Non-Markovian temporal networks with auto- and cross-correlated link dynamics

Oliver E. Williams,[1,*] Piero Mazzarisi,[2,†] Fabrizio Lillo,[3,2,‡] and Vito Latora[1,4,5,*]

[1]*School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom*
[2]*Scuola Normale Superiore, Piazza dei Cavalieri, 7, 56126 Pisa, Italy*
[3]*Department of Mathematics, University of Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy*
[4]*Dipartimento di Fisica ed Astronomia, Università di Catania and INFN, I-95123 Catania, Italy*
[5]*Complexity Science Hub Vienna (CSHV), A-1080 Vienna, Austria*

(Received 18 January 2021; revised 27 October 2021; accepted 3 February 2022; published 7 March 2022)

Many of the biological, social and man-made networks around us are inherently dynamic, with their links switching on and off over time. The evolution of these networks is often observed to be non-Markovian, and the dynamics of their links are often correlated. Hence, to accurately model these networks, predict their evolution, and understand how information and other relevant quantities propagate over them, the inclusion of both memory and dynamical dependencies between links is key. In this article we introduce a general class of models of temporal networks based on discrete autoregressive processes for link dynamics. As a concrete and useful case study, we then concentrate on a specific model within this class, which allows to generate temporal networks with a specified underlying structural backbone, and with precise control over the dynamical dependencies between links and the strength and length of their memories. In this network model the presence of each link is influenced not only by its past activity, but also by the past activities of other links, as specified by a coupling matrix, which directly controls the causal relations, and hence the correlations, among links. We propose a maximum likelihood method for estimating the model's parameters from data, showing how the model allows a more realistic description of real-world temporal networks and also to predict their evolution. Due to the flexibility of maximum likelihood inference, we illustrate how to deal with heterogeneity and time-varying patterns, possibly including also nonstationary network dynamics. We then use our network model to investigate the role that, both the features of memory and the type of correlations in the dynamics of links have on the properties of processes occurring over a temporal network. Namely, we study the speed of a spreading process, as measured by the time it takes for diffusion to reach equilibrium. Through both numerical simulations and analytical results, we are able to separate the roles of autocorrelations and neighborhood correlations in link dynamics, showing that not only is the speed of diffusion nonmonotonically dependent on the memory length, but also that correlations among neighboring links help to speed up the spreading process, while autocorrelations slow it back down. Our results have implications in the study of opinion formation, the modeling of social networks, and the spreading of epidemics through mobile populations.

DOI: 10.1103/PhysRevE.105.034301

I. INTRODUCTION

Much of the world we experience is governed by interactions. Networks provide a natural way of modeling these interactions, and as such the study of networks has been central to the understanding of both natural phenomena and man-made systems. Observably, many of the networks around us change over time, as the interactions and connections that define them come and go. Human contacts and social interactions do not last forever [1–3], roads between towns and cities can be closed or new ones build [4,5], financial or economic agents trade each day with different counterparts [6], and even our brains undergo significant changes throughout our lives [7–10]. Real-world examples of temporal networks are often found to have a set of very well-defined structural and temporal features, many of which play key roles in determining the dynamics and functioning of the systems for which they form the backbone [11–18]. Various models have been recently proposed to replicate such features. For instance, models of human face-to-face interactions often rely on the assumption that the agents move as random walkers in a physical space and create a link whenever they are closer than a certain distance [2,19]. Other models take a slightly more abstract approach, introducing the notion of node activity to control the presence of links [20–22]. The adaptations and extensions of these models do directly specify the presence of empirically observed features such as memory, by which we here mean a dependence on some finite number of past states. Indeed, memory has been seen to play an important role in many real-world networks [8,23–25]. It can affect the dynamics of social interactions [26] and the controllability of temporal networks [27] and can also turn useful in the definition of flow-based communities [28–30]. An area of study in

*Corresponding authors: o.edgar.williams@gmail.com; v.latora@qmul.ac.uk
†piero.mazzarisi@sns.it
‡fabrizio.lillo@unibo.it

which memory has received a large attention is its relation to spreading processes [31,32]. When considering the spreading of an infection over a network, the presence of memory in the link activities can have a considerable effect on the rate of spreading of the disease, and can even cause dramatic changes to the epidemic threshold [33–36]. In diffusive processes, memory directly induces the slowing, or acceleration, of the spread of information over the network [30,31,37–40]. This has been studied in the context of higher-order networks, and is often understood to be a result of the correlated bursts, and the induced lasting interactions that the nonexponential interevent times, which define memory, necessitate [31,41–45]. What has been done, however, does not form a full picture. The presence of memory in the links that make up a network naturally means that the state of each of these links at a given time can depend on the past activity of the link. It is common in real networks to have pairs of links which are correlated with each other. Indeed, it seems natural to assume that links in a temporal network can have memory of each others past, rather than simply their own. The connections between the rate at which information spreads across a network and the memory of links are deep, as are the connections between memory and link correlations. However, the way in which interlink correlations and memory interact, and the effects this interaction has on spreading and other dynamical processes occurring over temporal networks is not well understood.

The goal of this article is twofold. We first introduce a general class of models of temporal networks, which are based on a discrete autoregressive mechanism for link dynamics. Then, as a case study, we concentrate on a specific generative model for temporal networks within this class in which the backbone structure, temporal correlations, and memory are all taken into account but can be precisely and separately controlled. We will also present a method for inferring the key parameters of the model from empirical data, simultaneously highlighting both the ability of the model to describe real systems and the role played by both memory and cross-interactions of links in the dynamics of real-world networks and in the forecasting of links. Then we extend further the range of applicability by showing how to account for heterogeneity and time-varying patterns in link dynamics, again validating the proposed generalization on data. The second goal of the article is to investigate how the interplay of the three key properties of a temporal network, namely, the structure of its underlying backbone, the correlations between the evolution of its links, and the memory of its own past states, impacts dynamical processes over the network. In particular, we will study the way in which these properties affect a process of diffusion over a temporal network.

This paper is organized as follows. In Sec. II we introduce a general class of models of temporal networks based on discrete autoregressive processes. As a concrete case, in Sec. III we consider a specific model within this class that allows for a controlled description and treatment of the cross-interactions in link dynamics, the so-called *correlated discrete autoregressive network* model of order p, or in short the CDARN(p) model. We discuss our model in the context of other existing generative models for temporal networks, and we explain how the controllability, flexibility, and analytical tractability of the model fill an important gap in the literature. In Sec. IV we show how the CDARN(p) can be applied to model real temporal networks presenting a maximum likelihood estimation framework to infer the key parameters of the model from empirical data. In this section the role played by both memory and cross-interactions of links in the dynamics of real-world networks will be evident, as well as the ability of the CDARN(p) model to effectively reproduce real features of temporal networks. Hence, we point out that including cross-interactions allows us to better describe the evolution of real-world temporal networks, specifically by better predicting the appearance of a link between a given couple of nodes. Moreover, we show how heterogeneous or time-varying parameters can be considered in our setting, due to the flexibility of maximum likelihood inference. In particular we show the role played by both heterogeneous and time-varying patterns in estimation of and forecasting with the CDARN(p) model. In Sec. V we consider processes occurring over temporal networks. As an example of a network process, we study diffusion over temporal networks generated by the the CDARN(p) model. We implement the CDARN(p) model on a number of backbone topologies taken from real-world systems, and we present numerical and analytical results concerning how the various features of the temporal network affect the diffusion process occurring over it. In particular, we show that the average time taken for diffusion to reach equilibrium on these networks is generally nonmonotonically dependent on the memory length, in agreement with recent findings regarding a different type of process, namely, epidemic spreading in temporal networks with only self-correlated links activities [35]. Here, however, we find that the time taken to reach equilibrium is additionally highly dependent on how links in the temporal network are correlated. Moreover, and more importantly, we study in detail the effects of link cross-correlations on diffusion. We are able to explain the dependence of the time to reach equilibrium on the types of correlations between the activities of links in the temporal network. Specifically we find that when correlations between neighboring links are strengthened when compared to link autocorrelations, diffusion speeds up. This is a surprising complement to some recent works: while autocorrelation in links slows down diffusion, as explained by the induced burstiness of the link processes, correlations between neighbor links speeds it up [30,31,35,37–39,41,46–48].

Overall, our results demonstrate that the topology of a temporal network interacts in a complex way with the dynamical properties (correlation and memory) of its links. Our model provides a framework for systematic investigation of this delicate interplay and for the description of real systems: its simplicity allows for efficient numerical simulation and analytical tractability, and its flexibility allows us to explore and understand a wide range of observable phenomena relating to diffusion over temporal networks. Further, it proves to be useful when investigating temporal networks observed in the real world, where we cannot assume any ability to study the effects of temporal correlations and memory in isolation, thus making it an ideal building block for further studies of empirical systems.

II. A GENERAL CLASS OF DISCRETE AUTOREGRESSIVE NETWORK MODELS

Models for temporal networks in which links are governed by a possibly correlated set of stochastic processes allow for a great deal of control over various aspects of their output, but can run the risk of being too abstract, and thus their use in describing empirical systems can be limited. For example, temporal networks in which links are specified to have an interevent time with a Weibull distribution have been seen to reflect empirical findings with respect to infection spreading and clearly imply memory in the network; however, it is not clear that they are a good model for temporal networks in more general settings [49]. Activity-driven network models [50], in particular those versions with link reinforcement process [21], allow us to describe non-Markovian memory in links. Nevertheless, it is unclear how to estimate activity-driven models on empirical data. State-space models of temporal networks (see, for example, [51–53]) describe nodes as evolving in a latent Euclidean space and interacting depending on their "physical" distances in such a space. The resulting link dynamics can display both non-Markovian memory and cross-correlations among links. However, there is no explicit control on both features. Modeling temporal networks as Markov chains of generic memory order [18,34] allows us to characterize the memory patterns displayed by empirical data, however, at the expense of high computational costs and the use of a large number of parameters. Finally, maximum entropy models of temporal networks [54] permit us, in principle, to describe many patterns of link dynamics, also having a high level of control on the features of the output network, as well as allowing applications to empirical data. Recently introduced Markovian models of temporal networks [6,35] based on some opportune generalization of the discrete autoregressive process [55] are to all effect maximum entropy models, as shown in [56]. Here we show how the multivariate and non-Markovian generalization of the discrete autoregressive mechanism [55] is suited for a *general* description of the auto- and cross-correlation structure of temporal networks described as time series of adjacency matrices. Such a generalization allows us to define an alternative class of models of temporal networks, which is highly flexible, largely controllable, and analytically tractable at the same time.

The discrete autoregressive process DAR(p) [55], whose properties have been largely studied in the statistics and econometrics literature [57–59], describes the persistence pattern of a stochastic process by means of the discrete autoregressive (copying) mechanism as

$$X_t = Q_t X_{t-Z_t} + (1 - Q_t) Y_t, \tag{1}$$

with the following:

(1) $Q_t \sim \mathcal{B}(q)$ Bernoulli random variable with success probability q

(2) Non-Markovian memory described by a random variable Z_t, which picks an integer value τ ranging from 1 to p with probability z_τ (i.e., *memory kernel*), such that $\sum_{\tau=1}^{p} z_\tau = 1$

(3) Bernoulli marginal $Y_t \sim \mathcal{B}(y)$ with success probability y.

The DAR(p) model in Eq. (1) captures the *positive* autocorrelation of a binary time series with memory of generic order p by means of the copying mechanism mediated by the Bernoulli random variable Q_t. It is quite natural moving from the description of a single binary time series to the multivariate case of adjacency matrices, thus exploiting the flexibility of the framework, to account also for the cross-interactions of links and time-varying patterns in temporal networks. For practical reasons, the multivariate generalization of the DAR(p) process in Eq. (1) allows us to define an alternative class of temporal networks, the so-called *discrete autoregressive network models*, which combine the mechanism of copying from the past with the sampling of links according to some marginal, which is Bernoulli in the simplest case. In particular, the latter can be interpreted to all effects as the non-Markovian dynamic generalization of the Erdős-Rényi random graph model when one parameter is controlling for the density of the network.

In order to precisely define the discrete autoregressive network models, let us consider a temporal adjacency matrix $\underline{\underline{A}}_t = \{a_t^{ij}\}$, with $t = 1, 2, \ldots$. If each link (i, j) is labeled by a single index $\ell \equiv (i, j)$, we can consider the vectorization $\underline{X}_t \equiv \{a_t^\ell\}^{\ell=1,\ldots,L}$ of the adjacency matrix $\{a_t^{ij}\}^{(i,j)\in B}$ of the network snapshot at time t, where L is the number of *possible* links belonging to some subset B (i.e., the so-called *backbone* of the temporal network) of all the $N(N-1)/2$ couples of nodes. We then consider the following discrete autoregressive multivariate process ($\ell = 1, \ldots, L$):

$$X_t^\ell = Q_t^\ell X_{t-Z_t^\ell}^{M_t^\ell} + \left(1 - Q_t^\ell\right) Y_t^\ell \tag{2}$$

with the following:

(1) $Q_t^\ell \sim \mathcal{B}(q_t^\ell)$ Bernoulli random variable with, in general, link-specific time-varying probability q_t^ℓ

(2) Non-Markovian memory described by a set of random variables Z_t^ℓ which pick value τ running from 1 to p with probability z_τ (i.e., *memory kernel*), such that $\sum_{\tau=1}^{p} z_\tau = 1$;

(3) Link cross-interactions described by a set of random variables M_t^ℓ which pick values from 1 to L according to each row of a coupling matrix $\underline{\underline{C}} \equiv \{c^{\ell\ell'}\}$, a row stochastic (i.e., $\sum_{\ell'} c^{\ell\ell'} = 1$) matrix, which characterizes the correlations between pairs of links.

(4) Bernoulli marginals $Y_t^\ell \sim \mathcal{B}(y_t^\ell)$ with, in general, link-specific time-varying probability y_t^ℓ.

The formulation in Eq. (2) is very general, accounting for time-dependent persistence patterns of links, with non-Markovian memory, cross-interactions mediated by the coupling matrix $\underline{\underline{C}}$, and possibly time-varying marginal probabilities. In practice, some parametrization needs to be considered, and it is possible to reduce the complexity of the model and use it to focus, one by one, on the various specific features of temporal networks.

Previous works have investigated the role of non-Markovian memory in models of temporal networks that can now be seen as limiting cases of the most general framework proposed in Eq. (2). However, a very important aspect, which has not yet received the deserved attention, is the modeling of cross-interactions of links. For instance, the dynamics of spreading processes on temporal networks with memory has been investigated in the DARN(p) model, a non-Markovian model that can be seen as a limiting case of the

model in Eq. (2) with constant parameters and with diagonal coupling matrix $\underline{\underline{C}}$, i.e., under the very strong assumption that only autocorrelations in the link dynamics are present [35]. The authors of Ref. [6] have instead proposed an empirical application of link inference to the interbank market. The have considered a model similar to that in Eq. (2) with Markovian link-persistence patterns, again without explicit cross-correlations of links, but combined with node-specific time-varying marginals. Finally, Granger causality has been investigated in a model with two time series, which corresponds to a *bivariate* case (i.e., $\ell = 1, 2$) of the model in Eq. (2) with constant parameters [60].

Here we focus on a crucial feature of real-world temporal networks that has received less attention from a modeling point of view, namely, the cross-interactions of links (i.e., the presence of dependencies in the time evolution of pairs of different links). In particular, we will consider the model in Eq. (2) with general memory kernels and opportune parametrizations of the coupling matrix $\underline{\underline{C}}$. In this way we will be able to study auto- and cross-correlated link dynamics combined with non-Markovian memory, and in the presence of a backbone network defining which links may be present or not.

We point out the richness, versatility, and controllability of our model of temporal networks, together with its low computational costs in empirical applications, due to maximum likelihood methods for inference. In Sec. III we will first consider the CDARN(p) model, a simplified version of the model in Eq. (2) with constant parameters $q_t^\ell = q$ and $y_t^\ell = y\ \forall t, \ell$. Then, in Sec. IV due to the high flexibility of the proposed framework, we will relax the last assumption by allowing for heterogeneous parameters (i.e., link-specific q^ℓ and y^ℓ) and we will exploit *local* likelihood methods [61] to deal with time-varying parameters q_t, c_t, and y_t. Finally, the analytical tractability of our approach will come to light in Sec. V in the study of the dynamics of spreading processes over temporal networks generated by the model.

III. The CDARN(p) MODEL OF TEMPORAL NETWORKS

Here we consider a simplified version of the general model of temporal networks in Eq. (2), which is rich enough to describe non-Markovian memory patterns, with precisely controlled strength and length of the memory, while also reproducing a key feature of real-world networks, namely, correlations between the evolutions of links over time, as produced by dependencies between their dynamics. Furthermore, we want to keep such a model as simple as possible, with a small number of parameters, thus permitting also easy application to empirical data by fitting the parameters of the model on graph sequences from the real world. Hence, we take the general setting given by Eq. (2) and consider a particular parametrization that reflects the presence of two key features of real systems [3,7,8,38,62]: (1) the existence of an underlying restriction, a so-called network "backbone" on which links can occur and (2) the presence of cross-correlations in link dynamics, i.e., of dynamical dependencies in the temporal activities of different links.

We introduce the so-called *correlated discrete autoregressive network* model of order p, or in short the CDARN(p) model, which describes the dynamics of links with an included mechanism for copying from the past: at each time, a link (or no-link) can be copied from the past, either of the link itself or the past of some other link on the backbone, or sampled according to a Bernoulli marginal (Erdős-Rényi model). This point in the past, from 1 to p steps behind, is then randomly selected with uniform probability. The model is, in effect, the multivariate generalization of the DARN(p) model [35] with non-Markovian memory and both self- and cross-interactions of links. In the following, for clarity, we briefly review DARN(p) before introducing the CDARN(p) model, while in the next section we show how to estimate the model on real data, to infer the key parameters of the network dynamics. Once estimated on data, the model can be used also for link prediction. Moreover in Sec. V we will show how the model parameters can be varied to study in a systematic and controlled way the role that memory in the underlying temporal network has on the rate at which information, or some other quantity spreads throughout a system whose interactions change over time.

A. The basic model

The DARN(p) model, originally introduced in [35], generates a temporal network with precisely controlled memory features in the temporal sequence of each link. Namely, the model considers N nodes and assigns to each of the $N(N-1)/2$ pairs of nodes the presence or absence of a link as ruled by independent, identical DAR(p) processes (discrete autoregressive processes of order p) [6,55,63,64]. In this way each link either will, at each time step, be generated randomly with some fixed probability, or will copy a randomly chosen state from its past p iterations. In terms of random variables, this gives us a temporal adjacency matrix $\underline{\underline{A}}_t = \{a_t^{ij}\}$, with $t = 1, 2, \ldots$, where each link (i, j), with $i, j = 1, \ldots, N$ is governed by the process

$$a_t^{ij} = Q_t^{ij} a^{ij}_{(t-Z_t^{ij})} + \left(1 - Q_t^{ij}\right) Y_t^{ij}, \tag{3}$$

where, for each link (i, j) and time t, Q_t^{ij}, Y_t^{ij}, and Z_t^{ij} are random variables. In particular, $Q_t^{ij} \sim \mathcal{B}(q)$ and $Y_t^{ij} \sim \mathcal{B}(y)$ are Bernoulli random variables, while Z_t^{ij} is some random variable which picks integers in the range $\{1, \ldots, p\}$. Note that no restriction is imposed to the memory kernel controlling for the probability of picking the integers in the range $\{1, \ldots, p\}$, as long as the probability sums to one. For example, an uniform kernel describes equal probabilities of picking past observations from lag 1 to lag p, whereas an exponential kernel describes probabilities exponentially decaying to zero as the lag is increasing. Without loss of generality, here we take $Z_t^{ij} \sim \text{Uniform}(1, p)$. The networks created by the DARN(p) model are undirected, and clearly non-Markovian, with precise memory p.

The DARN(p) model assumes that links can occur between any two nodes. This is not always the case in real-world networks, where certain links may be unfeasible, or simply impossible. For example, a plane may not be allowed to fly between two particular airports, or a doctor may be responsible for a small number of patients, and therefore not interact with others. We therefore say that these temporal networks

have a "backbone": a fixed set of possible links which restrict the networks evolution. With this in mind we make our first modification leading to a more general framework. First, we define a *backbone network* with L links described by a static $N \times N$ adjacency matrix $\underline{\underline{B}} = \{b^{ij}\}$. Then a temporal network on this backbone is represented by a $N \times N$ time-varying adjacency matrix $\underline{\underline{A}}_t = \{a_t^{ij}\}$, so that $a_t^{ij} = 0$ for all t if $b^{ij} = 0$, while if $b^{ij} = 1$ then the link (i, j) can exist for any value of t. In this way the presence of links can be appropriately limited to reflect reality.

Since links in the DARN(p) model are generated by independent processes, there can be autocorrelations in the temporal activity of each link but no cross-correlations between different links. Conversely, correlations among different links are a natural feature of many systems. To further our earlier analogy, an airline is unlikely to schedule two flights between the same airports in close proximity to each other but may prefer to schedule flights at appropriate times to make connections. Similarly doctors may see patients in a particular order each day, even if the duration of each interaction is not so consistent. In order to allow for such correlations, we introduce our second modification: when a link in a DARN(p) model would pick from its own memory, we now allow it to pick a link from the network at random, possibly itself again, and copy a randomly chosen state of that link instead. In this way, the dynamics of each link (i, j) that belongs to the network backbone is governed by the process

$$a_t^{ij} = Q_t^{ij} a^{M_t^{ij}}_{(t-Z_t^{ij})} + \left(1 - Q_t^{ij}\right) Y_t^{ij} \tag{4}$$

with $i, j = 1, \ldots, N$ and such that $b^{ij} = 1$, and where at each time t, M_t^{ij} is a (categorical) random variable which associates to link (i, j) another link (i', j') among links which are present in the backbone $\underline{\underline{B}}$, with an assigned probability distribution. Note that for each time t and link (i, j), M_t^{ij} is independent and identically distributed. That is to say, if a link is copied from the past of another link, then which link it chooses is completely independent on either the time, or the existence of any other link.

