_l\rangle\lambda_l^n\langle v_l|. \tag{D18}
$$

Hence, the expected value of the product of spins at different times is given by

$$
\begin{aligned}
&\langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle \\
&= \frac{1}{Z(B_{ij},J_{ij})}\sum_{\sigma_t,\sigma_{t+\tau}}\sum_{i,j\in\{+,-\}} \\
&\quad\times\left[\sigma_t\langle\sigma_t|v_i\rangle\lambda_i^{t+\tau}\langle v_i|\sigma_{t+\tau}\rangle\sigma_{t+\tau}\langle\sigma_{t+\tau}|v_j\rangle\lambda_j^{T-\tau}\langle v_j|\sigma_t\rangle\right].
\end{aligned}
$$

$$\text{(D19)}$$

We can then rewrite Eq. (D19) by utilizing the matrix representation of spin states S_i, in terms of its eigenvector basis:

$$
S_i = \sum_{\sigma_i}|\sigma_i\rangle\sigma_i\langle\sigma_i|. \tag{D20}
$$

In this way moving $\langle v_j|\sigma_t\rangle$ (is just a number) and summing over $\sigma_{ij}(t)$ and $\sigma_{ij}(t+\tau)$, we obtain

$$
\langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle = \frac{\sum_{i,j}\langle v_j|S_t|v_i\rangle\lambda_i^\tau\langle v_i|S_{t+\tau}|v_j\rangle\lambda_j^{T-\tau}}{\sum_i\lambda_i^T}, \tag{D21}
$$

where $S_i = S\,\forall\,i$. Hence, after writing explicitly the summation, Eq. (D21) becomes

$$
\langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle = \frac{1}{\lambda_+^T+\lambda_-^T}[\lambda_+^T\langle v^+|S|v^+\rangle^2 + \lambda_-^T\langle v^-|S|v^-\rangle^2 + [(\lambda_{ij}^-)^\tau(\lambda_{ij}^+)^{T-\tau} + (\lambda_{ij}^+)^\tau(\lambda_{ij}^-)^{T-\tau}]\langle v^+|S|v^+\rangle(1-\langle v^+|S|v^+\rangle)]].
$$

$$\text{(D22)}$$

Finally, incorporating the results from Eqs. (D14) and (D13), and performing some algebraic manipulations, we get

$$
\langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle = \frac{1}{Z}\mathrm{Tr}[SV^\tau SV^{N-\tau}] = \langle\sigma_{ij}(t)\rangle_{tl}^2 + \langle\sigma_{ij}(t)\rangle_{tl}[1-\langle\sigma_{ij}(t)\rangle_{tl}]\left(\frac{(\lambda_{ij}^-)^\tau(\lambda_{ij}^+)^{T-\tau} + (\lambda_{ij}^+)^\tau(\lambda_{ij}^-)^{T-\tau}}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T}\right), \tag{D23}
$$

that in the limit $T\to\infty$ (corresponding to long graph trajectories) with τ fixed and finite (which necessarily implies $\tau\ll T$) becomes

$$
\langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle = \langle\sigma(t)\rangle_{tl}^2 + \langle\sigma(t)\rangle_{tl}[1-\langle\sigma(t)\rangle_{tl}]\left(\frac{\lambda_{ij}^+}{\lambda_{ij}^-}\right)^\tau. \tag{D24}
$$

Now, it should be noted that the model exhibits a manifest "translational" invariance, which in the context of our discussion translates to *temporal* invariance. This invariance implies that both $\langle\sigma_{ij}(t)\rangle$ and $\langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle$ are independent of t. As a result, the expectation values of

time-averaged quantities coincide with those of the quantities themselves:

$$
\overline{\langle\sigma_{ij}(t)\rangle} = \langle\sigma_{ij}(t)\rangle, \tag{D25}
$$

$$
\overline{\langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle} = \langle\sigma_{ij}(t)\sigma_{ij}(t+\tau)\rangle. \tag{D26}
$$

A similar result holds if we go back to the original variables $a_{ij}(t)$ using Eq. (B4). By doing so, we can express the quantities in terms of $a_{ij}(t)$, leading to

$$
p_{ij} \equiv \overline{\langle a_{ij}(t)\rangle} = \langle a_{ij}(t)\rangle, \tag{D27}
$$

$$
q_{ij}(\tau) \equiv \overline{\langle a_{ij}(t)a_{ij}(t+\tau)\rangle} = \langle a_{ij}(t)a_{ij}(t+\tau)\rangle. \tag{D28}
$$