Hence, the CDARN(p) model in Eq. (4) relies on the following input parameters. The first ingredient is the $N \times N$ adjacency matrix $\underline{\underline{B}}$ describing the structure of the underlying network backbone of N nodes and L links, i.e., defining which pairs of nodes can be connected by links and which pairs cannot. The backbone has density $D = 2L/N(N-1)$, if the network is undirected. However, at each time not all the links of the backbone are necessarily present, and the average link density within the backbone is controlled by the parameter $0 < y < 1$. Moreover $0 \leqslant q \leqslant 1$ and $p = 1, 2, \ldots$ are respectively the strength and length of the memory component of the dynamics of the temporal network. Finally, the structure of interactions among links is captured by the random variable M_t^{ij}, which can be described by a $L \times L$ link coupling matrix $\underline{\underline{C}}$, characterizing the correlations between pairs of links. Labeling links with a linear index $(i, j) \mapsto \ell$, $(i', j') \mapsto \ell'$, with $\ell, \ell' = 1, 2, \ldots, L$ (see the Appendix A 1 for a full explanation), then M_t^{ij} can be characterized by the probabilities

$$\text{Prob}(\ell \text{ draws from } \ell') = c^{\ell\ell'}.$$

These probabilities define a $L \times L$ row-stochastic matrix $\underline{\underline{C}} = \{c^{\ell\ell'}\}$, which we call the *coupling matrix*. By tuning the entries of this matrix we can specify the dependencies among links existing in our temporal network. In practice, for each possible link (i, j) and at each time t, M_t^{ij} will select another link (i', j') among a set of possible links associated to (i, j), as given by matrix $\underline{\underline{C}}$. Then the presence of the term $a^{M_t^{ij}}_{(t-Z_t^{ij})}$ in Eq. (4) represents the state of link (i', j') at one of the previous p temporal steps, and so will allow link (i, j) to copy its state at time t from one of the p past states of link (i', j'). This is similar to building the line graph associated with the original network, but in the temporal case and restricting to pairs of links which are on the backbone. The choice of the coupling matrix $\underline{\underline{C}}$ is a crucial part of the CDARN(p) model, as this defines which links dynamics are correlated. Since the matrix has a large number of entries, it is advisable to choose a parsimonious representation depending on a small number of parameters. There are many ways one could structure the matrix $\underline{\underline{C}}$, which we refer to as "coupling models." Here we will focus on the following three simple approaches: (1) only link autocorrelations but no cross-correlations between different links, (2) links are coupled to all other neighboring links in the network backbone (as defined by $\underline{\underline{B}}$) with equal strength, and (3) links are coupled to all other links in the backbone with equal strength.

To summarize, given a backbone $\underline{\underline{B}}$ with L links, we have the three following coupling models:

(1) The *no cross-correlation (NCC)* coupling model, where the coupling matrix reads $\underline{\underline{C}} = \underline{\underline{I_d}}$ (the identity matrix).

(2) The *local cross-correlation (LCC)* coupling model, where the entries of the coupling matrix can be written as $c^{\ell\ell'} = (1-c)\delta(\ell, \ell') + \chi(\ell' \in \partial_B \ell)\, c/|\partial_B \ell|$, for coupling strength c. Here χ is the indicator function, $\delta(\ell, \ell') = 1$ if $\ell = \ell'$ and 0 otherwise, and $\partial_B \ell$ is the neighborhood of link ℓ in backbone $\underline{\underline{B}}$, i.e., for $\ell = (i, j)\partial_B \ell = \{\ell' = (i', j') : b^{i'j'} = 1$ and $i' \in \ell$ or $j' \in \ell\}$.

(3) The *uniform cross-correlation (UCC)* coupling model, where the entries of the coupling matrix can be written as $c^{\ell\ell'} = (1-c)\delta(\ell, \ell') + [1 - \delta(\ell, \ell')]c/(L-1)$, for coupling strength c.

Notice that the parameter $0 \leqslant c \leqslant 1$ in the second and third coupling model allows us to tune the contribution of the cross-correlations with respect to that of the autocorrelations. In particular the NCC model is the special case of the LCC and UCC models with coupling strength $c = 0$.

Considering all the building blocks, we then have our full model, which we name the *correlated discrete autoregressive network* model of order p, or in short CDARN(p) model. This model has the advantage of being able to introduce both auto- and cross-correlations in the link activities in a controlled way, allowing for a more realistic description of real-world systems. It also retains much of the simplicity and tractability of the DARN(p) model. Indeed, three of the key features of the DARN(p) model that allow us to study a range of phenomena are exactly the same. Namely, the (unconditional) probability

of observing a link (restricted to the feasible connections over the backbone) is y, similarly to ER graphs (see the Appendix A 2) [65]. Moreover, in the limit of long memory, as given by large p, the model is identical to a sequence of uncorrelated ER graphs (see the Appendix A 3). Finally, the interevent time distribution, also known as intercontact time in human communication networks, is (*approximately*) exponential, with a timescale that is bounded from above by the DARN(p) model (see the Appendix A 18).

In summary, our model generates temporal networks $\underline{\underline{A}}_t, t = 1, 2, \ldots$, with precisely controlled coupling among links, given the following set of control parameters: network backbone as specified by matrix $\underline{\underline{B}}$, link density y, memory strength q, memory length p, and link coupling matrix $\underline{\underline{C}}$. For our purposes we will assume that the links in the temporal network are undirected, we do this by identifying $a_t^{ij} = a_t^{ji}$. Implicitly the backbone in any network will be taken as undirected, implying that only symmetric matrices $\underline{\underline{B}}$ will be considered. The extension to directed networks is, however, straightforward. Finally, an important assumption of the CDARN(p) model is that the parameters p, q, and y are the same for all the links of the backbone. Of course, this choice is a simplification motivated only by the need for control over the dynamics with a small number of parameters, for both empirical application to real-world networks and analytical study of the dynamics of spreading processes over temporal networks. The CDARN(p) model can in fact be easily generalized to the case of link-specific or time-varying parameters, as done in the section below or, for instance, in the simpler DARN(1) model presented in [6].

B. Model generalizations

Real-world networked systems can be characterized by heterogenous, i.e., link- or node-specific, and/or time-varying, patterns of link dynamics. Realistic models of temporal networks need to be able to capture such patterns when we aim to replicate empirically observed network dynamics. Such a generalization can easily be accounted for within our framework by considering link-specific parameters $y \to y^\ell$ and $q \to q^\ell$, or promoting constant parameters q, c, and y to time-varying parameters q_t, c_t, and y_t, then introducing a method to estimate them.

We have therefore considered an entire set of models that generalize the simplified version of the CDARN(p) model, by relaxing, step by step, some homogeneity assumptions:

(1) Ceteribus paribus, the probability of success y of the Bernoulli marginal is not anymore equal for all links, but each link ℓ is described by a different probability y^ℓ, this allowing for a different existence probability for each link

(2) Ceteribus paribus, the probability of copying from the past q is not anymore equal for all links, but each link ℓ is more or less persistent depending on a specific parameter q^ℓ

(3) Parameters q, c, and y are not constant anymore during the evolution of the network, but they can change in time, in order to capture the presence of time-varying, possibly nonstationary, patterns in link dynamics, e.g., link density and/or correlations depending on the time of the day.

IV. NETWORK MODEL INFERENCE AND LINK PREDICTION

In this section we introduce a method for estimating the parameters of our models from real data. We will first focus on the simplified version of the CDARN(p) model defined in Eq. (4). The strength of our approach comes to light, since the CDARN(p) model can be estimated on data by using maximum likelihood methods [61], thus inferring case by case the role played in the real world by both auto- and cross-correlations of links. Then, due to the flexibility of the maximum likelihood approach for inference, we show that our method can also be used to estimate the parameters of generalized models accounting for *heterogeneous* or *time-varying* parameters, that can better capture the dynamics of real-world networked systems. Last but not least, we prove empirically that the inclusion of cross-interactions in the description of the link dynamics is no small matter: cross-interactions of links are essential in describing various types of real-world temporal networks. In particular, we show, through a link prediction study, that such correlation patterns are, indeed, present in networks from the real world.

A. Parameter estimation

Assume we have observed a time series of network snapshots $\{a_t^{ij}\}_{t=p+1,\ldots,T}^{i,j=1,\ldots,N}$ with given initial p conditions $\{a_t^{ij}\}_{t=1,\ldots,p}^{i,j=1,\ldots,N}$, then we ask for the values of parameters in Eq. (4) which best describe the evolution of the temporal network. Here we aim to obtain a *point estimate* of the parameters, which is, from a Bayesian inference perspective, the value maximizing the posterior probability of parameters given the data. By referring to the set of parameters as θ, due to the Bayes theorem, we can write

$$\mathbb{P}(\theta|\underline{\underline{A}}) \propto \mathbb{P}(\underline{\underline{A}}|\theta)\mathbb{P}(\theta),$$

where $\underline{\underline{A}} \equiv \{\underline{\underline{A}}_t\}_{t=1,\ldots,T}$. Without prior information on the parameters, we can assume uniform prior distribution $\mathbb{P}(\theta)$. Thus, the point estimation corresponds to maximizing the likelihood of data under the model with parameters θ, i.e., $\mathbb{P}(\underline{\underline{A}}|\theta)$, namely, the maximum likelihood estimator (MLE) of the CDARN(p) model. The *likelihood* of data under the CDARN(p) model reads as

$$\mathbb{P}(\underline{\underline{A}}_{p+1}, \ldots, \underline{\underline{A}}_T | \underline{\underline{A}}_1, \ldots, \underline{\underline{A}}_p, q, c, y) = \prod_{t=p+1}^{T} \mathbb{P}(\underline{\underline{A}}_t | \underline{\underline{A}}_{t-1}, \ldots, \underline{\underline{A}}_{t-p}, q, c, y), \quad (5)$$

by using the Markov property, and with $\{q, c, y\}$ the model parameters. The likelihood of the Markov chain then corresponds to the product of $T - p$ conditionally independent transition probabilities, each one describing the likelihood of a network snapshot given the previous p observations, because of the non-Markovian memory of the process. The MLE of the parameters is thus obtained by maximizing Eq. (5), or equivalently the *log-likelihood* $\mathbb{L}(q, c, y) \equiv$

$\log \mathbb{P}(\{\underline{\underline{A}}_t\}_{t=p+1,\dots,T} | \{\underline{\underline{A}}_\tau\}_{\tau=1,\dots,p}, q, c, y)$, that is,

$$\arg\max_{y,c,q} \sum_{t=p+1}^{T} \log \mathbb{P}(\underline{\underline{A}}_t | \underline{\underline{A}}_{t-1}, \dots, \underline{\underline{A}}_{t-p}, q, c, y),$$
$$q, c, y \in [0, 1]. \tag{6}$$

The solution to Eq. (6) is the MLE $\{\hat{q}, \hat{c}, \hat{y}\}$ of the CDARN(p) model. The explicit formulas for the MLE are found in the Appendix A 17. Notice that the order p of the memory of the Markov chain can be selected by finding the integer value which maximizes the likelihood of data under CDARN(p), since the number of parameters of the CDARN(p) model is the same, independently from the order p, and thus there is no need to penalize the use of more parameters. See [66] for a study on the optimal selection of the (non-Markovian) memory in temporal networks.

In empirical applications, some networked systems may display some time-varying density pattern, related, for example, to the activity of nodes, which may, crucially, affect the estimation of the parameters q and c. For example, in the presence of a seasonality pattern, i.e., a network density depending on the time of the day, considering a constant density parameter tends to overestimate correlations, thus the MLE $\hat{q}$ and $\hat{c}$. In our framework, any variation of network density can be taken into account by letting y become a time-varying parameter, $y \to y_t$, and using the (suboptimal) estimator $\hat{y}_t = L^{-1} \sum_{(i,j)\in B} a_t^{ij}$. Hence, when network density is clearly not constant, a two-step estimation procedure can be implemented. First, we estimate the time series of density parameters $\{\hat{y}_t\}_{t=p+1,\dots,T}$. Second, the MLE $\hat{q}$ and $\hat{c}$ are obtained by solving Eq. (6), but conditioning on the values $\{\hat{y}_t\}_{t=p+1,\dots,T}$. In the presence of some density pattern, we use this method to obtain a genuine estimation of the memory parameters.

B. Heterogenous and time-varying parameters

In the case of heterogenous parameters, the MLE problem in Eq. (6) can be generalized and solved, similarly to what has been done in [6]. We study explicitly such MLE problem in the Appendix A 16.

In the case of time-varying parameters, we can use a nonparametric technique based on *local* likelihood estimation to infer the dynamics of q_t, c_t, and y_t. The main idea relies on considering observation weights, which are decaying in time, in the maximum likelihood equations. Thus, to obtain a local (in time) estimate of parameters at time t, we fit the model by using those observations that are closer to the time snapshot t. This localization is achieved via a weighting function or kernel $K_\lambda(t, s)$ with bandwidth λ, which assigns a weight to s based on the time difference $|s - t|$. Here we use the Epanechnikov quadratic kernel

$$K_\lambda(t, s) = \begin{cases} \frac{3}{4}\left[1 - \left(\frac{|s-t|}{\lambda}\right)^2\right] & \text{if } \frac{|s-t|}{\lambda} < 1 \\ 0 & \text{otherwise} \end{cases}$$

insert math with $\lambda = 40$. For further details on the local likelihood method see Ref. [61], while for a study of the optimization of the bandwidth of the kernel see Ref. [67].

Hence, the local MLE problem at time t reads as

$$\arg\max_{y_t, c_t, q_t} \sum_{s=p+1}^{T} K_\lambda(t, s) \log \mathbb{P}(\underline{\underline{A}}_s | \underline{\underline{A}}_{s-1}, \dots, \underline{\underline{A}}_{s-p}, q_t, c_t, y_t), \tag{7}$$

with $q_t, c_t, y_t \in [0, 1]$. The maximum likelihood equations to solve follow as similar to the standard ones. Then, by rolling the kernel over time t, a nonparametric reconstruction of the dynamics of time-varying parameters is obtained.

C. Application to real temporal networks

In this empirical section, we consider the application of the CDARN(p) model to temporal networks, described as time series of adjacency matrices, each one capturing the links between the nodes of the network, within a given time resolution. Each link describes a particular interaction, typical of the networked system under investigation. The following data sets are considered:

(1) Transportation networks, i.e., bus (B), (underground) rail (R), and train (T), designed for the public transport in Berlin (B), Dublin (D), Helsinki (H), Paris (P), Rome (R), Sydney (S), Venice (V), Winnipeg (W), namely, records for the movements of public transport systems from stop to stop [68], with time resolution of 1 minute. In particular, a connection between two stops is associated with a time interval, from the departure to the arrival, thus, within our framework, a link appears at the network snapshot corresponding to departure, then lasts up to the snapshot which includes the arrival.

(2) Online social communication networks, in particular.

(a) Text message interactions between 101 college students (MSG), namely, messages sent between (anonymized) student users of an online communication platform at the University of California, Irvine, over a period of seven months, with a time resolution of 1 hour.

(b) Email communications (EM), namely internal email communications between 65 employees of a mid-sized manufacturing company over a period of 9 months, with time resolutions of 5, 10, 30 minutes, 1 hour, and 24 hours.

In both cases, a link is an instantaneous communication between two nodes, and it is described by the entry of the adjacency matrix associated with the network snapshot of all communications within the considered time window.

(3) Social interaction or contact network (CN), which can be seen also as a off-line social communication network, namely, the interactions (measured by Bluetooth devices, phones, carried by) of 94 students at MIT over eight months, with time resolutions of 10, 30, 60, and 120 minutes. Here a link is a contact between two nodes, lasting for the time of the interaction, measured as the number of network snapshots at which the link is present.

(4) Football networks, i.e., the temporal networks formed by footballers (F) over a match, for two different games (1, 2), and for both sides separately (home h, away a), with time resolutions of 1, 2, 10, 30, and 60 seconds. A link is a contact between two players, measured as a physical distance between each other below a threshold of 10 meters. Thus, a

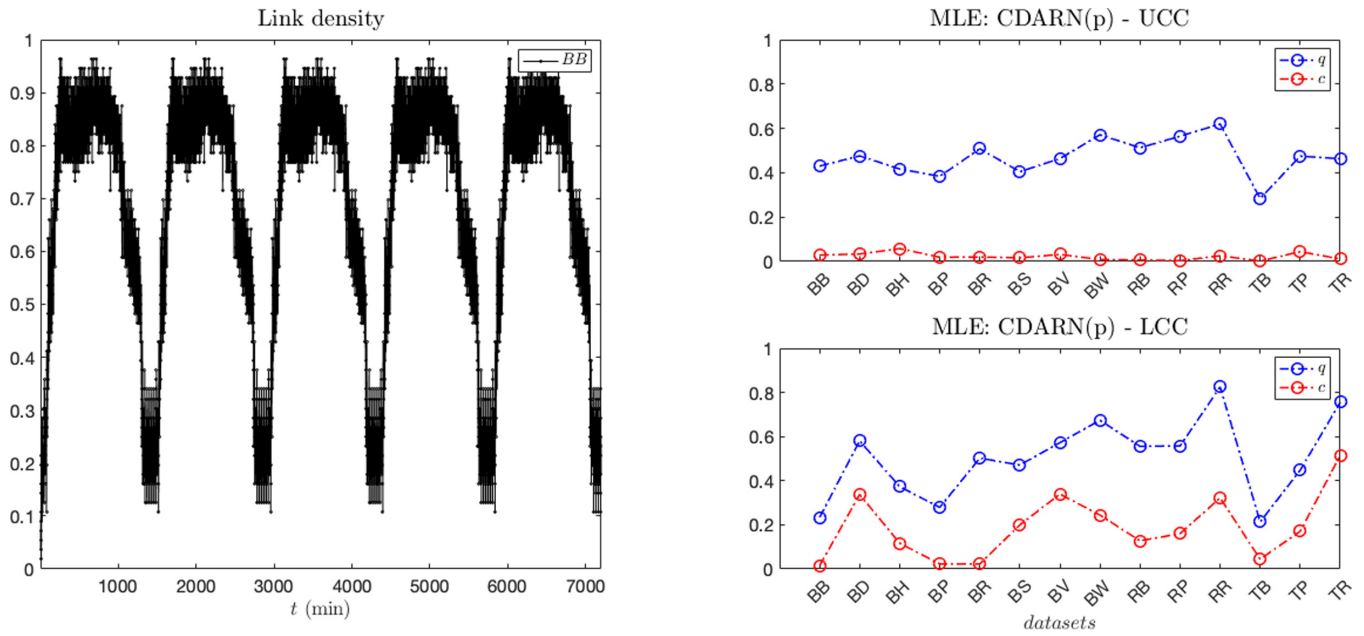

FIG. 1. Link density as a function of time (left) and MLE of the parameters q and c of the CDARN(p) model (right), considering both UCC (top right) and LCC (bottom right) coupling models, for the transportation network data sets.

link is described by an entry of the adjacency matrix associated with the network snapshot at which the contact occurs. Then a link lasts for all the time snapshots at which the contact is present.

Finally, the backbone is built by considering the network of all pairs connected at least once in the whole time period, for each temporal network.

We then estimate the CDARN(p) model on these network data sets, by considering the correction for the seasonality or nonstationarity patterns displayed by link density for the first three types of temporal networks (see the left panels of Figs. 1–3), while no correction is applied to the football networks, as supported by empirical evidence (see the left panel of Fig. 4), but we have nevertheless solved the original problem stated in Eq. (6). The maximum likelihood estimators of the parameters q and c are shown in the right panels of Figs. 1–4, for both the uniform cross-correlation (UCC) and local cross-correlation (LCC) coupling models. In both cases, the order p is selected by maximizing the likelihood of observing the given data under the CDARN(p) model.

(1) For transportation networks, the mechanism of copying from the past captures the observed link persistence patterns (long-lasting connections between two stops) as well as cross-interactions (transport connections at the stop), as testified by the large values of q. However, cross-interactions become significant only restricting to neighbor links over the backbone; see the estimated values of c (red points) in the right panels of Fig. 1, which are close to zero for UCC, significantly larger than zero for LCC (for almost all data sets). This behavior is consistent with the underlying dynamics of the considered transportation systems, where a link is a connection between two physical stops, and interactions may arise only between incoming or outgoing transport connections at the same stops. Furthermore, we can notice a significant positive correlation between the estimated parameters q and c in this case. Finally, the order p of the CDARN(p) model is estimated as one (generating a Markovian network), for all types of transportation and for all cities.

(2) Online social networks display less important memory patterns, as testified by small values of q and c; see the right panels of Fig. 2. Here, differently from above, we do not notice much difference between UCC and LCC coupling models. However, non-Markovian effects characterize such networks. In fact, for the UCC model we obtain $p = 4$ for the MSG network, $p = 5$ for the EM data set with 5 minute resolution, $p = 4$ for the EM with 10 minute resolution, $p = 2$ for the EM with 30 minute resolution, $p = 1$ for the EM with both 1 hour and 1 day resolutions. For the LCC we obtain $p = 5$ (MSG), $p = 2$ (EM5m), $p = 2$ (EM10m), $p = 3$ (EM30m), and $p = 1$ (EM1h and EM1d), respectively.

(3) Contact networks display a very important link persistence pattern, as opposed to very small or zero cross interactions between links, at any time resolution, as verified by values of q close to one, but c close to zero (for both UCC and LCC); see the right panels of Fig. 3. In fact, this social network is an example of the stability pattern characterizing some social ties, such as friendship. Furthermore, such social systems are described by Markovian dynamics ($p = 1$).

(4) Football networks display both link-specific persistence and cross interactions between links, with the two patterns which are inversely correlated as functions of the time resolution. For high resolution (1–2 seconds), we measure large values of q, as opposed to small values of c. This is the result of contacts between players lasting for longer than the typical resolution and resulting in links persistent over several network snapshots, thus described by the mechanism of copying (itself) from the past. However, when time resolution becomes lower than the typical duration of a contact, link-specific persistence patterns disappear, in favor of some

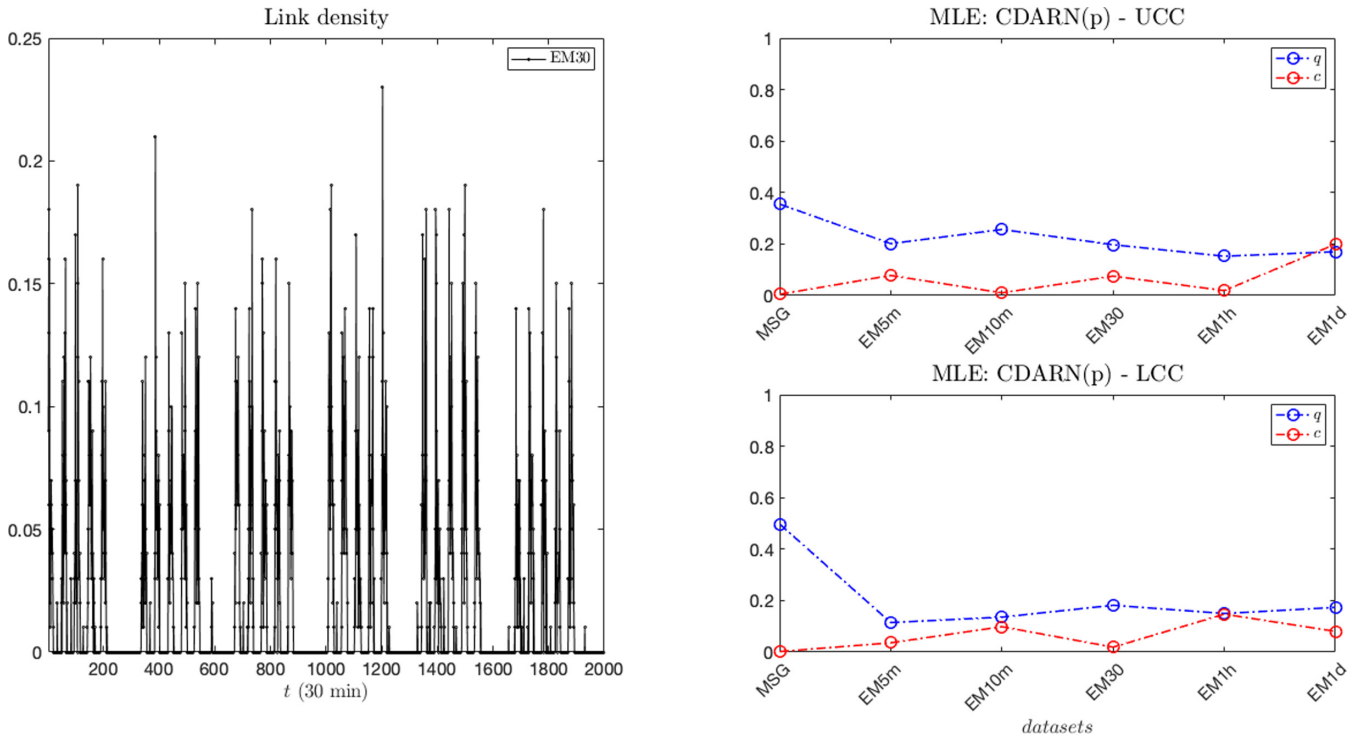

FIG. 2. Link density as a function of time (left) and MLE of the parameters q and c of the CDARN(p) model (right), considering both UCC (top right) and LCC (bottom right) coupling models, for the online social communication network data sets.

cross-interactions between links, probably related to game strategies in football which appear evident at that specific timescale. In particular, this behavior results in lagged cross-correlations for the link dynamics which are thus captured by large values of the parameter c. This is an example of how including cross-interactions is crucial to capture the dynamics of the system. This is further confirmed with a simple exercise of link prediction; see below. Finally, when time

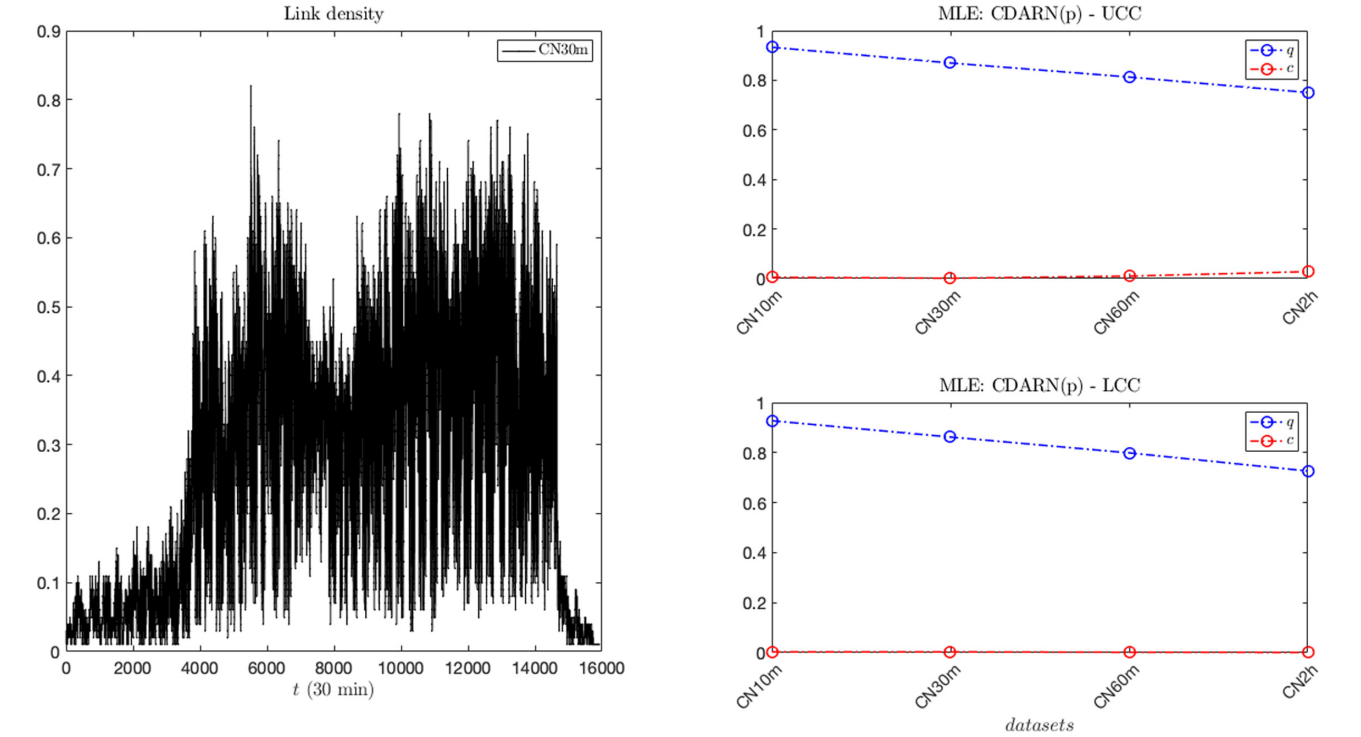

FIG. 3. Link density as a function of time (left) and MLE of the parameters q and c of the CDARN(p) model (right), considering both UCC (top right) and LCC (bottom right) coupling models, for the contact network data sets.

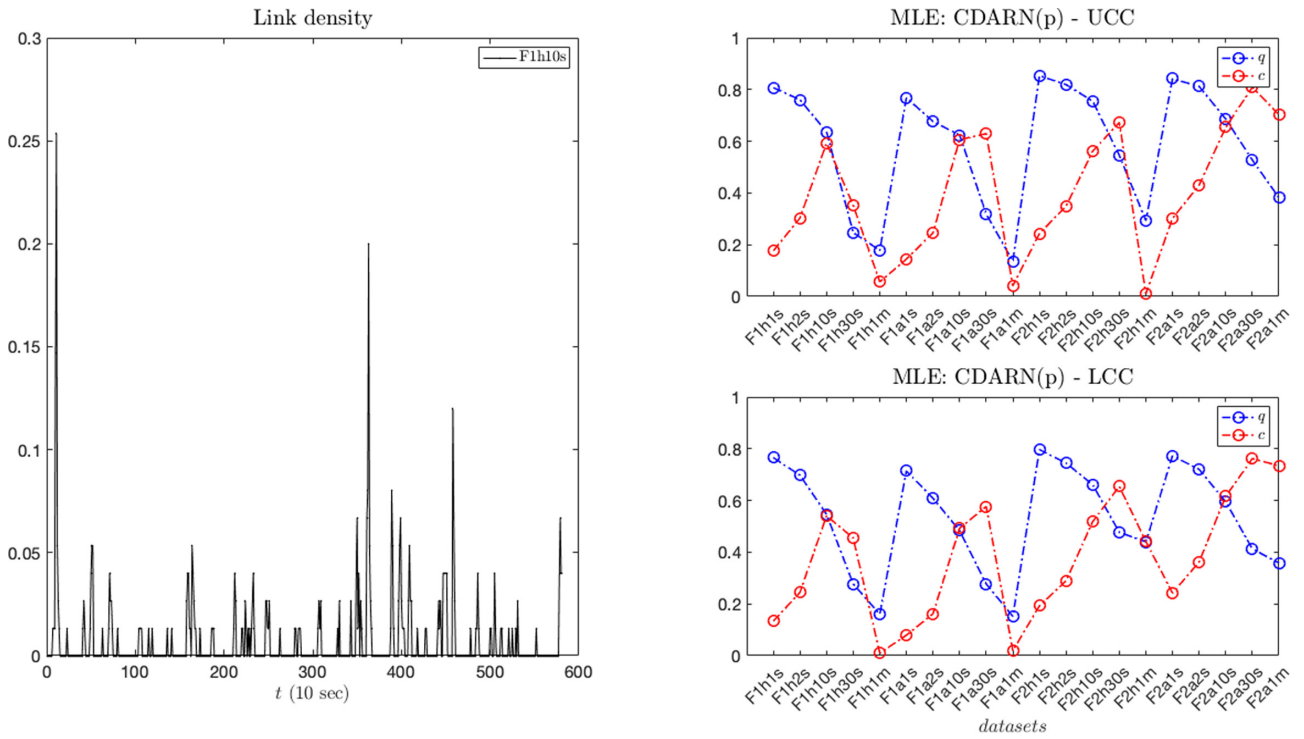

FIG. 4. Link density as a function of time (left) and MLE of the parameters q and c of the CDARN(p) model (right), considering both UCC (top right) and LCC (bottom right) coupling models, for the football network data sets.

resolution is too low (1 minute), the temporal information is destroyed and the estimated parameters q and c are small or close to zero for almost all data sets. For the football networks, we do not notice much difference between the UCC and the LCC coupling models, because of an almost full backbone graph. Finally, the order of the CDARN(p) model is selected equal to $p = 1$, for all football matches at any time resolution.

D. Heterogenous and time-varying patterns in real networks

Real-world networked systems may display heterogenous patterns in link dynamics. First of all, the probability for the appearance of a link is, in general, link-specific. For example, transports connecting different parts of a city are more or less frequent depending on people traffic, thus there are many buses crossing the main streets, while few buses connect the periphery. Second, auto- and cross-correlations of links may in general differ link by link. For example, in football contacts between midfielders tend to be persistent in time since the game is played largely in the midfield, while any contact between the forward and the defensive players is likely to be quick and short, for both scoring and defense.

In our framework, we can study such behaviors in link dynamics by considering the version of CDARN model with *heterogenous* parameters. In order to point out the relevance of the heterogeneous generalization, in the following we consider the local cross-correlation coupling CDARN(1) model with Markovian memory (as suggested by previous results) with heterogeneous parameters y^ℓ and q^ℓ applied to four network data sets: BD, MSG, CN30min, and F1h10s. The estimation method in such cases is described in the Appendix A 16. The results are shown in Fig. 5. It is interesting to notice that some networked systems display a similar marginal link probability among links, such as football or contact networks, while others, such as transportation and online social communication networks, are characterized by some degree of heterogeneity. A similar result is obtained by looking at the autocorrelation structure of networks, with similar autocorrelations of links for transportation and contact networks, while link-specific autocorrelation properties are observed in the other cases. This is an analysis that suggests time by time when the approximation with global parameters is enough or not for the precise description of a given network data set.

Heterogeneity in networked systems can be *spatial* as well as *temporal*. In general, the correlation structure of a network as well as link probability may change over time, thus displaying time-varying patterns in link dynamics. This behavior can be captured by using the CDARN model with time-varying parameters. In particular, we consider the LCC coupling model with Markovian memory and exploit local likelihood methods to estimate the dynamics of parameters. The results for three network datasets (BD, CN30min, and F1h10s) are shown in Fig. 6. Such a method is able to capture the (*smooth*) dynamics of the marginal link probability (see left panels of Fig. 6), similarly to the results of the previous section. Moreover, now we are able to describe the time-varying patterns of both auto- and cross-correlations of link dynamics (see right panels of Fig. 6). It is interesting to notice that some systems, such as transportation and contact networks, display a quite

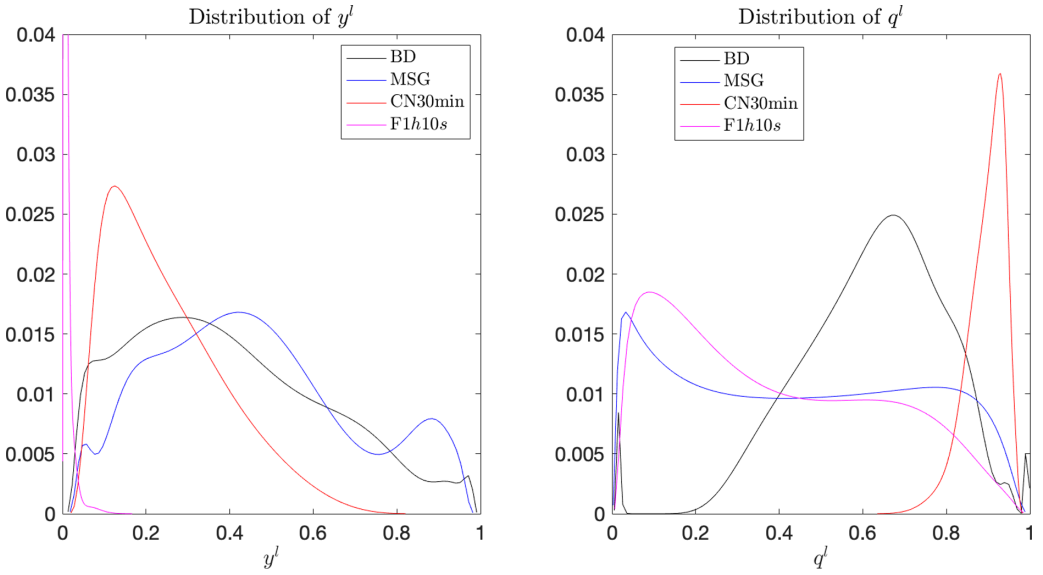

FIG. 5. Distribution of y^ℓ (left) and q^ℓ (right) of the *heterogeneous* CDARN(1) (LCC) model estimated on temporal network data built for four data sets as indicated in the legend. Parameters are estimated by solving Eqs. (A67) and (A68), respectively.

constant (around some mean value) correlation structure (except for the periods of no link activity when the estimation results as noisy, e.g., during night hours for transports). On the other hand, systems like football networks display significant time-varying patterns of link correlations, likely related to the different phases of the game.

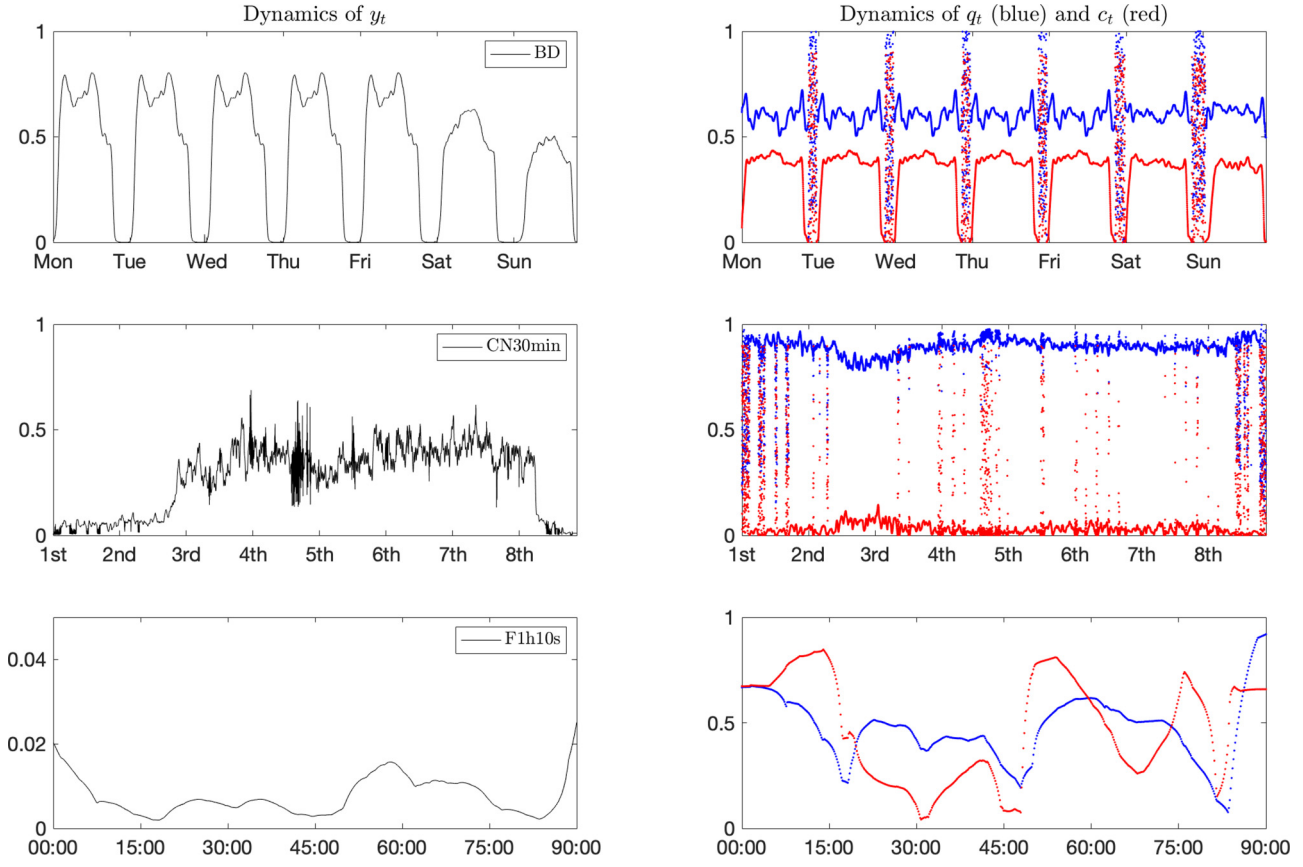

FIG. 6. Estimated dynamics of time-varying parameters of the CDARN(1) model, namely, y_t in the left panels and q_t and c_t in the right panels, obtained by using local likelihood methods for three network data sets, BD, CN30min, and F1h10s.

E. The role of cross-interactions and time-varying patterns in link prediction

Cross-interactions of links shape the dynamics of real-world networks in many ways. A preliminary indication of this comes from the values of the key parameter c we have obtained above. Moreover, real-world systems may display nonstationary patterns in link density as well as link correlations, which can be captured by time-varying parameters, as shown above. Such effects are significant not only for description, but also for forecasting. This can be made more evident by devising a simple study of link prediction in empirical networks based on the CDARN(p) model with constant and homogenous parameters q, c, and y, opposed to the case of heterogenous or time-varying parameters, as follows.

Assume that we have observed a temporal network up to time t (and also that the backbone does not change in time) and to try to predict the appearance of a link (i, j) at time $t+1$ based on the information up to time t. The one-step-ahead *forecast* (or *prediction*) is defined as the probability projected at time $t+1$ of observing the link (i, j),

$$S^{ij}_{t+1} \equiv \mathbb{P}(a^{ij}_{t+1} = 1 | \{\underline{\underline{A}}_s\}_{s=t,t-1,\dots,t-p+1}, q, c, y), \qquad (8)$$

for the CDARN(p) model with homogeneous and constant parameters. In the case of heterogenous parameters, it is

$$S^{ij}_{t+1} \equiv \mathbb{P}\left(a^{ij}_{t+1} = 1 | \{\underline{\underline{A}}_s\}_{s=t,t-1,\dots,t-p+1}, q^{(ij)}, c, y^{(ij)}\right), \qquad (9)$$

with link-specific parameters, as described above. In the case of time-varying parameters, it is

$$S^{ij}_{t+1} \equiv \mathbb{P}\left(a^{ij}_{t+1} = 1 | \{\underline{\underline{A}}_s\}_{s=t,t-1,\dots,t-p+1}, q^t, c, y^t\right), \qquad (10)$$

making sure to use a causal kernel (i.e., weighting only observations up to time t) in the estimation procedure. For the related explicit formulas, see the Appendix A 17. The time series of forecasts $\{S^{ij}_t\}$, together with the realizations $\{a^{ij}_t\}$, allow us to characterize the forecasting performance of the model by using some binary classifier. Here we consider the receiving operating characteristic (ROC) curve [61], which is the plot of the true positive rate (TPR) (sensitivity) against the false positive rate (FPR) (specificity) at various threshold values. In practical terms, the better the model performs in the forecasting, the higher the associated ROC curve is in the unit square, or, equivalently, the larger the area under the curve (AUC); see the Appendix A 17 for further details.