Therefore, using Eqs. (B4) and (D15) we derive for p_{ij} the following relationship:

$$
p_{ij} = \left(\frac{e^{2J_{ij}}\sinh B_{ij}}{2\sqrt{1+e^{4J_{ij}}\sinh^2 B_{ij}}}\right)\left(\frac{(\lambda_{ij}^+)^T - (\lambda_{ij}^-)^T}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T}\right) + \frac{1}{2} = \left(\frac{x_ix_jy_{ij}-1}{2\sqrt{4x_ix_j+(x_ix_jy_{ij}-1)^2}}\right)\left(\frac{(\lambda_{ij}^+)^T - (\lambda_{ij}^-)^T}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T}\right) + \frac{1}{2},
$$

$$\text{(D29)}$$

where we have introduced the parameters

$$x_i \equiv e^{-\alpha_i}, \quad y_{ij} \equiv e^{-\beta_{ij}}. \tag{D30}$$

While for $q_{ij}(\tau)$, a similar derivation leads to

$$q_{ij}(\tau) = \tilde{p}_{ij}^2 + \tilde{p}_{ij}(1 - \tilde{p}_{ij}) \\ \times \left(\frac{(\lambda_{ij}^-)^\tau (\lambda_{ij}^+)^{T-\tau} + (\lambda_{ij}^+)^\tau (\lambda_{ij}^-)^{T-\tau}}{(\lambda_{ij}^+)^T + (\lambda_{ij}^-)^T} \right), \tag{D31}$$

where $\tilde{p}_{ij}$ corresponds to the value of p_{ij} computed in the long limit trajectory:

$$\tilde{p}_{ij} = \left(\frac{e^{2J_{ij}} \sinh B_{ij}}{2\sqrt{1 + e^{4J_{ij}} \sinh^2 B_{ij}}} \right) + \frac{1}{2}. \tag{D32}$$

q_{ij} in the thermodynamic limit assumes the following expression:

$$\tilde{q}_{ij}(\tau) = \tilde{p}_{ij}^2 + \tilde{p}_{ij}(1 - \tilde{p}_{ij}) \left(\frac{\lambda_{ij}^+}{\lambda_{ij}^-} \right)^\tau. \tag{D33}$$

APPENDIX E: MAXIMUM LIKELIHOOD ESTIMATION

The above expressions enable the computation of all relevant expected properties for the time series generated by the model. This calculation is contingent upon setting the parameters **B** and **J** to their optimal values **B*** and **J***. These optimal values are those that maximize the likelihood $\mathcal{P}(\mathcal{G}^*|\mathbf{B}, \mathbf{J})$ of the observed graph trajectory $\mathcal{G}^*$, where the probability $\mathcal{P}(\mathcal{G}|\mathbf{B}, \mathbf{J})$ is defined in Eq. (C13). Similarly, this optimization process can be equivalently expressed in terms of the parameters $\vec{x}$ and $\vec{y}$. We seek the values $\vec{x}^*$ and $\vec{y}^*$ that maximize the likelihood function for the same observed graph trajectory.

As discussed in the main text, the probability $\mathcal{P}(\mathcal{G}|\mathbf{B}, \mathbf{J})$ is modeled as specified in Eq. (B13), where the Hamiltonian $\mathcal{H}(\mathcal{G})$ is described by Eq. (B1). According to a general theorem [25], the values $\vec{x}^*$ and $\vec{y}^*$ that maximize the likelihood satisfy a system of equations for $N + \frac{N(N-1)}{2}$ variables. These equations are

$$\langle \bar{k}_i \rangle = \sum_{j \neq i} p_{ij} = \bar{k}_i^* \quad \forall \, i, \tag{E1}$$

$$\overline{\langle a_{ij}(t) a_{ij}(t+1) \rangle} = q_{ij}(1) = \bar{h}_{ij}^* \quad \forall \, i, j, \tag{E2}$$

where p_{ij} and $q_{ij}(1)$, which are both functions of $\vec{x}$ and $\vec{y}$, as specified in Eqs. (D29) and (D31). These formulations not only ensure the consistency of model estimates with observed data but also enable the direct calculation of node and dyadic properties based on the fitted model parameters.