As case study we have considered the football matches network data set. We will show the results of the prediction analysis for the network of game 2 away, although similar results have been obtained for other matches. We aim to validate the model performance, in particular to verify the effect of including cross interactions to better capture the network dynamics of real-world systems, together with the role played by heterogenous and time-varying patterns in link prediction. Thus, we compare the no cross-correlation (NCC) coupling model, i.e., the DARN model, with the local cross-correlation (LCC) specification of the CDARN model, with either homogeneous, heterogeneous, or time-varying parameters.

The link prediction study is as follows: (1) we split the sample period in two, the first half of the match is used as a *training set* and the second half as *out-of-sample period*, then (2) we estimate the parameters of each coupling model on network data of the first half, by solving the MLE problem (6) for q, c, y with data $\{\underline{\underline{A}}_t\}_{t=1,\dots,T^{\text{half}}}$ (or the corresponding problems for heterogeneous or time-varying parameters), and finally (3) we construct the time series of forecasts, snapshot by snapshot, by considering a rolling window over the second half, i.e., from $T^{\text{half}}+1$ to $2T^{\text{half}}$, thus obtaining $\{S^\ell_t\}^{\ell\in B}_{t=T^{\text{half}}+1,\dots,2T^{\text{half}}}$. Notice that in this exercise the model parameters q, c, y, or q^ℓ, c, y^ℓ in the heterogenous case are estimated by using only data from the first period, and not updated each time the window rolls over new snapshots of the second half. On the contrary, in the case of time-varying parameters, every time the window rolls over a new observation the estimate of q^t, c^t, y^t is updated. The link prediction exercise is restricted to all pairs which can be connected on the backbone. In conclusion, we compare the time series of forecasts $\{S^\ell_t\}^{\ell\in B}_{t=T^{\text{half}}+1,\dots,2T^{\text{half}}}$ with the realizations $\{X^\ell_t\}^{\ell\in B}_{t=T^{\text{half}}+1,\dots,2T^{\text{half}}}$, by evaluating the ROC curve. The results are summarized in Fig. 7, for match 2-away with time resolution equal to 10 seconds (however, similar results are obtained for different matches and time resolution). We can notice that the LCC coupling specification of the CDARN model (blue line in the left panel of Fig. 7), which accounts for both autocorrelations and cross-interactions, always outperforms the DARN model (black line), accounting only for the autocorrelation of links. Moreover, accounting for the heterogeneity pattern of link probability y^ℓ (red line) plays an important role in link prediction of football data, largely outperforming the case with heterogeneous correlations q^ℓ (magenta line). The underperformance of the CDARN model with heterogeneous q^ℓ parameters w.r.t. the DARN model is a signal of overfitting for the specific case of football networks. Finally, when comparing the forecasting performances of CDARN with constant or time-varying parameters (see the right panel of Fig. 7), the benefit of accounting for time-varying patterns in link prediction depends on the degree of specificity, i.e., false positives, we are willing to accept to obtain some given degree of sensitivity, i.e., true positives. In any case, a more timely estimation of parameters, associated with a tighter kernel bandwidth λ, tends to produce better forecasting. In conclusion, the football network is an example of the importance of taking into consideration lagged cross-correlations of links, together with heterogenous or time-varying patterns, in the description of the dynamics of networked systems.

V. DIFFUSION PROCESSES ON CORRELATED TEMPORAL NETWORKS

One of the most important points when modeling a networked system is understanding how information, or some other quantity, spreads throughout the system, in particular the rate of the diffusion and the time in reaching the equilibrium. When links between nodes change over time, then the first interest is on the role either memory and link dynamics play in the diffusion process.

In order to study this in a systematic way, in this section we will exploit the flexibility of the CDARN(p) model introduced in Sec. III, which, as shown in Sec. IV, allows us to generate

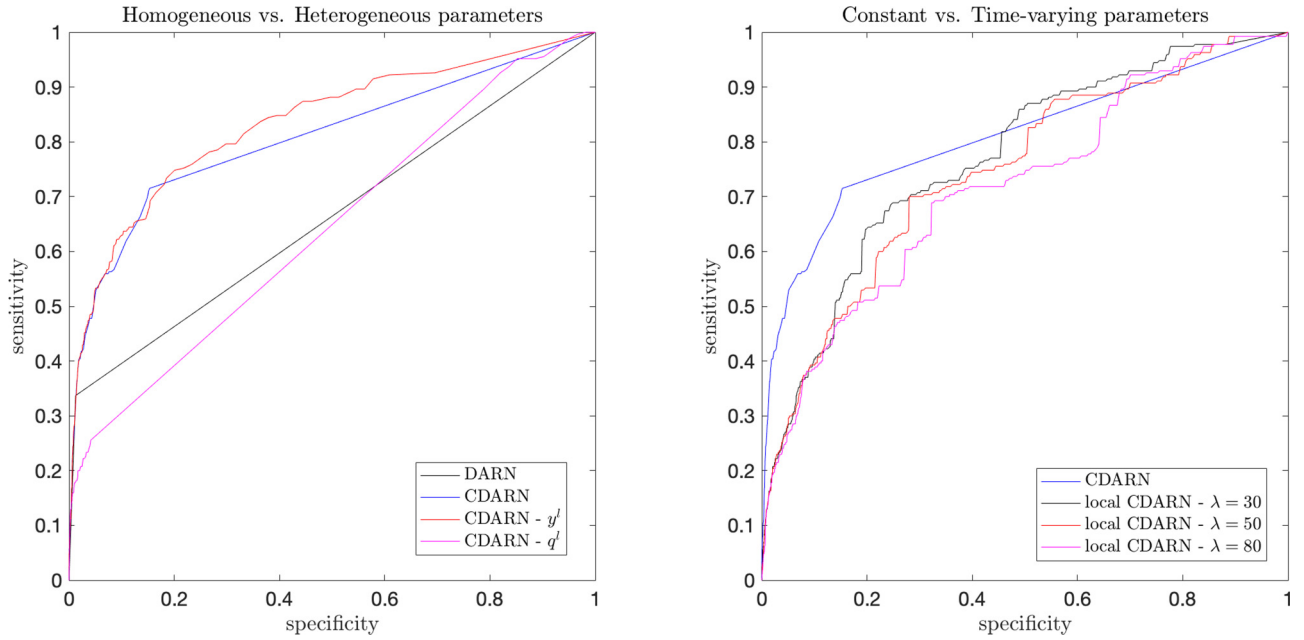

FIG. 7. Receiving operating characteristic curves built for the DARN(1) model (LCC) and the CDARN(1) with NCC specification, applied to the football network associated with game 2, away with time resolution of 10 sec. We compare the standard version of the CDARN model also with both the heterogeneous and the local generalizations. In the case of time-varying parameters, we use the causal Epanechnikov quadratic kernel (i.e., weighting only past observations) for different bandwidths λ.

realistic temporal networks. The model allows us to fine tune the strength and length of the memory, while also controlling a key feature of real-world networks, namely, correlations between the evolutions of links over time, as produced by dependencies between their dynamics.

A. Quantifying diffusion on a temporal network

Diffusion is, in its original sense, the physical process by which atoms and molecules move from regions of high concentration to regions of low concentration. This process has been seen as an analog to processes in several other areas, such as opinion formation [69], the motions and social interactions of people [70], and the movements of capital through a financial system [71], and as such is among the most common ways of describing spreading phenomena in these areas. Indeed, diffusion finds uses in many other areas, where it is used as a linear approximation to nonlinear systems, such at the Kuramoto model [72].

Complex networks often form the backbone of many real-world systems, and so it is natural to study diffusion over them [73,74]. In a diffusive process on a network the flow of information, or some material, over a link is proportional to the difference in its concentrations at the two nodes. The natural way to study diffusion on a network is in terms of the so called Laplacian matrix, which forms the network analog of the Laplace operator, which governs continuous-time, continuous-space, diffusion. Suppose we have a static undirected network with N nodes and adjacency matrix $\underline{\underline{A}} = \{a^{ij}\}$. The equation that governs the diffusion of some node-related quantity $\underline{d}(s) \in \mathbb{R}^N$ over (continuous) time s can be written as

$$\dot{\underline{d}}(s) = -\mu \underline{\underline{\mathcal{L}}}\,\underline{d}(s), \tag{11}$$

where μ is the diffusion coefficient, which controls the timescale of the diffusion process, and $\underline{\underline{\mathcal{L}}} = \{\mathcal{L}^{ij}\}$ is the graph Laplacian matrix, whose entries can be written in terms of the entries of $\underline{\underline{A}}$ as $\mathcal{L}^{ij} = \delta(i,j)k_i - a^{ij}$, where $k_i = \sum_j a^{ij}$ is the degree of node i [74]. Notice that this equation is in continuous time; as a convention when a variable is continuously dependent on time s, the time will be in brackets [e.g., $\underline{\underline{A}}(s)$], and for discrete time t it will be given as an index (e.g., $\underline{\underline{A}}_t$). On a temporal network the only thing that needs to be changed in this equation is that the Laplacian matrix must be allowed to vary over time, hence $\underline{\underline{\mathcal{L}}} \mapsto \underline{\underline{\mathcal{L}}}(s)$ where $\underline{\underline{\mathcal{L}}}(s)$ is the Laplacian matrix associated with the continuous time adjacency matrix $\underline{\underline{A}}(s)$. This system exists in continuous time, and so the temporal network that underlies it must also exist in continuous time. The solution of the above equation is then clearly

$$\underline{d}(T) = \exp\left(-\mu \int_0^T \underline{\underline{\mathcal{L}}}(s)ds\right)\underline{d}(0).$$

However, the vast majority of models for temporal networks are discrete in time, and so, given a model for a discrete time temporal network, we must first embed the network in continuous time. To this end, we assume that the adjacency matrix changes at discrete time steps of length Δt, taken, without loss of generality, to be equal to 1. Thus the Laplacian $\underline{\underline{\mathcal{L}}}(s)$ is piecewise constant and, according to the above notation, will be denoted by $\underline{\underline{\mathcal{L}}}_t$ $(t = 1, .., T)$. The solution of the diffusion equation hence becomes

$$\underline{d}_T = \exp\left(-\mu \sum_{t=1}^{T} \underline{\underline{\mathcal{L}}}_t\right)\underline{d}_0. \tag{12}$$

As stated, our purpose here is to study the effects that memory in a temporal network has on diffusion over that network. This is a very general aim, and so we must be more specific about what we wish to analyze. Rather than studying spreading in terms of the full dynamics of diffusion on a temporal network, i.e., the concentrations $\underline{d}_t$ of material at each node at each time step t, we can instead ask about how long it takes for this diffusion to reach equilibrium. In particular, since the changes in the network are responsible for any changes in the rate of spreading, we focus on the number of network evolutions (number of time steps Δt) before equilibrium. To formalize this concept we first note that in general we will not reach equilibrium in a finite number of time steps, and so we instead fix some small positive ϵ, so that the *time to equilibrium* is then defined as

$$\tau = \min_{t\in\mathbb{N}}(t : |\underline{d}(t) - \underline{u}| < \epsilon), \tag{13}$$

where the vector $\underline{u}$ is the uniform vector with $u^i = 1/N$, which corresponds to the equilibrium state of the diffusion process on a connected network with N nodes. For our purposes the norm $|\cdot|$ will be taken to be the Euclidian norm. For our purposes we will keep the value of ϵ fixed as $\epsilon = 10^{-3}$. The temporal networks we will use here are generated by discrete-time random processes, and so τ will be a random variable. Given this, we will focus on finding the average of this value, $\langle\tau\rangle$, over several realizations of the system. Unfortunately, $\langle\tau\rangle$ will be highly dependent on the structure or size of any temporal network being studied, and so it would be impossible to draw conclusions about the influence of any model parameters in these systems. Our goal here is to study the effects of memory on spreading rate, and so we must introduce some way of comparing the time to equilibrium as a function of this memory as given by different networks. To this end we normalize τ by expressing it in terms of the time taken for a diffusion to reach equilibrium on the same backbone, but with a memoryless temporal network. In other words, we define the *rescaled time to equilibrium*, $\mathfrak{T}^p$, given memory length p, in terms of $\langle\tau^p\rangle$, the average time to equilibrium given memory length p, as

$$\mathfrak{T}^p = \frac{\langle\tau^p\rangle}{\langle\tau^0\rangle}. \tag{14}$$

Notice that the CDARN(p) model does not directly allow for $p = 0$, and so we define τ^0 to be the case where $q = 0$, and so no memory is ever used. This allows us to compare the effects that changing the memory length p and the coupling matrix $\underline{\underline{C}}$ have on different backbones.

B. Numerical results

We have first investigated the rescaled time to equilibrium of a diffusion process on CDARN(p) temporal network models with different backbones by means of an extensive set of numerical simulations. The value of the parameter μ allows to tune the timescale of the diffusion process, while the three parameters controlling the link density y, memory strength q, and memory length p and the two matrices network backbone $\underline{\underline{B}}$ and link coupling matrix $\underline{\underline{C}}$ control the properties of the temporal network. To construct the backbones $\underline{\underline{B}}$ we have taken three real-world temporal networks, each with different structural properties, and we have aggregated their links over the extent of the available network data and discarded the link weights. The three real temporal networks we have considered are the following: (1) Flights between U.S. airports (Airport) [75]; (2) email interactions between employees at a manufacturing company (Email) [76]; and (3) journeys on the London underground (Tube) [77]. The key features of the three resulting backbones are summarized in Table I. The number of nodes in the three networks ranges from about 100 to 300. With 302 nodes and an average degree $\langle k\rangle = 2.3$ the Tube is the backbone with the smallest link density, while Email is a very dense backbone with links connecting 23% of the possible pairs of nodes. Our aim here is to study not only the effects of memory, but also the interplay between memory and correlations in the dynamics of links. In order for us to clearly observe the effects of these features we must be able to compare different models: one in which the evolution of links is correlated, and one in which links are independent. As such we have simulated our system on each of the three different backbones $\underline{\underline{B}}$ for a range of different parameters p, q, y, and μ, and, for the three different forms of the coupling matrix $\underline{\underline{C}}$; see Sec. III A.

TABLE I. Key structural features for each backbone. The number of nodes (N), average degree ($\langle k\rangle$), density (D), and dominant (λ_N), and smallest nonzero (spectral gap, λ_2) eigenvalues of the Laplacian matrix for each backbone.

Backbone	N	$\langle k\rangle$	D	λ_N	λ_2
Airport	143	2.030	0.0143	31.02	0.01696
Email	167	38.93	0.2345	140.0	0.3811
Tube	302	2.311	0.0078	8.432	0.005918

In our simulations, for each instance of diffusion on a CDARN(p) model, i.e., for each different set of parameters μ, and $p, q, y, \underline{\underline{B}}, \underline{\underline{C}}$, and c, we compute $\mathfrak{T}^p$. This is done directly by estimating $\langle\tau^p\rangle$ and $\langle\tau^0\rangle$, where the averages are taken from multiple realizations of the diffusion process. In each case the initial condition for the diffusion $\underline{d}_0$ is such that all of the material to be diffused is placed at a random node j: $d_0^i = \delta(i, j)$ where $j \in \{1, \ldots, N\}$. In this way we avoid any bias that might be introduced by repeatedly choosing the same starting node. Before any diffusion takes place on the temporal network, we allow the CDARN(p) temporal model to evolve until it has reached a steady state (see Appendix A 7).

In Fig. 8 we report the rescaled time to equilibrium $\mathfrak{T}^p$ given memory length p as a function of p, for each backbone and with a number of different sets of model parameters. Note that a semilog scale has been used. To ensure that memory plays a significant part in the evolution of the temporal network we have fixed the memory strength $q = 0.95$, and to ensure that the system has enough time for the effects of memory to be observable we have fixed the link density $y = 0.1$. We then vary the diffusion speed $\mu = 0.1, 0.5$. All of these results are shown for both the local cross-correlation model, with the three values of the coupling strength $c = 0.5, 0.3, 0.1$, and the no cross-correlation model, i.e. the case $c = 0$. For $\mu = 0.1$, and hence slow diffusion, we observe

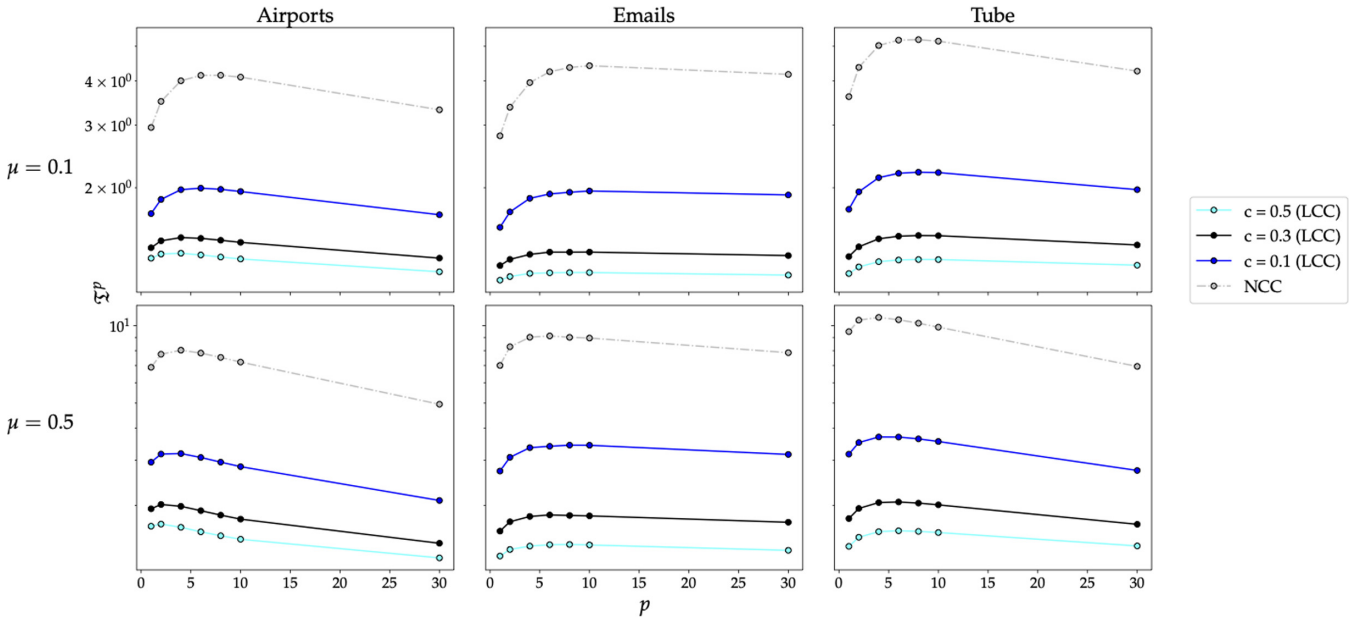

FIG. 8. Rescaled time to equilibrium for diffusion on different network backbones as a function of the memory length p for a CDARN(p) model with local (solid lines, LCC), and no (gray line, NCC) cross-correlations between different links. Memory strength q is kept constant at 0.95 to ensure that memory plays a significant role in the evolution of the network and link density y is kept at 0.1 to ensure that there is sufficient time for any effects of memory to be observed. The coupling strength c and diffusion speed μ are varied. The backbones were taken from a collection of real data sets. Averages were taken over 2×10^4 realizations of the process. Note that a semilog scale has been used.

that the equilibrium time is nonmonotonically dependent on the memory length for all backbones, coupling strengths, and for both coupling models. This nonmonotonicity is most prominent when $c = 0.1$, but far less so when the coupling is stronger. When we consider $\mu = 0.5$, and hence faster diffusion, the observed nonmonotonicity is far less apparent in all but the NCC model, there is, however, still a clear dependence on memory, particularly for lower values of c. Unsurprisingly, there is a significant difference between the results for the no correlation model and those with correlations: in all cases local correlations speed up diffusion. What we do notice though is that there is no marked difference between different backbones. Since we have normalized each set of results this is not entirely unexpected.

In summary, the rescaled equilibrium time shows a number of interesting features as a function of the memory length p, the coupling matrix $\underline{\underline{C}}$, and the backbone $\underline{\underline{B}}$. Most notable among these features are the following:

(1) The rescaled time to equilibrium $\mathfrak{T}^p$ is generally a nonmonotonic function of the memory length p.

(2) Stronger local correlations, i.e., larger values of the coupling strength c speed up diffusion.

(3) Correlations have a considerable effect on the influence of memory: when the coupling strength c is high then diffusion properties are weakly dependent on the memory properties of the network.

As we will show in the following, by understanding the behavior in the limit of no cross-correlations, and by isolating the effects of temporal correlations, we can get a clear picture of the causes of our observations.

C. Analytical results in the no cross-correlations limit

In light of our numerical results, we now study the theory which underpins both the CDARN(p) model and the diffusion of material over it. We first study diffusion on the simplest form of the CDARN(p) model, the limit of no cross-correlation between the dynamics of links. This is precisely the NCC coupling model that was previously introduced. In such a limit the links of the CDARN(p) model are independent processes, and so we can study them in isolation. In this case, as we will show below, the model is analytically tractable and it is possible to derive an analytical expression for the rescaled time to equilibrium.

In order to analyze the dynamics of diffusion over a single link of the CDARN(p) model, let us consider two nodes, one of which has an amount of a material, and the other of which has some other amount. The diffusion of this material out of the first node is given by

$$\dot{d}^1(s) = -\mu[d^1(s) - d^2(s)]a_t^{1\,2}, \tag{15}$$

where $t = \lfloor s \rfloor$, and the random variable $a_t^{1\,2}$ describes the presence of the link between node 1 and node 2 at discrete time $t = 0, 1, \ldots$ as governed by the DAR(p) process defined in Eq. (3). When combined with the conservation condition $d^2(s) = 1 - d^1(s)$ this describes the full dynamics of the diffusion process. Given any set of initial conditions we can first find the number τ of time steps before equilibrium is reached. By noticing that, since when the link is not present there can be no diffusion, we only need to count the number of times that the link is present. If we were to take $a_t^{1\,2} = 1$ for all t, then we can easily find $\tau = n$, and express n in terms of

μ, Δt and ϵ (see Appendix A 4). Now let us associate with a_t^{12} the counting process $F_t = \sum_{k=0}^{t} a_k^{12}$. We can then see that for a link that changes in time $\tau = \min_{t>0}(t : F_t = n)$. This allows us to rephrase our problem: we now want to find the average time taken until a link governed by a DAR(p) process has occurred n times. The DAR(p) process that governs the link can be thought of as a pth-order Markov process with the following transition matrix (see the Appendix A 5 for a full explanation and discussion):

$$T_{\alpha\beta} = \left[q\frac{h(\alpha)}{p} + (1-q)y\right]\delta\left(\beta, 2^{p-1} + \left\lfloor\frac{\alpha}{2}\right\rfloor\right) + \left[1 - q\frac{h(\alpha)}{p} - (1-q)y\right]\delta\left(\beta, \left\lfloor\frac{\alpha}{2}\right\rfloor\right). \quad (16)$$

Here α and β represent some indexing of the $S = 2^p$ possible memory states, $h(x)$ is the Hamming weight of the number x (the number of 1's in its binary representation), $\delta(x, y) = 1$ if $x = y$ and 0 otherwise, and $\lfloor x \rfloor$ is the largest integer value smaller than x. If we break this matrix up into two parts, $\underline{\underline{T}}^L = [1 - q\frac{h(\alpha)}{p} - (1-q)y]\delta(\beta, \lfloor\frac{\alpha}{2}\rfloor)$ and $\underline{\underline{T}}^R = \underline{\underline{T}} - \underline{\underline{T}}^L$, then we can find the average time $k_\alpha \in \mathbb{R}^S_{\geqslant 1}$ taken for a link to occur given that it started in state α as [78,79]

$$\underline{k} = (\underline{\underline{I_d}} - \underline{\underline{T}}^L)^{-1}\underline{1}, \quad (17)$$

and the probability $h_{\alpha\beta}$ that when a link occurs it will occur in state β, given that it started in state α as

$$\underline{\underline{h}} = (\underline{\underline{I_d}} - \underline{\underline{T}}^L)^{-1}\underline{\underline{T}}^R. \quad (18)$$

Now let us define ω_α as the probability that a link starts in state α. We can then find the average time taken until the nth link in a pth-order system as

$$\langle\tau^p\rangle = \underline{\omega}^T\left(\sum_{t=0}^{n-1}\underline{\underline{h}}^t\right)\underline{k}. \quad (19)$$

We now have an explicit formula for the average number of time steps to equilibrium. However, it is impossible to compare values of $\langle\tau^p\rangle$ directly, as such values will be heavily dependent on parameters of the model other than the memory length p. Because of this, we look at the rescaled time to equilibrium as defined in Eq. (14). The limiting behavior of this quantity can be studied analytically. First, we note that $\langle\tau^0\rangle$ can be found directly as n/y. Then, we observe that as $p \to \infty$, $\langle\tau^p\rangle \to n/y$ (see the Appendix A 3 and A 8), meaning that our large memory limit is exactly the same as the no memory case, and because of this $\mathfrak{T}^p$ is not intrinsically bounded above (see the Appendix A 8). We can also directly solve for $p = 1$, and in principle extend these calculations to solve for small p (see the Appendix A 8). Finally we can show that, when y is "small enough," as it is in all of our cases, $\langle\tau^p\rangle \geqslant \langle\tau^\infty\rangle$, and hence that

$$\mathfrak{T}^p \geqslant 1. \quad (20)$$

Hence, the rescaled equilibrium time in the large memory limit acts as a lower bound for the case of arbitrary p (see the Appendix 9), explaining the similar behavior observed in the full system. It should be noted that in cases where y is not "small enough" we will observe the opposite effect: the large memory limit will be an upper bound.

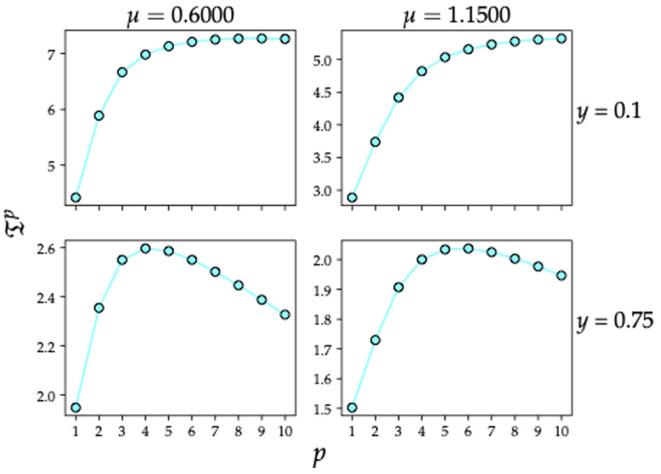

FIG. 9. Rescaled time to equilibrium for diffusion over a link in the limit of the CDARN(p) model with no cross-correlations as a function of the memory length p. The dynamics of the link is generated by a DAR(p) model with $q = 0.95$, for various values of y. Two different values of diffusion constant μ were used.

When plotting this rescaled time to equilibrium as a function of p for various μ and y, as in Fig. 9, we observe many of the same traits we found in Sec. V B for the full CDARN(p) model with cross-correlations. Principally, the following two similarities needs to be noted. First, we see evidence for the previously explained large memory limit, i.e. the rescaled time to equilibrium is bounded below by the value obtained in the limit of large p. Second, we see that $\mathfrak{T}^p$ can be highly nonmonotonic as a function of p.

In summary, the study of the CDARN(p) model in the limit of no cross-correlations provides us with a good understanding of the causes for two of the most notable phenomena observed in the full network systems, and allows us to focus on the role of correlations in inducing the remaining effects.

D. Derivation of the temporal correlation matrix

We will now present a general analytical approach to finding the lagged cross- and autocorrelations for an arbitrary coupling matrix $\underline{\underline{C}}$, which we will use in Sec. V E to explore the interplay between correlations and memory in more depth than would be possible through simulations alone. In particular, we will use it to isolate the effects that correlations among neighboring links in the LCC coupling model have on the time taken for diffusion processes on the temporal network to reach equilibrium.

The results of Sec. V B clearly indicate that the presence of coupling in the temporal dynamics of different links plays an important role in the behavior of the rescaled time to equilibrium for a diffusion process on a temporal network. Indeed, our claim is that, while the temporal autocorrelations of links slow diffusion [30,31,37,39,41], as evidenced by the limit of no cross-correlations case, temporal correlations among neighboring links speeds it up. Fortunately, the CDARN(p) model is analytically tractable enough for us to fully describe

the correlations that are present for a general coupling matrix $\underline{\underline{C}}$, without relying wholly on simulations.

Rather than working with the backbone network directly, we will instead consider the corresponding graph in which links are the nodes in the backbone, and we connect any two nodes in the new graph if their links in the backbone graph shared a node. For a backbone network with L possible links, we assign each of these links with a linear index. Then let us denote the correlations between link ℓ and ℓ' at time lag k as $\langle a_t^\ell a_{t-k}^{\ell'}\rangle = \rho_k^{\ell\ell'}$. Given a coupling matrix $\underline{\underline{C}}$, $\rho_k^{\ell\ell'}$ can be found as the solution to the following Yule-Walker equations [55,80]:

$$\underline{\underline{\rho}}_k = \frac{q}{p}\underline{\underline{C}}\sum_{a=1}^{p}\underline{\underline{\rho}}_{k-a}. \tag{21}$$

Note that we have dropped our indices, and so each element in the equation is a matrix. We can show that, for general $\underline{\underline{C}}$, this equation is solved by the composition of different functions over supports $k \in \{np, \ldots, (n+1)p\}$ for integer n. The first of these can be found to be constant, while the following are exponentially decaying (see Appendix A 10 and 11). Because of this, we can characterize the correlations at all values of k in terms of this initial constant, which we call $\underline{\underline{\rho}}$. We first define the following tensor:

$$\Delta^{\ell\ell'\ell''} = \frac{q}{p}\left((p-1)c^{\ell\ell''} + q\sum_{b\neq\ell'} c^{\ell b}c^{b\ell''}\right), \tag{22}$$

then $\rho^{\ell\ell'}$ can be found as the solution to the following system of linear equations (see the Appendix A 10):

$$\rho^{\ell\ell'} = \sum_{\ell''=1}^{L}\Delta^{\ell\ell'\ell''}\rho^{\ell''\ell'} + \frac{q}{p}c^{\ell\ell'}. \tag{23}$$

The system can be greatly simplified in special cases (see the Appendix A 11, 12, 13, and 14). For example, in the case of the UCC coupling model, we show that $\rho^{\ell\ell}$ is constant for all ℓ, and similarly $\rho^{\ell\ell'}$ is constant for all pairs ℓ, ℓ' such that $\ell \neq \ell'$, thus reducing the calculation of the correlation coefficients to solving a pair of linear simultaneous equations. Given this set of equations for $\rho^{\ell\ell'}$, we can also then find the correlations $\langle a_t^\ell a_t^{\ell'}\rangle = \rho_0^{\ell\ell'}$ when $\ell \neq \ell'$ as

$$\rho_0^{\ell\ell'} = q\sum_{\ell''=1}^{L} c^{\ell\ell''}\rho^{\ell''\ell'}. \tag{24}$$

This gives us a full picture of the correlations present in the CDARN(p) model and allows us to calculate them directly.

E. Cross-interactions speed up the diffusion

We saw in our study of diffusion in the limit of no cross-correlations that the rescaled time to equilibrium $\mathfrak{T}^p$ is a nonmonotonic function of the memory length p. It is also widely understood that a way of characterizing memory of a time series is by using the autocorrelation function. We find for a CDARN(p) process the memory p is precisely the value for the time lag k after which the correlation function $\underline{\underline{\rho}}_k$ decays exponentially (see the Appendix A 8 and results in [35]). With this in mind we can now focus on the comparison between autocorrelation coefficients of links in a CDARN(p) temporal model and the cross-correlation coefficients of neighboring links. This is done by studying the constant values for the auto- and cross-correlation coefficients at time lags $k \leqslant p$. In order to do this effectively for large networks we will average these quantities over all links (and neighbors where appropriate) to gain the averaged autocorrelation coefficient ρ_{ac} and the averaged neighborhood correlation coefficient ρ_{ncc}. These are defined, given the matrix of correlation coefficients $\rho^{\ell\ell'}$ for a backbone $\underline{\underline{B}}$ with L links derived in Sec. V D, as

$$\rho_{ac} = \frac{1}{L}\sum_{\ell=1}^{L}\rho^{\ell\ell}, \quad \rho_{ncc} = \frac{1}{L}\sum_{\ell=1}^{L}\frac{1}{|\partial_B\ell|}\sum_{\ell'\in\partial_B\ell}\rho^{\ell\ell'}, \tag{25}$$

where as before $\partial_B\ell$ is the set of links in the neighborhood of ℓ on the network backbone $\underline{\underline{B}}$.

For clarity, let us now restate our claim, as based on our observations of the numerical simulations displayed in Fig. 8: while autocorrelation of links slows down diffusion, correlations between neighboring links speeds diffusion. While in Fig. 8 we do see that diffusion is faster in the LCC model than in the NCC model, the autocorrelations of links in the two models are different. Further to this, while it would be possible to tune the parameters of the NCC model so that it produced links with the same autocorrelation coefficient as the LCC model, as the NCC model does not have a coupling strength, this could be achieved only by changing either the memory strength q or the memory length p. Because of this we cannot judge the influence of neighborhood correlations from our previous results, and we cannot use the NCC model to explore the effects of neighborhood correlations further. In order to give a valid point of comparison, we can now make use of our third coupling model, which allows us to precisely control the average link autocorrelation, but also removes any correlations between neighboring links. To recall, for a backbone with L links, the coupling matrix in the UCC model $\underline{\underline{C}} = \{c^{\ell\ell'}\}$ is given by $c^{\ell\ell'} = (1-c)\delta(\ell,\ell') + (1-\delta(\ell,\ell'))c/(L-1)$. The simplicity of this model lends itself well to analytical calculations, and so we can now use this model to isolate the effects of neighborhood correlations. Indeed, we can show that in the limit of large numbers of links L this model reduces to a DARN(p) temporal network on a fixed backbone, in which links are independent (see the Appendix A 14). First, we fix the parameters p, q, and y for both the LCC and UCC models, this ensures that there is the same memory strength and length, and the average degree of the temporal networks produced are the same. We can then fix the value of c for the LCC model, as shown in Fig. 10 (first row), and calculate the resulting value of ρ_{ac} and ρ_{ncc}, as defined by Eqs. (22), (23), and (25). By then varying the value of c used in the corresponding UCC model we obtain precisely the same value for ρ_{ac}, while leaving $\rho_{ncc} \approx 0$, because in the considered network backbones the number of links L is large. In Fig. 10 we plot both the values for ρ_{ac} and ρ_{ncc}, along with the rescaled time until equilibrium for a diffusion process on the corresponding temporal network. Note that the LCC and UCC models have, by construction, exactly the same value of ρ_{ac}, and so only LCC is plotted in the upper panels, and the

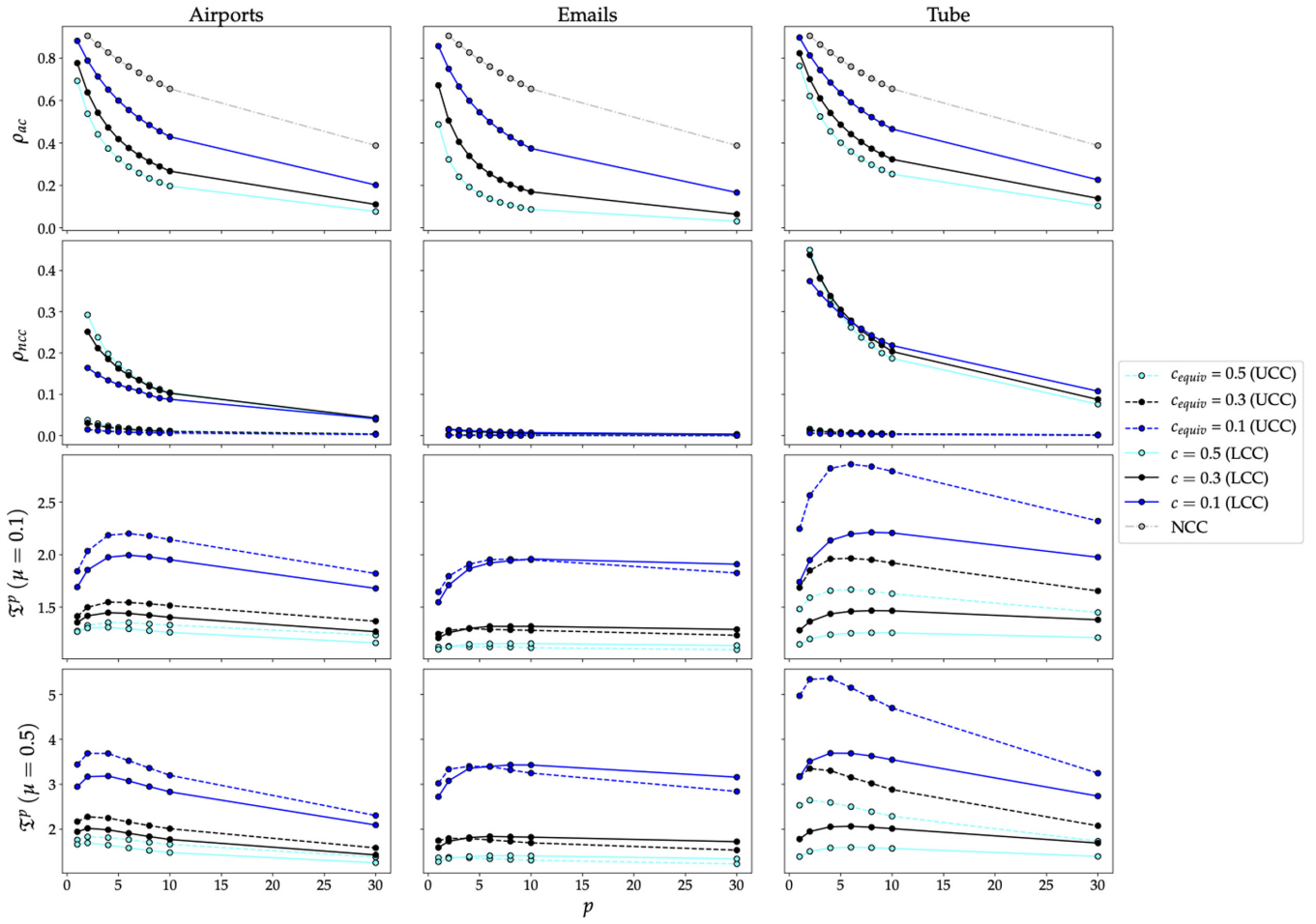

FIG. 10. Average autocorrelation and neighborhood correlation of a link, and rescaled time to equilibrium for diffusion for both the LCC and UCC coupling models on different backbones as a function of the memory length p. The value of the average ρ_{ac} (first row), and ρ_{ncc} (second row), where averages are taken over links in a CDARN(p) temporal network with local cross-correlation (solid line, LCC), uniform cross-correlation (dashed line, UCC), and no cross-correlation (dash-dot line, NCC) coupling, for each backbone. The third and fourth rows display the rescaled average time until equilibrium for a diffusion process on these networks with diffusion constants $\mu = 0.1$ and $\mu = 0.5$, respectively. Note that UCC is not included in the first row (ρ_{ac}) as its values are, by construction, precisely the same as those of the LCC model. The NCC model is not included in the second row (ρ_{ncc}) as its value is always zero. Memory strength q is kept constant at 0.95 to ensure that memory plays a significant role in the evolution of the network and link density y is kept at 0.1 to ensure that there is sufficient time for any effects of memory to be observed. The coupling strength c and diffusion speed μ are varied. Note that for the UCC model we assign the curves the value c_{equiv} rather than c; this is because we chose the values of c to match the value of ρ_{ac} for the UCC and LCC coupling models, as such c_{equiv} refers to the value of c in the LCC model that is being matched. The backbones were taken from a collection of real data sets. The rescaled times to equilibria were averaged over 2×10^4 realizations of the process.

NCC model must always have $\rho_{ncc} = 0$, and so it is not plotted in the lower panels.

We observe here that, as expected, the value of ρ_{ac} in the NCC model is significantly higher than for the LCC model. We also see that both ρ_{ac} and ρ_{ncc} decay as the memory length p increases, consistent with the DARN(p) temporal network model. Most notably, there are significant differences between the values of ρ_{ncc} given for each backbone. While both the Airport and Tube backbones display significant neighborhood correlations in the LCC coupling model (though the values are larger for the Tube backbone), the Emails backbone has notable neighborhood correlations only for low values of p and, indeed, at $p = 30$ is practically indistinguishable from the NCC model. Also, as expected the value of ρ_{ncc} for the UCC model is always approximately 0. We have tested our hypothesis by comparing the rescaled time to equilibrium (see the lower two rows of Fig. 10) $\mathfrak{T}^p$ for both the LCC and UCC models, both generated and plotted in exactly the same way as was done for Fig. 8. It is clear that when the value of ρ_{ncc} is large, as in the Tube backbone, diffusion on the LCC coupling model is always faster. When ρ_{ncc} is lower, as in the Airports backbone, this is still true. Finally, when there are little to no correlations between neighbors, as with the Email backbone, diffusion on the LCC coupling model is faster than on the UCC model only for small values of p; after this point the value of ρ_{ncc} is so small that its effects are no longer apparent. While the results for the Email backbone indicate that correlations among neighbors are not

the only influence on behavior, it is clear that their presence does act to speed diffusion processes over the temporal network.

Such a result is in effect more general and holds also for other spreading processes taking place over temporal networks with correlated link dynamics. For example, randomization-based studies [81] have shown that cross-correlations of links can facilitate epidemics described by the SI process [82] or more complex contagion dynamics [83]. A precise comparison between the two approaches to the problem is left for future research.

VI. CONCLUSIONS

The influence that memory in temporal networks has on the dynamics of the network links, as well as on processes that run on a temporal network is increasingly seen as fundamental to better understand our highly networked world. Different mechanisms can lead to links depending on each others with a given memory of the past in real-world networks. For example, dynamic patterns of node activities may induce link correlations, possibly non-Markovian, as well as link-specific dependencies. In general, a full understanding of these mechanisms is still lacking, thus the development of models to generate and study synthetic temporal networks with tunable memory patterns in their link dynamics is key.