For the models specified by $\mathcal{H}_3$, the system of $\frac{N(N-1)}{2}$ equations, as shown in Eq. (E2), is replaced with different sets of equations depending on the constraints applied. Specifically, we solve either N equations for each node given by

$$\langle \bar{h}_i \rangle = \sum_{j \neq i} q_{ij}(1) = \bar{h}_i^* \quad \forall \, i, \tag{E3}$$

or a single aggregated equation for the entire network:

$$\frac{1}{N} \sum_{i=1}^{N} \langle \bar{h}_i \rangle = \sum_{i \neq j} q_{ij}(1) = \frac{1}{N} \sum_{i=1}^{N} \bar{h}_i^*. \tag{E4}$$

It is important to note that even if we impose a uniform β across all nodes, q_{ij} remains distinct for each pair due to its dependence on the values of $\vec{\alpha}$.

In general, the expressions for p_{ij} and q_{ij} are the ones described in Eqs. (D29) and (D31), although if the long trajectory limit is satisfied, the thermodynamic limit versions can be implemented.

APPENDIX F: FROM MEMORY TO MEMORYLESS

Here we demonstrate how, starting from the full model and deactivating the memory component, we can derive the same expression for p_{ij} as in the memoryless case. The absence of memory in the system can be effectively modeled by setting $\beta_i = 0$ for all i, which translates to $J_{ij} = 0$ for all pairs i, j. This setting leads to the simplification of the Hamiltonian, implying

$$\lambda_{ij}^+ = 2 \cosh B_{ij}, \tag{F1}$$

$$\lambda_{ij}^- = 0, \tag{F2}$$

$$\sqrt{\sinh^2(B_{ij}) + e^{-4J_{ij}}} = \cosh B_{ij}. \tag{F3}$$

Using these equations in the expression of $\langle \sigma_{i,j}(t) \rangle$ [Eq. (D15)], we have

$$\langle \sigma_{ij}(t) \rangle = \frac{\sinh B_{ij}}{\cosh B_{ij}} = \frac{e^{2B_{ij}} - 1}{e^{2B_{ij}} + 1}. \tag{F4}$$

Considering that $B_{ij} = \frac{1}{2}[\ln(x_i) + \ln(x_j)]$, we obtain

$$\langle \sigma_{ij}(t) \rangle = \left(\frac{x_i x_j - 1}{x_i x_j + 1} \right). \tag{F5}$$

The primary quantity of interest is $\langle a_{ij} \rangle$. Recalling the relationship between the spin variable and the adjacency matrix element $\langle a_{ij} \rangle = \frac{\langle \sigma_t \rangle + 1}{2}$, we get

$$\langle a_{ij} \rangle = \frac{x_i x_j}{x_i x_j + 1}. \tag{F6}$$

This expression for $\langle a_{ij} \rangle$ matches exactly with that expected in the memoryless case.

APPENDIX G: DERIVING THE STOCHASTIC MATRICES

The model constructs each link within the network through a stochastic matrix, which encapsulates the dynamics and transition probabilities of link formation and dissolution over time. In this section, we describe the formulation of this matrix, grounding our discussion in the quantities previously defined in the model. We aim to explain the coupled probabilities of events for each link by utilizing the defined parameters and relationships. We begin by connecting these quantities to formally introduce the stochastic matrix for each link.

(1) The probability that two nodes have a connection at time t and $t + \tau$ (with $\tau = 1$):

$$P[a_{ij}(t + \tau) = 1 \text{ and } a_{ij}(t) = 1]$$
$$= \langle a_{ij}(t)a_{ij}(t+1)\rangle = q_{ij}. \qquad \text{(G1)}$$

(2) The probability that two nodes have a connection at time t and not $t + \tau$:

$$P[a_{ij}(t + \tau) = 0 \text{ and } a_{ij}(t) = 1]$$
$$= \langle a_{ij}(t)[1 - a_{ij}(t+1)]\rangle = p_{ij} - q_{ij}. \quad \text{(G2)}$$

(3) The probability that two nodes do not have a connection at time t having a connection a time $t + \tau$:

$$P[a_{ij}(t + \tau) = 1 \text{ and } a_{ij}(t) = 0]$$
$$= \langle [1 - a_{ij}(t)]a_{ij}(t+1)\rangle = p_{ij} - q_{ij}. \quad \text{(G3)}$$