In the context of spreading processes on temporal networks, such as the transmission of diseases or the diffusion of information, it has been observed that the presence of memory can either speed up or slow down the spreading relative to some memoryless case. Similar results in this direction have been found here, in a manner that is reminiscent of other recent findings. What is generally less well understood is precisely how memory causes this change in the speed of spreading and the role played by the multivariate structure of interactions of the link dynamics. A great deal of work has gone into the studying of how the correlated bursts in link activity that are the result of nonexponential inter link times, and hence memory, slow spreading processes. This, however, does not yet give us a full picture.

In this article we have introduced a flexible and controllable class of generative models of temporal networks in which the activity of a link is ruled by a discrete autoregressive mechanism of copying the activity of a link from the past, either itself (autocorrelations), or a different link (cross-correlations). Such discrete autoregressive network models allow for arbitrary backbone topologies, and precise control over the memory strength, memory length, average degree, and coupling strength. Within this class of models, we have considered and studied in detail a simplified model with a few parameters, which can be interpreted as the temporal generalization of the Erdős-Renyi random graph with non-Markovian memory, backbone structure, and both self- and cross-interactions of links. Not only the model is simple enough for some analytical results, but thanks to maximum likelihood estimation methods, it also allows applications to empirical data. Together with this baseline version of the model, we have studied other generalizations that well capture the heterogeneous and time-varying patterns that can characterize real-world systems.

Hence, first of all, we have empirically proved that, despite the simplicity of the baseline model, it is possible to infer many memory patterns observed in the real world. By estimating the model on a number of data sets, we have shown that: transportation networks are Markovian systems with significant cross-correlations of links capturing the presence of transport connections between the physical nodes, i.e., stops, of the network; social networks (both online and offline) are in general non-Markovian with a memory order larger than one, and display autocorrelations of links, as a result of the stability pattern characterizing some social ties, such as friendship; football networks are Markovian and display both link-specific persistence and cross interactions between links, with the two memory patterns which are inversely correlated as functions of the time resolution, as a result of the underlying contact dynamics of the game. We have then shown how heterogenous patterns of memory strengths can be captured by generalizing the baseline model with link-specific parameters. This allows us, among other things, to point out that social networks display spatial heterogeneity, i.e., heterogenous degree distribution, but homogeneous and persistent contacts. We have studied also the case of time-varying patterns in the memory of links, by considering a generalized model with dynamic parameters. While some empirical systems, such as transportation and contact networks, display almost constant memory over time, we have found instead that football networks are highly complex also from a dynamic point of view. Finally, within our framework, we have further proved the importance of describing the full structure of correlations of links with a study of link prediction, in particular showing that we are able to improve significantly the forecasting performance when such effects are included.

The model is rich but flexible enough to allow the analytical treatment of a number of theoretical problems, including finding the exact correlations between any two links. Given this we have been able to study exactly how the memory and coupling of the dynamics of different links in our model influence a spreading processes on the network. In doing this we have provided a solution to the time taken for a diffusion process to reach equilibrium in the limit of no cross-correlations and as a function of the memory length p. This shows us that the spreading time is nonmonotonically dependent on p, and allows us to infer that the equivalent memoryless process provides the fastest possible diffusion in our model. Looking at networks we have shown that correlations play a subtler role than might previously been expected. While we find, in accordance with previous works, that nonexponentially decaying autocorrelations among links do slow down diffusion, we, surprisingly, see that the opposite is true of local correlations. When links that share a node are correlated, this tends to speed up diffusion. This is made clear by the fact that when we observe a system in which links have fixed autocorrelation, but the correlations between neighboring links varies (while all other parameters are kept constant), then diffusion is faster when correlations among neighbors are higher. This has strong implications for real-world systems. While it is understood that memory and correlations between links have an effect on the spreading of information, the observation that correlations between neighbors and autocorrelations behave in opposite ways directly contributes to our understanding

of many empirical systems. For example, when considering the diffusion of information over a social network, and any consequent formation of opinions, correlations between two different social ties must be considered as important as the correlation of a social tie with its own history. In a more general sense our findings also suggest that considering the evolution of links as independent processes in a temporal network means we loose a significant amount of information. Hence, when assessing the properties of an empirical network, correlations between the evolutions of links must be taken into account. Finally, we have been able to test our model using as backbones the topologies of real-world systems. The differences in spreading behavior demonstrated among these backbones show the important role that such topologies play. It is clear that memory, correlations and backbone interact in a complex manner, and when considering the study of real-world systems one can not assume to study of any of these features in isolation. Here, however, we have provided a framework in which the interplay between these features can be studied systematically, and how surprising results occur when we do.

The codes for both simulation and estimation of the CDARN(p) models proposed in this paper are available online at [84]. Documentation can be found at [85].

ACKNOWLEDGMENTS

The work of V.L. was funded by the Leverhulme Trust Research Fellowship "CREATE: The network components of creativity and success." F.L. and P.M. acknowledge support by the European Union's Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019—Integrating Activities for Advanced Communities," Grant No. 871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" [86]. We also thank Lucas Lacasa for useful discussions.

All the authors designed the research. O.W. and P.M. performed calculations and generated figures. All the authors discussed intermediate and final results and wrote the paper. The authors declare no competing interests, financial or otherwise.

APPENDIX

1. Linear indexing for links

Throughout this work we frequently make use of a linear indexing for links in a network, and, equivalently, entries in a matrix. In practice this is a way to take a pair of indices for either network nodes, or rows and columns of the adjacency matrix, say (i, j), and map it to a single number ℓ. As the only requirement we need for such a mapping to be bijective, that is each unique pair (i, j) corresponds to a unique value ℓ. As an example, the simplest way this can be done, assuming $i, j \in \{1, \ldots, N\}$ for some value N, is to take $\ell = N(i-1) + j$ element of a matrix as 1, then proceeding left to right line by line.

2. Average degree of the CDARN(p) network

Our aim is to show that a CDARN(p) network on a complete backbone has the same average degree as a DARN(p) network, and hence as an Erdős-Renyi (ER) random graph. To do this we need only show that the average value of an arbitrary link is given by $\langle a_t^{ij} \rangle = y$. Let us proceed by first averaging over the left- and right-hand sides of Eq. (4) to get

$$\left\langle a_t^{ij} \right\rangle = q\left\langle a^{M_t^{ij}}_{(t-Z_t^{ij})} \right\rangle + (1-q)y, \tag{A1}$$

where the symbols $\langle \cdot \rangle$ denote temporal averages. If we now label the link (i, j) with its linear index $\ell = 1, 2, \ldots, L$, we obtain the following:

$$\begin{aligned} \left\langle a_t^{\ell} \right\rangle &= \frac{q}{p} \sum_{s=1}^{p} \sum_{\ell'=1}^{L} c^{\ell\ell'} \left\langle a_{t-s}^{\ell'} \right\rangle + (1-q)y, \\ &= q \sum_{\ell'=1}^{L} c^{\ell\ell'} \left\langle a_t^{\ell'} \right\rangle + (1-q)y. \end{aligned} \tag{A2}$$

where $c^{\ell\ell'}$ are the entries of the coupling matrix $\underline{\underline{C}}$. In the above we have made use of the stationarity of the sequence a_t^{ℓ} to say that $\langle a_{t-a}^{\ell} \rangle = \langle a_t^{\ell} \rangle$. One can see also that $\langle a_t^{\ell} \rangle = \bar{a}$, for some constant $\bar{a}$, is a solution to the above equations. The fact that $\underline{\underline{C}}$ is row stochastic, and so its rows sum to 1, then gives us that $\bar{a} = y$ is the unique solution. Hence, we have obtained $\langle a_t^{\ell} \rangle = y$, that is, we have shown that the CDARN(p) produces networks with the same average degree of the DARN(p) model.

3. The infinite memory limit

One of the key features of the DARN(p) model is that, as $p \to \infty$, it produces temporal networks that are indistinguishable from a sequence of independent ER graphs. Our aim now is to show that this is true for the CDARN(p) model as well. We start by writing the conditional probability for a single link with linear index ℓ:

$$\mathrm{Prob}\left(a_t^{\ell} = 1 | \{\underline{\underline{A}}_s\}_{s=t-1}^{t-p}\right) = (1-q)y + q\phi_t(p), \tag{A3}$$

where $\{\underline{\underline{A}}_s\}$ is the random matrix representing the adjacency matrix at time s, which we say has observed values a_s^{ℓ}. We hence see that our problem can be reduced to a study of the properties of some kernel function ϕ, defined as the probability that a 1 is drawn from any point in the memory, i.e.,

$$\phi_t(p) = \sum_{\ell'} c^{\ell\ell'} \frac{1}{p} \sum_{k=1}^{p} a_{t-k}^{\ell'}. \tag{A4}$$

We recognize the sample expectation over the past p steps of the time series, and so can see that $\phi_t(p) \to y$ as $p \to \infty$. For completeness we must also check that any fluctuations away from the mean can be ignored at finite times. First, we see the following:

$$\begin{aligned} \phi_{t+1}(p) - \phi_t(p) &= \sum_{\ell'} c^{\ell\ell'} \frac{1}{p} \sum_{k=1}^{p} \left(a_{t-k+1}^{\ell'} - a_{t-k}^{\ell'}\right), \\ &= \frac{1}{p} \sum_{\ell'} c^{\ell\ell'} \left(a_{t-k+1}^{\ell'} - a_{t-k}^{\ell'}\right). \end{aligned} \tag{A5}$$

We then have that $a^{\ell'}_{t-k+1} - a^{\ell'}_{t-k} \in \{-1, 0, 1\}$, and so

$$-\frac{1}{p} \leqslant \phi_{t+1}(p) - \phi_t(p) \leqslant \frac{1}{p}, \tag{A6}$$

$$\Rightarrow y - \frac{t}{p} \leqslant \phi_t(p) \leqslant y + \frac{t}{p}. \tag{A7}$$

Hence, in the large p limit then the memory kernel ϕ tends to 0, and so the system is equivalent to one in which there is no memory. This displays exactly the same behavior as is found for the DARN(p) model: the CDARN(p) model does indeed tend to a memoryless model in the limit of large memory. In the memoryless case we note that $\text{Prob}(a^\ell_t = 1) = y$, and so must have an expected interlink time of $1/y$, and correspondingly the expected time until the nth link is n/y.

4. Time to equilibrium in a two-node system with a permanent link

Consider the equations defining diffusion in continuous time between two nodes, for which the link between them is permanent (always present):

$$\dot{d}^1(t) = -\mu[d^1(t) - d^2(t)]. \tag{A8}$$

If we impose conservation, i.e., $d^2(t) = 1 - d^1(t)$ [and drop the 1 so that $d^1(t) \to d(t)$], we can rewrite this as

$$\dot{d}(t) = \mu - 2\mu d(t). \tag{A9}$$

Assuming $d(0) = 1$, the solution is

$$d(t) = \tfrac{1}{2}(e^{-2\mu t} + 1). \tag{A10}$$

We say that this system has reached equilibrium at the first time (in a continuous sense) $t = \tau_c$ where $|d^1(t) - d^2(t)| < \epsilon$ for some small positive ϵ. Again, imposing conservation this can be rewritten as the first value of t such that $2d(t) - 1 = \epsilon$. With Eq. (A10) we can then find the (continuous) time to equilibrium directly as

$$\tau_c = \frac{-\log \epsilon}{2\mu}. \tag{A11}$$

Hence, the number of full time steps τ of length Δt which must occur before equilibrium is reached is given by

$$\tau = \left\lfloor \frac{-\log \epsilon}{2\mu \Delta t} \right\rfloor. \tag{A12}$$

Note that for this system, for any given values of μ and Δt we can always find $\bar{\mu} = \mu \Delta t$, meaning that we may fix $\Delta t = 1$ and still recover the full range of possible values for $\bar{\mu}$ by varying μ.

5. The transition matrix for a DAR(p) variable

Consider a stochastic process where the random variable X_t is governed by the the DAR(p) model:

$$X_t = Q_t X_{(t-Z_t)} + (1 - Q_t) Y_t, \tag{A13}$$

where, for each t, $Q_t \sim \mathcal{B}(q)$ and $Y_t \sim \mathcal{B}(y)$ are Bernoulli random variables, while Z_t picks integers uniformly from the set $\{1, \ldots, p\}$. This can be thought of as a pth-order Markov chain, and so is equivalent to a first-order Markov chain in an enlarged state space [80]. Accordingly, we define the so-called "p-state" of link (i, j) at time t, by combining the state of the link at time t along with its previous $p - 1$ states in the vector $\underline{S}_t = (X_t, X_{t-1}, \ldots, X_{t-p+1})$. If we now define the set $\mathcal{S}$ as the set containing all 2^p possible p-states, then for any $\alpha, \beta \in \mathcal{S}$ we can look at the conditional probability $\text{Prob}(\underline{S}_{t+1} = \beta | \underline{S}_t = \alpha)$. This defines the entries $T_{\alpha\beta}$ of the pth-order $2^p \times 2^p$ transition matrix. More details on the transition matrix of the DAR(p) model can be found in [35].

6. State indexing

When writing the matrix element $T_{\alpha\beta}$ we are implicitly associating an index to the p states α and β. Since elements of a matrix are usually labeled by values $i, j \in \{0, \ldots, I-1\}$ (or $i, j \in \{1, \ldots, I\}$) for some value of I, we must hence impose an ordering on the states $\alpha, \beta \in \mathcal{S}$. This is done by associating a linear index $l(\alpha)$ to each possible state $\alpha \in \mathcal{S}$ (and similarly for β). The simplest form of this labeling function in our case, given a memory length of p, is

$$l(\alpha) = \sum_{k=0}^{p} 2^k \alpha_k, \tag{A14}$$

where α_k is the kth entry in the p-state vector associated with α. In practice this is equivalent to consider the sequence of 0's and 1's, representing the link history corresponding to state α, as a binary number and converting in into to a decimal number. We will implicitly assume that wherever we use α, or any state in $\mathcal{S}$, we are referring to the label $l(\alpha)$, and that the labeling function is as given in Eq. (A14).

7. Initializing CDARN(p) model simulations

When generating realizations of the CDARN(p) model for the purposes of Monte Carlo simulation (or any other simulation) it is important to ensure that the model is appropriately initialized. In all of the simulations and calculations here we require that the model be in a steady state, and so before any simulated diffusion starts we do the following:

(1) For each link $\ell \in \{1, \ldots, L\}$ and for each time $s \in \{0, \ldots, p-1\}$ assign to link state a^ℓ_s the value taken by a random variable $X^\ell_s \sim \mathcal{B}(y)$. This gives us a set of preinitial conditions from which simulation can be started.

(2) Simulate the CDARN(p) model for times $p, \ldots, T_0$ using the previously generated states of the network, for some large T_0. T_0 is chosen so that the network has reached a steady state, as approximated by the point where the autocorrelations $\langle a^\ell_p a^\ell_{T_0} \rangle$ have decayed below some suitable threshold. Functionally this has been set at approximately $T_0 = 500$.

(3) Any simulation on top of the network can now start. This process must be repeated for every simulated realization.

8. Average time to equilibrium for a single Markov link

The value $\langle \tau^1 \rangle$ of the average time until diffusion across a single DAR(1) link reaches equilibrium, is central to any analysis of the rescaled times to equilibrium. Hence, we calculate it explicitly here. We know that, given the value $\tau = n$ from Eq. (A12) giving the number of time steps until equilibrium in the two-node system where the link is always present, $\langle \tau^1 \rangle$

will be precisely the time taken for a DAR(1) link to occur n times. This can be found as the solution to the following equation:

$$\langle \tau^1 \rangle = \underline{\omega}^T \left(\sum_{t=0}^{n-1} \underline{\underline{h}}^t \right) \underline{k}, \tag{A15}$$

where $\underline{\underline{h}}^t$ denotes the t_{th} power of the matrix $\underline{\underline{h}}$. Now $\underline{\underline{h}}$ will be a 2×2 matrix, and $\underline{\omega}$ and $\underline{k}$ will be two-dimensional vectors. Given the definition of $h_{\alpha\beta}$ as the probability that a system starting in state α ends in state β, we can easily see that the following must be true:

$$\underline{\underline{h}} = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \tag{A16}$$

and so

$$\sum_{t=0}^{n-1} \underline{\underline{h}}^t = \begin{pmatrix} 1 & n-1 \\ 0 & n \end{pmatrix}. \tag{A17}$$

Similarly we may find that $\omega_1 = 1-y$ and $\omega_2 = y$, and $k_1 = [(1-q)y]^{-1}$ and $k_2 = y^{-1}$. This gives us the equation

$$\langle \tau^1 \rangle = \left(\frac{1}{1-q} + n - 1 \right) \frac{1-y}{y} + n. \tag{A18}$$

It is then a simple matter to extend this result and calculate the value of the rescaled time to equilibrium $\mathfrak{T}^1$ directly. Given that we know $\langle \tau^p \rangle \to n/y$ as $p \to \infty$, and this is precisely the value of $\langle \tau^0 \rangle$, we then find

$$\mathfrak{T}^1 = \frac{\langle \tau^1 \rangle}{\langle \tau^0 \rangle} = \left(\frac{1}{1-q} + n - 1 \right) \frac{1-y}{n} + y. \tag{A19}$$

It is easy to see that the maxima and minima of this function in terms of n and y are finite and occur at their limiting values ($y = 0, 1$ and $n = 1, \infty$ respectively) if $q \neq 1$. However, in the limit $q \to 1$ we see that $\mathfrak{T}^1 \to \infty$. In the $q = 0$ limit we obtain the value $\mathfrak{T}^\infty = 1$, as expected.

9. Rescaled time to equilibrium in the limit of large p

In the main text we claim that $\mathfrak{T}^p \geqslant 1$ for suitably sparse initial conditions, i.e., when y is small, but that for $y \approx 1$ the opposite can be true. To understand this we first formalize our statement: given an initial probability vector $\underline{\omega}$, we have $\langle \tau^p \rangle \geqslant \langle \tau^\infty \rangle$, provided that the entries representing states in which no link is present (in our case ω_α for $\alpha \in \{0, \ldots, 2^{p-1}\}$) contain the majority of the probability mass. Recall, first, that we are implicitly labeling our states $\alpha \in \mathcal{S}$ according to the labeling function given in Eq. (A14). Now, notice that, since $\underline{\omega}$ is a probability vector and $\underline{\underline{h}}$ is a stochastic matrix, we can define a vector $\underline{k}^0$ such that $k^0_\alpha = 1/y$ for all α, and we can write the following equation:

$$\langle \tau^\infty \rangle = \underline{\omega}^T \left(\sum_{t=0}^{n-1} \underline{\underline{h}}^t \right) \underline{k}^0. \tag{A20}$$

Hence, we can write

$$\langle \tau^p \rangle - \langle \tau^\infty \rangle = \underline{\omega}^T \left(\sum_{t=0}^{n-1} \underline{\underline{h}}^t \left(\underline{k} - \underline{k}^0 \right) \right). \tag{A21}$$

By construction $h_{\alpha\beta} = 0$ if $\beta < 2^{p-1}$, and so if we define

$$\underline{\tilde{\omega}}^T = \underline{\omega}^T \left(\sum_{t=0}^{n-1} \underline{\underline{h}}^t \right), \tag{A22}$$

then, when $\beta < 2^{p-1}$, we have $\tilde{\omega}_\beta = \omega_\beta$. Hence, we see that if $\omega_0 \approx 1$ (the entry in ω representing an initial state with no links), then we need only check that $k_0 > 1/y$ to show that $\mathfrak{T}^p \geqslant 1$. This can be checked directly by analyzing the average time taken to reach equilibrium k_α, given some starting state α, as defined by the following set of linear equations:

$$k_\alpha = 1 + T_{\alpha\alpha'} k_{\alpha'}, \tag{A23}$$

where $\alpha' = \lfloor \alpha/2 \rfloor$. From this we can directly obtain

$$k_0 = \frac{1 - qp^{-1}}{(1-q)y}, \tag{A24}$$

and hence confirm that $k_0 > 1/y$ when $\omega_0 \approx 1$. We now want to understand the conditions in which this breaks down, and we instead observe $\mathfrak{T}^p < 1$. One can manually check that, for any values of p or q, $k_\alpha > 1/y$ for $\alpha = 0, 1$, but that this inequality does not generally hold for $\alpha = 4$. As a specific example of this, if we fix $p = 3$, $y = 0.01$ and $q = 0.1$, then $1/y = 100$, but $k_4 \approx 98.52$. To understand this behavior, we can then make use of the following two facts about k_α. Given a memory state $\alpha \in \mathcal{S}$,

(1) If by α_n we indicate the memory state with a 1 in the nth entry, and zeros elsewhere, then the values of k_{α_n} are given by solutions to the equation $x_{n+1} = 1 + a x_n$, with appropriate values for a and x_1.

(2) If β is the memory state obtained by taking memory state α and replacing any of its 0 states with 1, then $k_\beta < k_\alpha$.

To prove the first of these statements, we directly analyze Eq. (A23). This equation, in our α notation becomes $k_{\alpha_{n+1}} = 1 + T_{\alpha_{n+1},\alpha_n} k_{\alpha_n}$, but we also notice that $T_{\alpha_{n+1},\alpha_n}$ is invariant of n, always taking the value $T_{\alpha_{n+1},\alpha_n} = 1 - q/p - (1-q)y$, which we will now denote as T. To obtain the desired form of difference equation, we now simply identify $x_n = k_{\alpha_n}$ and $a = T$. This can be easily solved to give the following:

$$k_{\alpha_n} = \frac{T^n}{(1-q)y} + \frac{1 - T^n}{1 - T}. \tag{A25}$$

This equation must clearly be decreasing with n.