(4) The probability that two nodes do not have a connection at time t and $t + \tau$:

$$P[a_{ij}(t + \tau) = 0 \text{ and } a_{ij}(t) = 0]$$
$$= \langle [1 - a_{ij}(t)][1 - a_{ij}(t+1)]\rangle = 1 - 2p_{ij} + q_{ij}.$$
$$\text{(G4)}$$

Defining these elements, we can write the transition matrix P_{ij} equal in the form for each couple of nodes

$$P_{ij} = \begin{bmatrix} \frac{q_{ij}}{p_{ij}} & \frac{p_{ij}-q_{ij}}{p_{ij}} \\ \frac{p_{ij}-q_{ij}}{1-p_{ij}} & \frac{1-2p_{ij}+q_{ij}}{1-p_{ij}} \end{bmatrix}. \qquad \text{(G5)}$$

Once the values of α_i^* and β_i^* are determined, each transition matrix operates independently, signifying that each link trajectory evolves in isolation, unaffected by the behaviors of other links. This independence allows for a modular approach to analyzing the network dynamics. We can separately analyze the transition matrix for each link, studying its unique behavior and characteristics. Subsequently, we can aggregate these individual analyses to construct a comprehensive view of the entire system's evolution.

The eigenvalues and eigenvectors of these stochastic matrices are particularly informative, as they encapsulate critical aspects of the link dynamics, such as the response speed to changes within the network. To find the eigenvalues we solve the characteristic equation

$$\text{Det}(P_{ij} - \mu^{ij}I) = \begin{vmatrix} \frac{q_{ij}}{p_{ij}} - \mu^{ij} & \frac{p_{ij}-q_{ij}}{p_{ij}} \\ \frac{p_{ij}-q_{ij}}{1-p_{ij}} & \frac{1-2p_{ij}+q_{ij}}{1-p_{ij}} - \mu^{ij} \end{vmatrix} = 0, \quad \text{(G6)}$$

from which we obtain

$$\mu_{(1)}^{ij} = 1, \quad \mu_{(2)}^{ij} = \frac{p_{ij}^2 - q_{ij}}{p_{ij}^2 - p_{ij}} \equiv \mu_{ij}. \qquad \text{(G7)}$$

The stationary distribution, that is given by the eigenvector related to the unitary eigenvalue, is

$$\pi_{ij} = (p_{ij}, 1 - p_{ij}), \qquad \text{(G8)}$$

while the second eigenvalue can be interpreted as an indication of how fast the link converges to its stationary distribution, as remarked in the main text.

APPENDIX H: TEMPORAL EXPONENTIAL RANDOM GRAPHS AS MAXIMUM-ENTROPY PROBABILITY

In this Appendix, we demonstrate how temporal exponential random graph models (TERGMs), introduced by Hanneke in 2010 [13], are inherently structured as maximum-entropy probability distributions. Hanneke's seminal work proposed a dynamic extension of traditional exponential random graph models (ERGMs) to accommodate time-varying networks, employing a Markovian assumption to model the temporal evolution of networks:

$$\mathcal{P}(\mathbf{G_t}, \mathbf{G_{t-1}}, \mathbf{G_{t-2}}, \ldots, \mathbf{G_1}|\mathbf{G_0})$$
$$= P(G_t|\mathbf{G_{t-1}})P(G_{t-1}|\mathbf{G_{t-2}})\ldots P(G_1|\mathbf{G_0}). \quad \text{(H1)}$$

In this way, Hanneke develops a nontrivial generalization of traditional ERGMs. Specifically, he proposes that the conditional probability of observing a network state $\mathbf{G_t}$ given its previous state $\mathbf{G_{t-1}}$ can be modeled using an ERGM-like framework. The model is formalized as follows:

$$\mathcal{P}(\mathbf{G_t}|\mathbf{G_{t-1}}, \vec{\theta}) = \frac{e^{\vec{\theta}C(\mathbf{G_t}, \mathbf{G_{t-1}})}}{Z_t(\vec{\theta}, G_{t-1})}, \qquad \text{(H2)}$$

where $C(\mathbf{G_t}, \mathbf{G_{t-1}})$ is a vector of some properties depending on G_t and G_{t-1}, and $\vec{\theta}$ are the associated Lagrange multipliers. Similar to ERGMs, TERGMs can also be derived from an entropy maximization framework. We begin with a general consideration, and then we show how to reproduce the functional form of the TERGMs. In general the entropy of the trajectory reads as

$$\mathcal{S}[\mathcal{P}] \equiv -\sum_{\mathcal{G}} \mathcal{P}(\mathcal{G}) \ln \mathcal{P}(\mathcal{G}), \qquad \text{(H3)}$$

where $\mathcal{P}(\mathcal{G})$ is the probability distribution over all possible graph trajectories $\mathcal{G}$.

The goal is to find a probability distribution $\mathcal{P}(\mathcal{G})$ that maximizes this entropy subject to the constraint that the expected values of certain network properties $C_i(\mathcal{G})$ match their observed values C_i^*, and that the distribution is properly normalized. This optimization problem is typically solved using the method of Lagrange multipliers:

$$\frac{\partial}{\partial P(\mathcal{G})}\left[\mathcal{S}[\mathcal{P}(\mathcal{G})] + \alpha\left(1 - \sum_{\mathcal{G}} P(\mathcal{G})\right)\right.$$
$$\left. + \sum_i \theta_i\left(C_i^* - \sum_{\mathcal{G}} P(\mathcal{G})C_i(\mathcal{G})\right)\right]. \qquad \text{(H4)}$$

The solution of this problem is given by

$$P(\mathcal{G}|\vec{\theta}) = \frac{e^{-\sum_i \theta_i C_i(\mathcal{G})}}{Z(\vec{\theta})}, \qquad \text{(H5)}$$

that represents an ensemble of trajectories, where

$$Z(\vec{\theta}) = \sum_{\mathcal{G}} e^{-\sum_i \theta_i C_i(\mathcal{G})} \qquad \text{(H6)}$$

is the partition function.

Now, we can note that if

$$C_i(\mathcal{G}) = \sum_t^T \frac{C_i(G_t)}{T} \tag{H7}$$

is the sufficient statistic, Eq. (H5) becomes

$$P(\mathcal{G}|\vec{\theta}) = \prod_t \frac{e^{-\sum_i \theta_i \frac{C_i(G_t)}{T}}}{Z(\vec{\theta})} = \prod_t P(G_t|\vec{\theta}), \tag{H8}$$

where

$$P(G_t|\vec{\theta}) = \frac{e^{-\sum_i \theta_i \frac{C_i(G_t)}{T}}}{Z_t(\vec{\theta})} \tag{H9}$$

and

$$Z_t(\vec{\theta}) = \sum_{G_t} e^{-\sum_i \theta_i \frac{C_i(G_t)}{T}}. \tag{H10}$$

Equation (H8), corresponds to the scenario described in (9).

On the other hand, if we have

$$C_i(\mathcal{G}) = \sum_t^T \frac{C_i(G_t|G_{t-1})}{T}, \tag{H11}$$

we obtain that

$$P(\mathcal{G}|\vec{\theta}) = \prod_t P(G_t|G_{t-1}, \vec{\theta}), \tag{H12}$$

where

$$P(G_t|G_{t-1}, \vec{\theta}) = \frac{e^{-\sum_i \theta_i \frac{C_i(G_t|G_{t-1})}{T}}}{Z_t(\vec{\theta}, G_{t-1})}, \tag{H13}$$

with

$$Z_t(\vec{\theta}, G_{t-1}) = \sum_{G_t} e^{-\sum_i \theta_i \frac{C_i(G_t|G_{t-1})}{T}}. \tag{H14}$$

Equation (H13) is a particular case of the scenario indicated in Eq. (11) of the main text, and corresponds to the functional form of the TERGMs (H2).

Maintaining time-independent parameters is equivalent to assuming a stationary data-generating process. This assumption implies that we are dealing with a system in equilibrium, thereby inducing the resulting model described in (H12) a homogeneous Markov process.

As discussed in the main text, even if we consider the system at the equilibrium, we can still say something related to the behavior out of the equilibrium (i.e., relaxation time).