To prove the second of these statements consider two possible memory states α and β, where β is given by taking α and replacing one of the zeros in its memory with a one. Let us label the position of the state which we change with t. Now let us now denote $\alpha^{(n)} = \lfloor \alpha^{(n-1)}/2 \rfloor$, where $\alpha^0 = \alpha$, and similarly for β. To clarify, we can think of $\alpha^{(n)}$ as being the memory state α shifted back n times, or similarly what happens to the memory state of a DAR(p) process if it starts in state α and generates n zeros. We can then write the following equation directly from Eq. (A23):

$$k_{\alpha^{(n)}} - k_{\beta^{(n)}} = T_{\alpha^{(n)}\alpha^{(n+1)}} k_{\alpha^{(n+1)}} - T_{\beta^{(n)}\beta^{(n+1)}} k_{\beta^{(n+1)}}. \tag{A26}$$

Given our definition in Eq. (16), and the fact that β is α with a 1 added, we can rearrange this to give

$$k_{\alpha^{(n)}} - k_{\beta^{(n)}} = T_{\alpha^{(n)}\alpha^{(n+1)}} (k_{\alpha^{(n+1)}} - k_{\beta^{(n+1)}}) + \frac{q}{p} k_{\beta^{(n+1)}}. \tag{A27}$$

From this we can see that if $k_{\alpha^{(n+1)}} \geqslant k_{\beta^{(n+1)}}$ then we must have $k_{\alpha^{(n)}} \geqslant k_{\beta^{(n)}}$. Now, by construction we know that $\alpha^{(n)} = \beta^{(n)}\, \forall n \geqslant t$, since this is the point at which the additional 1 in the memory is removed. In turn this means that $k_{\alpha^{(n)}} = k_{\beta^{(n)}}\, \forall n \geqslant t$. Inductively this gives us that

$$k_{\alpha^{(t-1)}} \geqslant k_{\beta^{(t-1)}}, \ldots, k_{\alpha^{(n+1)}} \geqslant k_{\beta^{(n+1)}}. \tag{A28}$$

Thus, we have that $k_{\alpha^{(n)}} \geqslant k_{\beta^{(n)}}$. Hence, we have proved that k_α is decreased by adding a one at any point in the memory, and, equally, increased by adding a zero at any point in the memory.

The first statement gives us that, since we can not guarantee that $k_4 > 1/y$, we can not guarantee that, for any state α with a single one in any position other than 1, $k_\alpha > 1/y$. The second statement then tells us that, since any state α can be generated by taking a state with only a single 1 somewhere, and adding more 1's to it, we can never guarantee that $k_\alpha > 1/y$ for any $\alpha \geqslant 4$.

Because of this we see that for small y we must have $\langle \tau^p \rangle \geqslant \langle \tau^\infty \rangle$, and hence we must have $\langle \tau^p \rangle \geqslant \langle \tau^0 \rangle$, finally giving us that $\mathfrak{T}^p \geqslant 1$. However, for larger y this may not be the case.

10. Correlations in the CDARN(p) model

By introducing the possibility that a link in a DARN(p) network can draw from the memory of another link, and hence creating the CDARN(p) model, we have introduced correlations among the activities of different links. As we will show in this section, the extent of these correlations can be completely characterized analytically. If we have a network with L possible links, each with its own linear index, let us denote the correlations between link ℓ and ℓ' at time lag k as $\langle a_t^\ell a_{t-k}^{\ell'} \rangle = \rho_k^{\ell\ell'}$. Following the procedures in [35,55], we can derive the Yule-Walker equations:

$$\rho_k^{\ell\ell'} = \frac{q}{p} \sum_{a=1}^{p} \sum_{b=1}^{L} c^{\ell b} \rho_{k-a}^{b\ell'}, \tag{A29}$$

where the elements $c^{\ell b}$ are taken from the coupling matrix assigning the probability of a link ℓ drawing from the memory of link b. This can be written more compactly in terms of the corresponding matrices $\underline{\underline{\rho}}_k = \{\rho_k^{\ell\ell'}\}$ and $\underline{\underline{C}} = \{c^{\ell\ell'}\}$ as

$$\underline{\underline{\rho}}_k = \frac{q}{p} \underline{\underline{C}} \sum_{a=1}^{p} \underline{\underline{\rho}}_{k-a}. \tag{A30}$$

These equations can be solved given a suitable closure. Following [35], we can rewrite this expression for values of $k < p$ as

$$\underline{\underline{\rho}}_k = \frac{q}{p} \underline{\underline{C}} \left(\sum_{a=1}^{k-1} \underline{\underline{\rho}}_a + \sum_{a=1}^{p-k} \underline{\underline{\rho}}_a + \underline{\underline{\rho}}_0 \right), \tag{A31}$$

for some value $\underline{\underline{\rho}}_0$. This equation can be seen to have a constant solution $\underline{\underline{\rho}}$, which satisfies

$$\underline{\underline{\rho}} = \frac{q}{p} \underline{\underline{C}} ((p-1)\underline{\underline{\rho}} + \underline{\underline{\rho}}_0). \tag{A32}$$

Now we need to find a suitable expression for $\underline{\underline{\rho}}_0$. We know that, by definition, $\rho_0^{\ell\ell} = 1$. The off-diagonal entries, however, are given by the Yule-Walker equation

$$\rho_0^{\ell\ell'} = \frac{q}{p} \sum_{a=1}^{p} \sum_{b=1}^{L} c^{\ell b} \rho_a^{b\ell'}. \tag{A33}$$

But we know that the value of $\underline{\underline{\rho}}_a$ must be a constant $\underline{\underline{\rho}}$, and so the off-diagonal elements of $\underline{\underline{\rho}}_0$ will be the same as the off-diagonal elements of

$$\begin{aligned} \bar{\underline{\underline{\rho}}}_0 &= \frac{q}{p} \sum_{a=1}^{p} \underline{\underline{C}} \underline{\underline{\rho}}, \\ &= q \underline{\underline{C}} \underline{\underline{\rho}}. \end{aligned} \tag{A34}$$

Putting everything together we get the equation

$$\rho^{\ell\ell'} = \frac{q}{p} \left((p-1) \sum_{b=1}^{L} c^{\ell b} \rho^{b\ell'} + q \sum_{b \neq \ell'} \sum_{\ell''=1}^{L} c^{\ell b} c^{b\ell''} \rho^{\ell''\ell'} + c^{\ell\ell'} \right). \tag{A35}$$

This can be rearranged to give

$$\rho^{\ell\ell'} = \frac{q}{p} \left[\sum_{\ell''=1}^{L} \left((p-1) c^{\ell\ell''} + q \sum_{b \neq \ell'} c^{\ell b} c^{b\ell''} \right) \rho^{\ell''\ell'} + c^{\ell\ell'} \right]. \tag{A36}$$

This can be further simplified by constructing the tensor $\underline{\underline{\underline{\Delta}}}$ as

$$\Delta^{\ell\ell'\ell''} = \frac{q}{p} \left((p-1) c^{\ell\ell''} + q \sum_{b \neq \ell'} c^{\ell b} c^{b\ell''} \right). \tag{A37}$$

The system of equations given in Eq. (A32) can then be written as

$$\rho^{\ell\ell'} = \sum_{\ell''=1}^{L} \Delta^{\ell\ell'\ell''} \rho^{\ell''\ell'} + \frac{q}{p} c^{\ell\ell''}. \tag{A38}$$

This form is more easily dealt with in numerical applications, as a simple dimensional reduction (flattening) yields a more traditional form for a system of linear equations. Importantly, this solution relies on no properties of the coupling matrix other than stochasticity, which it must have by definition. In special cases, such as those where coupling is uniform or symmetric, we can simplify these equations further by analyzing the symmetries that arise in $\underline{\underline{C}}$ and $\underline{\underline{\underline{\Delta}}}$.

11. Evolution of the autocorrelation function

We wish to now show that the full extent of the autocorrelations in our model are described by the constant value $\underline{\underline{\rho}}$, given over the first p time steps. We first notice that from Eq. (A30) we can obtain the following:

$$\begin{aligned} \underline{\underline{\rho}}_k - \underline{\underline{\rho}}_{k-1} &= \frac{q}{p} \underline{\underline{C}} \left(\sum_{t=1}^{p} \underline{\underline{\rho}}_{k-t} - \sum_{t=1}^{p} \underline{\underline{\rho}}_{k-t-1} \right), \\ &= \frac{q}{p} \underline{\underline{C}} (\underline{\underline{\rho}}_{k-1} - \underline{\underline{\rho}}_{k-p-1}), \end{aligned} \tag{A39}$$

and hence

$$\underline{\underline{\rho}}_k - \left(\underline{\underline{I_d}} + \frac{q}{p} \underline{\underline{C}} \right) \underline{\underline{\rho}}_{k-1} = -\frac{q}{p} \underline{\underline{C}} \underline{\underline{\rho}}_{k-p-1}. \tag{A40}$$

However, we know that for $k \in \{1, \ldots, p\}$ the autocorrelation is a constant $\underline{\underline{\rho}}_k = \underline{\underline{\rho}}$, meaning that for $k \in \{p+1, \ldots, 2p+1\}$ Eq. (A40) becomes

$$\underline{\underline{\rho}}_k - \left(\underline{\underline{I_d}} + \frac{q}{p}\underline{\underline{C}}\right)\underline{\underline{\rho}}_{k-1} = -\frac{q}{p}\underline{\underline{C}}\underline{\underline{\rho}}. \tag{A41}$$

This is now a first-order inhomogeneous difference equation with the solution

$$\underline{\underline{\rho}}_k = \underline{\underline{\rho}} - e^{k \log\left(1+\frac{q}{p}\underline{\underline{C}}\right)}\underline{\underline{R}}, \tag{A42}$$

where $\underline{\underline{R}}$ is a constant matrix. By noticing that $\underline{\underline{\rho}}_{p+1} = q\underline{\underline{C}}\underline{\underline{\rho}}$ we obtain the expression

$$\underline{\underline{R}} = (1-q)\underline{\underline{\rho}}\left(1 + \frac{q}{p}\right)^{-(p+1)}. \tag{A43}$$

With this solution we see that the equation governing the values of $\underline{\underline{\rho}}_k$, for k in the range $p+1$ to $2p+1$, is of the form

$$\underline{\underline{\rho}}_k - \bar{q}\underline{\underline{\rho}}_{k-1} = -\underline{\underline{A}} + e^{-\underline{\underline{\lambda}}k}\underline{\underline{B}}, \tag{A44}$$

where $\bar{q}$, $\underline{\underline{A}}$, $\underline{\underline{B}}$, and $\underline{\underline{\lambda}}$ are constant matrices. This equation has a general solution

$$\underline{\underline{\rho}}_k = \underline{\underline{A}}' - e^{-\underline{\underline{\lambda}}k}\underline{\underline{B}}', \tag{A45}$$

where $\underline{\underline{A}}'$ is a constant matrix, and $\underline{\underline{B}}'$ is a matrix that is constant over every interval $k \in [np+1, (n+1)p+1]$. Moreover, in our specific case we find that $\underline{\underline{A}}' = \underline{\underline{\rho}}$ and $\underline{\underline{\lambda}} = \log(1 + \frac{q}{p}\underline{\underline{C}})$. This implies that not only is the autocorrelation function for the CDARN(p) process exponentially decreasing for all values of k larger than $p+1$, but also that this decay varies according to a single parameter $\underline{\underline{B}}'$ every p time steps. This gives us a full picture of the autocorrelations for a CDARN(p) process.

12. Special case: Totally symmetric cross-correlation

The simplest type of correlation in the CDARN(p) model occurs when the coupling matrix $\underline{\underline{C}}$ is such that $c^{\ell\ell'} = 1/L \forall \ell, \ell'$, i.e. regardless of the pair of links in question. We can immediately notice that our tensor $\underline{\underline{\underline{\Delta}}}$ now takes the form

$$\Delta^{\ell\ell'\ell''} = \frac{q}{p}\left((p-1)\frac{1}{L} + q\frac{L-1}{L^2}\right), \tag{A46}$$

which is invariant over the three indexes ℓ, ℓ', and ℓ''. Consequently $\rho^{\ell\ell'}$ must be invariant over ℓ and ℓ'. Hence all of the lagged correlations have the same value, and we can write $\rho^{\ell\ell'} = \rho$ and $\Delta^{\ell\ell'\ell''} = \Delta$, giving us the following equation:

$$\rho = L\Delta\rho + \frac{q}{Lp}. \tag{A47}$$

Solving for ρ gives

$$\rho = \left[Lp\left(\frac{1}{q} - 1\right) + (1-q)L + q\right]^{-1}. \tag{A48}$$

This provides a full picture of the lagged correlations present in the system. All that remains is to find the time 0 correlations $\rho_0^{\ell\ell'}$ when $\ell \neq \ell'$. This can be done as follows:

$$\rho_0^{\ell\ell'} = \frac{q}{Lp}\sum_{a=1}^{p}\sum_{b=1}^{L}\rho_a^{b\ell'} = \frac{q}{Lp}L\sum_{a=1}^{p}\rho_a = q\rho. \tag{A49}$$

Here the last line is given by the fact that for $a = 1, \ldots, p$ we have $\rho_a = \rho$, the constant value given in Eq. (A48). Hence, when $\ell \neq \ell'$, $\rho_0^{\ell\ell'} = q\rho$.

Note first that if $L = 1$ then we recover the autocorrelation function of a DAR(p) process. Also note that as L increases this value must decrease, meaning that for large networks both correlations and autocorrelations are removed, and so memory no longer has any effect on the evolution of the system.

13. Special case: Uniform cross-correlation (UCC)

The second special case we will consider is that of uniform cross-correlation (UCC), as induced by a symmetric coupling matrix. Specifically this means that we require that $\underline{\underline{C}}$ be symmetric, with $c^{\ell\ell} = 1 - c$ for all values of ℓ and some given value of c, and $c^{\ell\ell'} = \bar{c}$ for $\ell \neq \ell'$ with $\bar{c} = c/(L-1)$. Going back to the general case in Eq. (A38) we notice that these conditions ensure that $\rho^{\ell\ell}$ is invariant with respect to ℓ, and when $\ell \neq \ell' \rho^{\ell\ell'}$ is invariant with respect to ℓ and ℓ'. This means that all of the values of $\rho^{\ell\ell'}$ can be found as the solutions to the two following equations (note that $\ell \neq \ell'$ is assumed here):

$$\begin{aligned}
\rho^{\ell\ell} &= \sum_{\ell''\neq\ell}\Delta^{\ell\ell\ell''}\rho^{\ell''\ell} + \Delta^{\ell\ell\ell}\rho^{\ell\ell} + \frac{q}{p}(1-c),\\
\rho^{\ell\ell'} &= \sum_{\ell''\neq\ell,\ell'}\Delta^{\ell\ell'\ell''}\rho^{\ell''\ell'} + \Delta^{\ell\ell'\ell}\rho^{\ell\ell'} + \Delta^{\ell\ell'\ell'}\rho^{\ell'\ell'} + \frac{q}{p}\bar{c}.
\end{aligned} \tag{A50}$$

Hence we need only find the relevant values of Δ to proceed. Given the definition of Δ and c and $\bar{c}$ we can find the following:

$$\begin{aligned}
\Delta^{\ell\ell\ell} &= \frac{q}{p}[(p-1)(1-c) + q(L-1)\bar{c}^2] =: \Delta^1,\\
\Delta^{\ell\ell\ell''} &= \frac{q}{p}\{(p-1)\bar{c} + q\bar{c}[(1-c) + (L-2)\bar{c}]\} =: \Delta^2,\\
\Delta^{\ell\ell'\ell} &= \frac{q}{p}\{(p-1)(1-c) + q[(1-c)^2 + (L-2)\bar{c}^2]\}\\
&=: \Delta^3,\\
\Delta^{\ell\ell'\ell'} &= \Delta^{\ell\ell\ell''} =: \Delta^2,\\
\Delta^{\ell\ell'\ell''} &= \frac{q}{p}\{(p-1)\bar{c} + q\bar{c}[2(1-c) + (L-3)\bar{c}]\} =: \Delta^4.
\end{aligned} \tag{A51}$$

Noticing that $\Delta^{\ell\ell\ell''}$ and $\Delta^{\ell\ell'\ell''}$ are invariant of ℓ, ℓ', and ℓ'' (down to excluded values) we then obtain the following pair of equations:

$$\begin{aligned}
\rho^{\ell\ell} &= (L-1)\Delta^2\rho^{\ell\ell'} + \Delta^1\rho^{\ell\ell} + \frac{q}{p}(1-c),\\
\rho^{\ell\ell'} &= [\Delta^3 + (L-2)\Delta^4]\rho^{\ell\ell'} + \Delta^2\rho^{\ell\ell} + \frac{q}{p}\bar{c}.
\end{aligned} \tag{A52}$$

This gives us a simple, solvable pair of equations. Note that while the full solution in terms of q, p, c, and L is easy to

obtain now, we will not write it here due to its length. To find the time 0 correlations we can use Eq. (A34) to obtain

$$\rho_0^{\ell\ell'} = q\{[(1-c)+(L-2)\bar{c}]\rho^{\ell\ell'} + \bar{c}\rho^{\ell\ell}\}. \tag{A53}$$

Competing the description of the correlations for the UCC model.

14. Uniform cross-correlation in the large network limit

The uniform cross-correlation (UCC) model for CDARN(p) networks was introduced to evenly distribute any temporal cross-correlations between links over the entire network. In doing this we minimize the cross correlations, i.e., $\rho^{\ell\ell'}$ where $\ell \neq \ell'$, for fixed values of p, q, y, and c. In turn this minimizes the influence that any such cross-correlations have on the diffusion process over the network. What we now show is that when the backbone of the temporal network has a large number of links then the UCC model is indistinguishable from a DARN(p) model that has been restricted to the same backbone, and hence temporal cross-correlations between links are completely removed.

Consider a temporal network given by time varying adjacency matrix $\underline{\underline{A}}_t$ with observed value $\{a_t^\ell\}$, and with L links, generated by the CDARN(p) model with link density y, memory strength q, memory length p and coupling matrix $\underline{\underline{C}}$ as in the UCC case. The conditional probability of a link ℓ occurring at time t, given the past p states of the network can be thought of in terms of contributions from the memory of the link itself, the memory of all other links, and some background contribution. This can hence be written as follows:

$$\text{Prob}\big(a_t^\ell | \{\underline{\underline{A}}_s\}_{s=t}^{t-p}\big) = (1-q)y + q[(1-c)\phi_{\text{self}} + c\phi_{\text{other}}], \tag{A54}$$

where ϕ_{self} and ϕ_{other} represent the contributions to the conditional probability $\text{Prob}(a_t^\ell | \{\underline{\underline{A}}_s\}_{s=t}^{t-p})$ from the past p states of the link ℓ and every other link, respectively.

For links to be effectively independent then we require that as $L \to \infty$, ϕ_{other} tends to a constant, and hence the link ℓ has no memory of the past states of any other link. To show this we study the memory kernels ϕ_{self} and ϕ_{other} directly as

$$\phi_{\text{self}} = \frac{(1-c)}{p}\sum_{k=1}^{p} a_{t-k}^{\ell}, \quad \phi_{\text{other}} = \frac{c}{(L-1)p}\sum_{\ell'\neq\ell}\sum_{k=1}^{p} a_{t-k}^{\ell'}. \tag{A55}$$

We need focus only on ϕ_{other}. First, let us consider the average value $\langle a_{t-k}^{\ell'}\rangle_{\ell'}$. The CDARN($p$) network is taken to be in a stationary state, and so the symmetry of the links under any relabeling guarantees us that $\text{Prob}(a_{t-k}^{\ell'})$ is the same for each link ℓ' and for each time $t-k$. Hence, we can write $\text{Prob}(a_{t-k}^{\ell'}) = \bar{a}\forall\ell'$ for some constant $\bar{a}$. Then we must have, for any of the $L-1$ possible values of ℓ',

$$\left\langle a_{t-k}^{\ell'}\right\rangle_{\ell'} = \text{Prob}\big(a_{t-k}^{\ell'}\big) = \bar{a}. \tag{A56}$$

Now ϕ_{other} can be rewritten as follows:

$$\phi_{\text{other}} = \frac{1}{p}\sum_{k=1}^{p}\frac{1}{L-1}\sum_{\ell'\neq l} a_{t-k}^{\ell'}. \tag{A57}$$

Then by the law of large numbers we can express this in terms of the sample average:

$$\begin{aligned}\phi_{\text{other}} &= \frac{1}{p}\sum_{k=1}^{p}\langle a_{t-k}^{\ell'}\rangle_{\ell'},\\ &= \frac{1}{p}\sum_{k=1}^{p}\bar{a},\\ &= \bar{a}.\end{aligned} \tag{A58}$$

Hence, $\phi_{\text{other}} \to \bar{a}$ as $L \to \infty$. Indeed, we can further see that $\bar{a} = y$. Since there are no terms containing links other than ℓ in ϕ_{self}, then we can conclude that the conditional probability is such that, in the same limit $L \to \infty$,

$$\text{Prob}\big(a_t^\ell = 1 | \{\{\underline{\underline{A}}_s\}_{s=t}^{t-p}\big) \to \text{Prob}\big(a_t^\ell = 1 | \{a_s^\ell\}_{s=t}^{t-p}\big), \tag{A59}$$

and so any memory of other links is lost. To show that this is equivalent to a DARN(p) network we need only look at the conditional probability of obtaining a link in such a network with memory strength $\bar{q}$, memory length p, link density $\bar{y}$, and adjacency matrix $\underline{\underline{E}}_t$ with observed values $\{e_t^\ell\}$:

$$\text{Prob}\big(e_t^\ell = 1 | \{\{\underline{\underline{E}}_s\}_{s=t}^{t-p}\big) = (1-\bar{q})\bar{y} + \frac{\bar{q}}{p}\sum_{k=1}^{p} e_{t-k}^{\ell}. \tag{A60}$$

Now, by setting the values of $\bar{q}$ and $\bar{y}$, in terms of the values q, y, and c from the CDARN(p) model, to be

$$\bar{q} = q(1-c), \quad \bar{y} = y, \tag{A61}$$

we obtain that

$$\text{Prob}\big(e_t^\ell = 1 | \{\{\underline{\underline{A}}_s\}_{s=t}^{t-p}\big) = \text{Prob}\big(a_t^\ell = 1 | \{\{\underline{\underline{A}}_s\}_{s=t}^{t-p}\big). \tag{A62}$$

Hence, the UCC model is precisely a DARN(p) model in the limit of $L \to \infty$.