APPENDIX I: STRUCTURAL BREAK DETECTION: SYNTHETIC EXPERIMENTS

In this Appendix, we present the results from a series of synthetic experiments designed to evaluate the efficacy of our structural break detection approach. These experiments involve simulating temporal networks in a controlled environment where we can precisely manage external shocks. We begin by defining two matrices:

$$\begin{aligned} \mathbf{P} &\equiv (p_{ij}) \quad \forall\, i, j \in N, \\ \mathbf{Q} &\equiv (q_{ij}) \quad \forall\, i, j \in N. \end{aligned} \tag{I1}$$
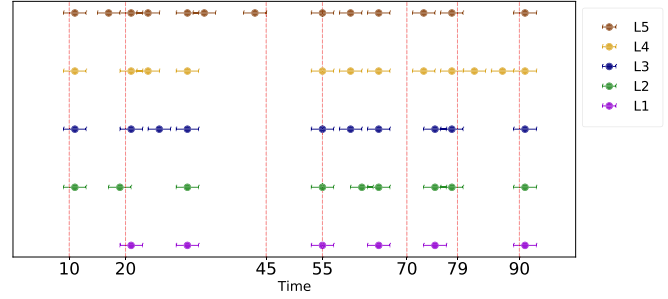

FIG. 4. Figure shows the impact of varying levels of perturbations for a synthetic temporal network. These perturbations are categorized into five levels: L1 (100%), L2 (500%), L3 (1000%), L4 (3000%), and L5 (5000%), as indicated in the legend. Points represent the predicted breaks for each level of perturbation, accompanied by corresponding error bars. The red vertical lines denote external shocks, which remain consistent across data generated with different intensities of the shock.

From Eq. (I1), we can derive the stochastic matrix for each pair of nodes, as described in Eq. (41). These matrices govern the evolution of links over time. Initially, we generate a graph based on **P** and, subsequently, we let the connections evolve according to the stochastic matrix defined for each link from elements i, j of **P** and **Q**.

At a designated point in the timeline, we introduce an external shock by altering the values in the matrices **P** and **Q** [Eq. (I1)]. Following this perturbation, the connectivity of the links continues to evolve based on the modified matrices. The aim of these experiments is to determine whether our structural break detection algorithm can effectively identify these induced perturbations. We introduce stochastic perturbations to the elements of a matrix to simulate shocks or noise. This is achieved by modifying each element by either increasing or decreasing its value by a random factor determined by a specified maximum percentage p. The perturbation process can be outlined as follows:

(i) Each element in the matrix undergoes a random adjustment process. With equal probability, the element is either increased or decreased. The magnitude of this adjustment is randomly selected up to a maximum of $p\%$ of the element's current value.

(ii) For increments, the adjustment factor is calculated as $1 + \frac{\text{random value}}{100}$, where *random value* is a uniformly distributed random number between 0 and p. For decrements, it is $1 - \frac{\text{random value}}{100}$.

(iii) The calculated factor is then multiplied by the element i, j of the matrix to apply the adjustment.

(iv) Postadjustment, any element values exceeding 1 are capped at 1, and those falling below or equal to 0 are set to a minimum threshold of 0.05 to maintain a defined range and avoid negative values.

The resultant matrix represents the original matrix subjected to random perturbations, simulating a shock.

Given the nature of the data-generating process behind the synthetic graph trajectory, the approach we choose to test is the one that considers memory effects; therefore, in the structural break detection algorithm, we use the likelihood related to the model described by the Hamiltonian $\mathcal{H}_2$, considering constraints at the node level.

Figure 4 demonstrates a robust alignment between predicted and actual shocks, underscoring the model's effectiveness in identifying structural breaks. The algorithm successfully captures numerous breaks in various scenarios but struggles with weaker perturbations and often overpredicts change points under stronger disturbances. Across all cases, there are instances where the model inaccurately forecasts change points, false positives that become more common as perturbation intensity increases. Intense shocks seem to induce significant system instability, leading to prolonged periods where the system's behavior deviates from the expected norm. These deviations often drive the algorithm to suggest breaks, even in the absence of actual external shocks. Therefore, the identification of such points does not necessarily signify errors; rather, it reflects our model's recommendation to apply different parameters for these segments, highlighting its adaptability in handling dynamic changes within the network.

[1] A. Vasilakos, Y. Zhang, and T. Spyropoulos, *Delay Tolerant Networks* (CRC Press, Boca Raton, FL, 2016).

[2] A. Pentland, R. Fletcher, and A. Hasson, Daknet: Rethinking connectivity in developing nations, Computer **37**, 78 (2004).

[3] M. E. Newman, The structure of scientific collaboration networks, Proc. Natl. Acad. Sci. USA **98**, 404 (2001).

[4] M. E. J. Newman, Coauthorship networks and patterns of scientific collaboration, Proc. Natl. Acad. Sci. USA **101**, 5200 (2004).

[5] G. Ghoshal, V. Zlatić, G. Caldarelli, and M. E. J. Newman, Random hypergraphs and their applications, Phys. Rev. E **79**, 066118 (2009).

[6] N. Masuda and R. Lambiotte, *A Guide to Temporal Networks* (World Scientific, Singapore, 2016).

[7] P. Grindrod and D. J. Higham, A dynamical systems view of network centrality, Proc. R. Soc. A **470**, 20130835 (2014).

[8] A. Clauset and N. Eagle, Persistence and periodicity in a dynamic proximity network, in *Proceedings of the DIMACS/DyDAn Workshop on Computational Methods for Dynamic Interaction Networks* (IEEE, Piscataway, NJ, 2007), arXiv:1211.7343.

[9] P. Leifeld, S. J. Cranmer, and B. A. Desmarais, Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals, J. Stat. Software **83**, 1 (2018).

[10] S. Wasserman and P. Pattison, Logit models and logistic regressions for social networks: I. an introduction to markov graphs and *p*, Psychometrika **61**, 401 (1996).

[11] P. W. Holland and S. Leinhardt, A dynamic model for social networks, J. Math. Sociol. **5**, 5 (1977).

[12] T. A. Snijders, The statistical evaluation of social network dynamics, Sociol. Methodology **31**, 361 (2001).

[13] S. Hanneke, W. Fu, and E. P. Xing, Discrete temporal models of social networks, Electron. J. Stat. **4**, 585 (2010).

[14] A. E. Clementi, C. Macci, A. Monti, F. Pasquale, and R. Silvestri, Flooding time in edge-markovian dynamic graphs, in *Proceedings of the Twenty-seventh ACM Symposium on Principles of Distributed Computing* (ACM, New York, 2008), pp. 213–222.

[15] P. Mazzarisi, P. Barucca, F. Lillo, and D. Tantari, A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market, Eur. J. Oper. Res. **281**, 50 (2020).

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

[17] N. Eagle, A. Pentland, and D. Lazer, Inferring friendship network structure by using mobile phone data, Proc. Natl. Acad. Sci. USA **106**, 15274 (2009).

[18] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina *et al.*, High-resolution measurements of face-to-face contact patterns in a primary school, PLoS ONE **6**, e23176 (2011).

[19] H. Akaike, A new look at the statistical model identification, IEEE Trans. Automat. Contr. **19**, 716 (1974).

[20] T. Squartini and D. Garlaschelli, *Maximum-entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics* (Springer, Berlin, 2017).

[21] E. T. Jaynes, Information theory and statistical mechanics, Phys. Rev. **106**, 620 (1957).

[22] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, The statistical physics of real-world networks, Nat. Rev. Phys. **1**, 58 (2019).

[23] J. Park and M. E. J. Newman, Statistical mechanics of networks, Phys. Rev. E **70**, 066117 (2004).

[24] D. Garlaschelli and M. I. Loffredo, Fitness-dependent topological properties of the world trade web, Phys. Rev. Lett. **93**, 188701 (2004).

[25] D. Garlaschelli and M. I. Loffredo, Maximum likelihood: Extracting unbiased information from complex networks, Phys. Rev. E **78**, 015101(R) (2008).

[26] P. Sarkar and A. W. Moore, Dynamic social network analysis using latent space models, ACM SIGKDD Explorations Newsletter **7**, 31 (2005).

[27] P. Erdős and A. Rényi, On random graphs, Publicationes Mathematicae **6**, 290 (1959).

[28] S. De Ridder, B. Vandermarliere, and J. Ryckebusch, Detection and localization of change points in temporal networks with the aid of stochastic block models, J. Stat. Mech. (2016) 113302.

[29] L. Peel and A. Clauset, Detecting change points in the large-scale structure of evolving networks, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 29 (AAAI, Washington, DC, 2015).

[30] M. Bhattacharjee, M. Banerjee, and G. Michailidis, Change point estimation in a dynamic stochastic block model, J. Machine Learning Res. **21**, 1 (2020).

[31] Z. Zhao, L. Chen, and L. Lin, Change-point detection in dynamic networks via graphon estimation, arXiv:1908.01823.

[32] A. Scott and M. Knott, A cluster analysis method for grouping means in the analysis of variance, Biometrics **30**, 507 (1974).

[33] N. J. Gotelli, Null model analysis of species co-occurrence patterns, Ecology **81**, 2606 (2000).

[34] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Dover, New York, 2007).