15. MLE of the CDARN(p) model

The CDARN(p) model can be estimated by the maximum likelihood method. First of all, consider the vectorization $\underline{\underline{X}}_t \equiv \{a_t^\ell\}^{\ell=1,\ldots,L}$, with L the number of links on the backbone, of the adjacency matrix $\{a_t^{ij}\}^{(i,j)\in B}$ of the network snapshot at time t. That is, $\{\underline{\underline{X}}_t\}_{t=1,\ldots,T}$ describes the binary random sequences associated with the dynamics of the L links on the backbone. Then the log likelihood of data (by conditioning on the first p observations) under CDARN(p), as defined in Sec. IV, reads as

$$\begin{aligned}\mathbb{L}(q,c,y) &\equiv \log\mathbb{P}(\{\underline{\underline{X}}_t\}_{t=p+1,\ldots,T} | \{\underline{\underline{X}}_s\}_{s=1,\ldots,p}, q, c, y)\\ &= \sum_{t,l}\log\big[q\big[(1-c)D_t^\ell + cC_t^\ell\big]\\ &\qquad + (1-q)y^{X_t^\ell}(1-y)^{1-X_t^\ell}\big],\end{aligned} \tag{A63}$$

where t runs from $p+1$ to T, while ℓ from 1 to L, with

$$D_t^\ell = \sum_{\tau=1}^{p} z_\tau \delta\big(X_t^\ell, X_{t-\tau}^\ell\big), \quad C_t^\ell = \sum_{\ell'\neq\ell}\lambda^{\ell\ell'}\sum_{\tau=1}^{p} z_\tau\delta\big(X_t^\ell, X_{t-\tau}^{\ell'}\big),$$

where $\delta(a,b)$ is the Kronecker delta, taking a value equal to one if $a=b$, zero otherwise, z_τ is the probability of picking τ in the range of integers $(1,\ldots,p)$ (it is $z_\tau = 1/p$ if we assume uniform probability), and $\lambda^{\ell\ell'}$ is the following:

(1) $\lambda^{\ell\ell'} = 0 \ \forall \ell, \ell' = 1, \ldots, L$ with $\ell \neq \ell'$, for the no cross-correlation (NCC) coupling model;

(2) $\lambda^{\ell\ell'} = 1/|\partial_B \ell|$ with $|\partial_B \ell|$ the number of neighbors of link ℓ if $\ell' \in \partial_B \ell$, zero otherwise, for the *local cross-correlation (LCC)* coupling model;

(3) $\lambda^{\ell\ell'} = 1/(L-1)$ with $\ell \neq \ell'$, for the *uniform cross-correlation (UCC)* coupling model.

The MLE of the CDARN(p) model can be then obtained by maximizing the log likelihood in Eq. (A63), or, equivalently, by solving the following system of nonlinear equations:

$$\begin{aligned}
\frac{\partial \mathbb{L}}{\partial y} &= \sum_{t,\ell} \frac{2X_t^\ell - 1}{q\left[(1-c)D_t^\ell + cC_t^\ell\right] + (1-q)y^{X_t^\ell}(1-y)^{1-X_t^\ell}} = 0, \\
\frac{\partial \mathbb{L}}{\partial q} &= \sum_{t,\ell} \frac{((1-c)D_t^\ell + cC_t^\ell) - y^{X_t^\ell}(1-y)^{1-X_t^\ell}}{q\left[(1-c)D_t^\ell + cC_t^\ell\right] + (1-q)y^{X_t^\ell}(1-y)^{1-X_t^\ell}} = 0, \\
\frac{\partial \mathbb{L}}{\partial c} &= \sum_{t,\ell} \frac{C_t^\ell - D_t^\ell}{q\left[(1-c)D_t^\ell + cC_t^\ell\right] + (1-q)y^{X_t^\ell}(1-y)^{1-X_t^\ell}} = 0.
\end{aligned} \tag{A64}$$

The system of nonlinear equations can be solved iteratively by adopting an *iterative proportional fitting procedure*. This consists in solving one by one each equation for each parameter, but conditioning on the values of the other parameters, up to convergence. Such a procedure can be initialized randomly in the parameter space; however, a natural initialization for the parameter y is the average link density of the network. For further details on the method see also [6].

16. MLE of the heterogeneous CDARN(p) model

In the case of link-specific parameters y^ℓ or q^ℓ with $\ell = 1, \ldots, L$, the log likelihood of data under CDARN(p) with heterogenous parameters is generalized quite naturally as

$$\begin{aligned}
\mathbb{L}(q, c, \underline{y}) &\equiv \log \mathbb{P}(\{\underline{\underline{X}}_t\}_{t=p+1,\ldots,T} | \{\underline{\underline{X}}_s\}_{s=1,\ldots,p}, q, c, \underline{y}) \\
&= \sum_{t,\ell} \log \left[q\left[(1-c)D_t^\ell + cC_t^\ell\right] \right. \\
&\qquad \left. + (1-q)(y^\ell)^{X_t^\ell}(1-y^\ell)^{1-X_t^\ell} \right]
\end{aligned} \tag{A65}$$

and

$$\begin{aligned}
\mathbb{L}(\underline{q}, c, y) &\equiv \log \mathbb{P}(\{\underline{\underline{X}}_t\}_{t=p+1,\ldots,T} | \{\underline{\underline{X}}_s\}_{s=1,\ldots,p}, \underline{q}, c, y) \\
&= \sum_{t,\ell} \log \left[q^\ell\left[(1-c)D_t^\ell + cC_t^\ell\right] \right. \\
&\qquad \left. + (1-q^\ell)y^{X_t^\ell}(1-y)^{1-X_t^\ell} \right],
\end{aligned} \tag{A66}$$

respectively, for y^ℓ and q^ℓ. Then, similarly to before, the MLE of the CDARN(p) model with heterogenous parameters is obtained by solving

$$\begin{aligned}
\frac{\partial \mathbb{L}}{\partial y^\ell} &= \sum_t \frac{2X_t^\ell - 1}{q\left[(1-c)D_t^\ell + cC_t^\ell\right] + (1-q)(y^\ell)^{X_t^\ell}(1-y^\ell)^{1-X_t^\ell}} \\
&= 0, \\
\frac{\partial \mathbb{L}}{\partial q} &= \sum_{t,\ell} \frac{\left((1-c)D_t^\ell + cC_t^\ell\right) - (y^\ell)^{X_t^\ell}(1-y^\ell)^{1-X_t^\ell}}{q\left[(1-c)D_t^\ell + cC_t^\ell\right] + (1-q)(y^\ell)^{X_t^\ell}(1-y^\ell)^{1-X_t^\ell}} \\
&= 0,
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathbb{L}}{\partial c} &= \sum_{t,\ell} \frac{C_t^\ell - D_t^\ell}{q\left[(1-c)D_t^\ell + cC_t^\ell\right] + (1-q)(y^\ell)^{X_t^\ell}(1-y^\ell)^{1-X_t^\ell}} \\
&= 0,
\end{aligned} \tag{A67}$$

and

$$\begin{aligned}
\frac{\partial \mathbb{L}}{\partial y} &= \sum_{t,\ell} \frac{2X_t^\ell - 1}{q^\ell\left((1-c)D_t^\ell + cC_t^\ell\right) + (1-q^\ell)y^{X_t^\ell}(1-y)^{1-X_t^\ell}} \\
&= 0, \\
\frac{\partial \mathbb{L}}{\partial q^\ell} &= \sum_t \frac{\left((1-c)D_t^\ell + cC_t^\ell\right) - y^{X_t^\ell}(1-y)^{1-X_t^\ell}}{q^\ell\left((1-c)D_t^\ell + cC_t^\ell\right) + (1-q^\ell)y^{X_t^\ell}(1-y)^{1-X_t^\ell}} \\
&= 0, \\
\frac{\partial \mathbb{L}}{\partial c} &= \sum_{t,\ell} \frac{C_t^\ell - D_t^\ell}{q^\ell\left((1-c)D_t^\ell + cC_t^\ell\right) + (1-q^\ell)y^{X_t^\ell}(1-y)^{1-X_t^\ell}} \\
&= 0,
\end{aligned} \tag{A68}$$

respectively.

17. Link prediction in the CDARN(p) model

Once estimated on data by solving the MLE problem (6), the CDARN(p) model can be used for link prediction: assume that we observe a temporal network up to time t and are asking for the prediction of the network snapshot at time $t+1$ (by using only the information up to time t). The one-step-ahead *forecast* (or *prediction*) of link ℓ is defined as

$$\begin{aligned}
S_{t+1}^\ell &\equiv \mathbb{P}\left(X_{t+1}^\ell = 1 | \{\underline{\underline{X}}_s\}_{s=t,t-1,\ldots,t-p+1}, q, c, y\right) \\
&= q\left[(1-c)\tilde{D}_{t+1}^\ell + c\tilde{C}_{t+1}^\ell\right] + (1-q)y,
\end{aligned} \tag{A69}$$

with

$$\begin{aligned}
\tilde{D}_{t+1}^e ll &= \sum_{\tau=1}^{p} z_\tau \delta\left(1, X_{t-\tau+1}^\ell\right), \\
\tilde{C}_{t+1}^\ell &= \sum_{\ell' \neq \ell} \lambda^{\ell\ell'} \sum_{\tau=1}^{p} z_\tau \delta\left(1, X_{t-\tau+1}^{\ell'}\right).
\end{aligned}$$

The one-step ahead forecast in Eq. (A69) is described by a real value in the unit interval, representing the probability projected at time $t+1$ of observing a link, then the prediction itself, namely, the binary value $\tilde{X}_{t+1}^\ell \in \{0, 1\}$, is obtained according to some threshold value. The time series of forecasts $\{S_t^\ell\}$, together with the realizations $\{X_t^\ell\}$, allow us to characterize the forecasting performance of the model by using some binary classifier. A possibility is constructing the receiving operating characteristic (ROC) curve [61], which is the plot of the true positive rate (TPR) (*sensitivity*) against the false positive rate (FPR) (*specificity*) at various threshold values of the link probability. In particular, the threshold values are selected implicitly by the inputs themselves: by moving from zero to one in the unit interval, each time the sensitivity is increasing or the specificity is decreasing, the corresponding value is considered as a threshold. In practical terms, the better the model performs in the forecasting, the higher the associated ROC curve is in the

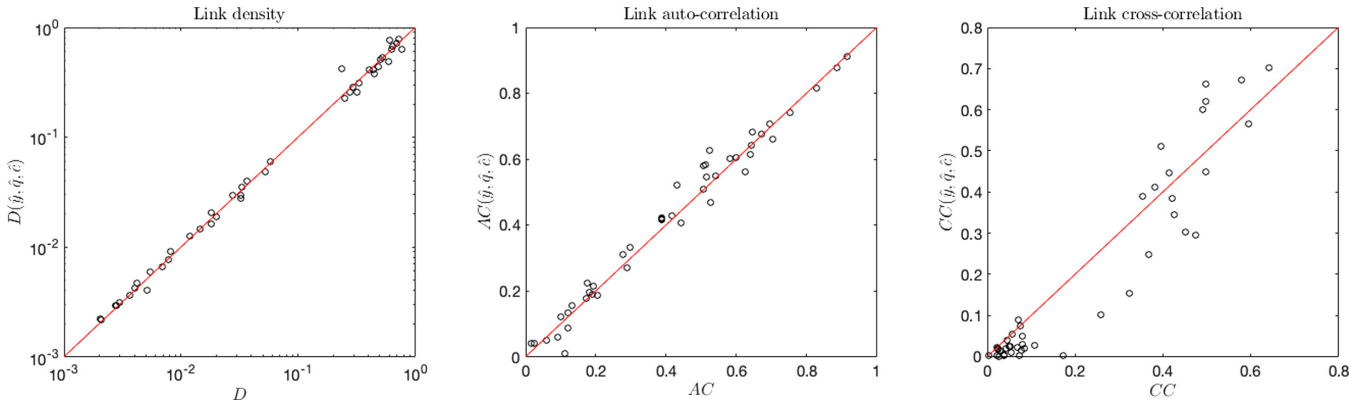

FIG. 11. Scatter plots of the mean link density (left), autocorrelation (middle), and cross-correlation (right) (at lag equal to one) of the empirical temporal networks built with the data sets described in the main text vs the same statistics computed on simulations of the CDARN(1) model (LCC) with constant parameters, which are estimated on the temporal networks themselves.

unit square, or, equivalently, the larger the area under the curve (AUC).

18. Network statistics

Once estimated on data, the CDARN(p) model in its standard version with constant parameters and for a given coupling specification, captures the average link density of a temporal network, together with the average auto- and cross-correlations of links. In particular, it captures the cross-correlations of neighboring links on the backbone, e.g., all links that are incident to the same nodes for the LCC specification. Thus, the observed statistics match (on average) their expectations according to the CDARN(p) model, eventually computed by using simulations. For validation purposes, this is shown in Fig. 11 for all the datasets described in Sec. IV by considering the CDARN(1)-LCC model with constant parameters, which are estimated on data by using maximum likelihood methods described above.

More interestingly, other networks statistics that are not explicitly described by CDARN models, can be computed within the proposed framework, in particular the interevent time between the occurrence of two subsequent links between two nodes at two different times. For instance, it has been observed in temporal networks of human communication that the duration between two contacts is often bursty and deviates from the uniform distribution expected in the case of memoryless processes [12]. This quantity is of great interest, e.g., when studying a spreading process over a temporal network, such as the diffusion process considered in the present work. In our setting, the interevent time can be defined as the number τ of observed time snapshots $X^{\ell}_{t+1}, \ldots, X^{\ell}_{t+\tau}$ for which

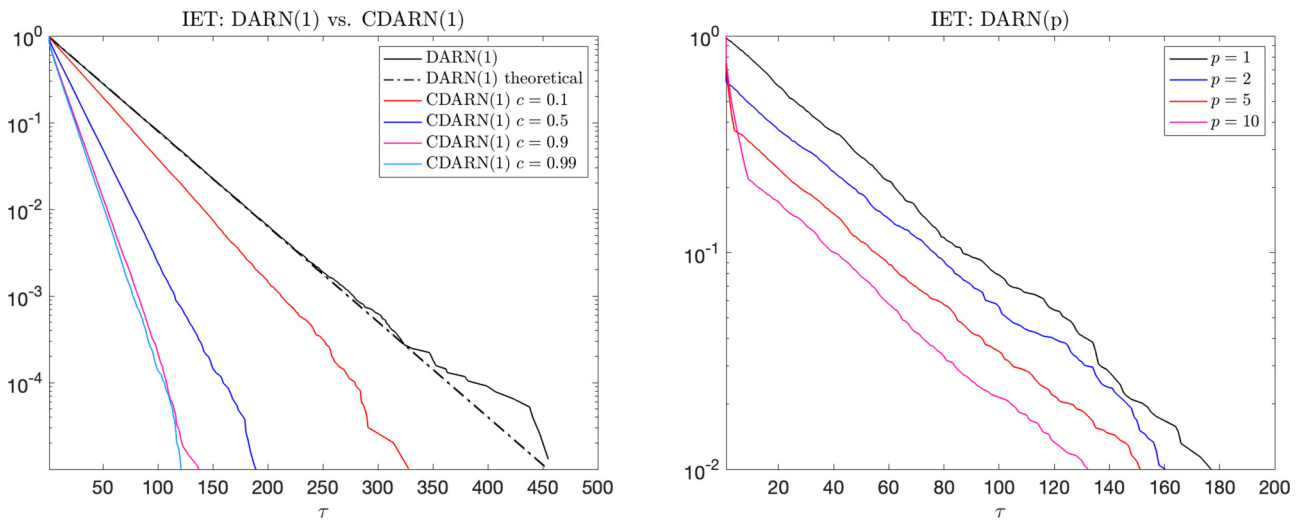

FIG. 12. Comparison of the interevent time (IET) distributions between DARN(1) and CDARN(1) (LCC) models (left), and of the DARN(p) models for different p (right). IET is obtained by using simulations of time series of length $T = 10^5$ of the models, with $y = 0.1$, $q = 0.75$, and c as indicated in the plots (for CDARN).

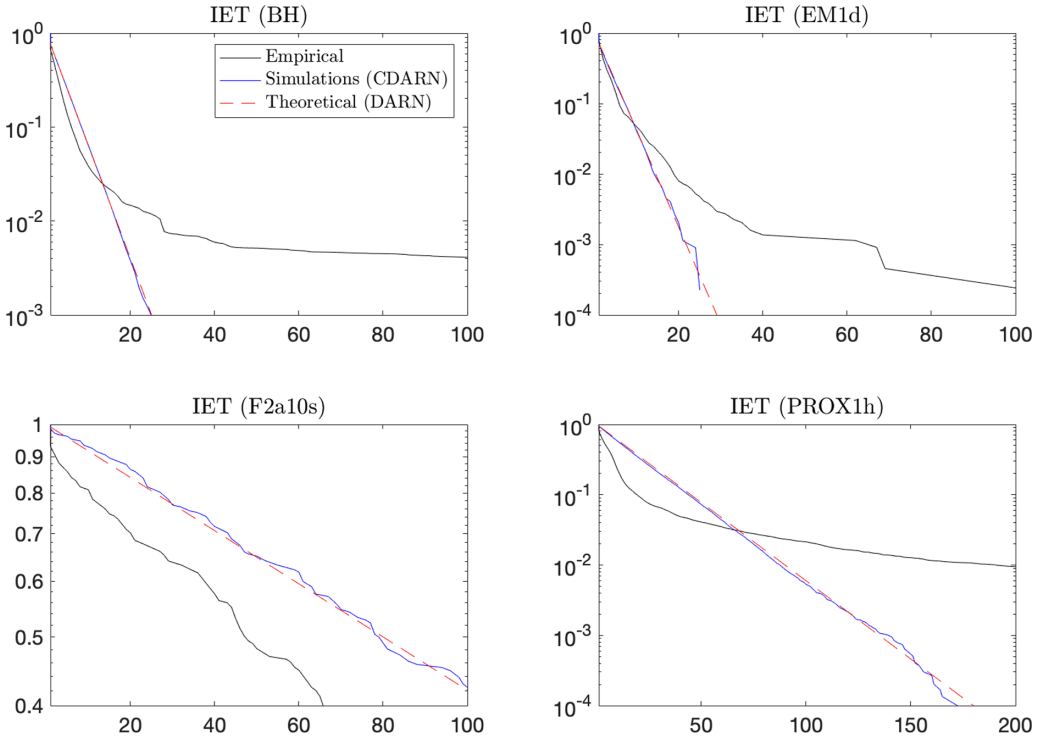

FIG. 13. Interevent time (IET) distribution for four temporal networks, BH, EM1d, F2a10s, and PROX60min, compared with IET distributions for the CDARN(1) model with LCC specification based on numerical simulations, and with the theoretical IET distribution for the DARN(1) model, which as we explain can be computed analytically.

the generic link ℓ is zero, after an observation $X_t^\ell = 1$. The distribution of τ conditional to the observation $X_t^\ell = 1$ for the CDARN(1) model (with constant parameters) is equivalent to compute the following joint probability:

$$\begin{aligned} &p\big(\tau|X_t^\ell = 1\big) \\ &\quad\equiv P\big(X_{t+1}^\ell = 0,\ldots,X_{t+\tau}^\ell = 0\big|X_t^\ell = 1\big) \\ &\quad= P\big(X_{t+1}^\ell = 0\big|X_t^\ell = 1\big)\prod_{i=1}^{\tau-1} P\big(X_{t+1+i}^\ell = 0\big|X_{t+i}^\ell = 0\big) \\ &\quad= \left[qc\sum_{\ell'\neq\ell}\lambda^{\ell\ell'}\delta\big(0,X_t^{\ell'}\big)+\beta\right] \\ &\qquad\times\prod_{i=1}^{\tau-1}\left\{q\left[(1-c)+c\sum_{\ell'\neq\ell}\lambda^{\ell\ell'}\delta\big(0,X_{t+i}^{\ell'}\big)\right]+\beta\right\}, \end{aligned} \tag{A70}$$

with $\beta = (1-q)(1-y)$.

In general, a closed form solution cannot be obtained because of nondiagonal interaction terms mediated by the couplings matrix λ. Except for the NCC specification of the model, i.e., the DARN(1) model, which sets to zero the cross-correlations, i.e., $c = 0$. In this case, the probability distribution of the interevent time can be easily derived and is equal to

$$p(\tau) = e^{-\frac{\tau-1}{\sigma}}, \quad \tau = 1,2,\ldots, \tag{A71}$$

with $\sigma = (\log\frac{1}{q+\beta})^{-1}$.

Notice that the probability distribution (A71) of the DARN(1) model represents an upper bound for the CDARN(1) model: when $c > 0$ in Eq. (A70), there exists always a probability larger than zero of copying one past neighboring link (whatever the coupling matrix λ), instead of copying the past itself, i.e., a zero, with probability one. This reduces the probability of observing a number τ of successive zeros, thus resulting in an (approximate) exponential distribution of interevent times with a timescale smaller than σ, see the left panel of Fig. 12. Finally, the non-Markovian case $p > 1$ is not analytically tractable as long as p increases further and further. However, for the NCC specification of the CDARN(p) model, i.e., DARN(p), it is easy to gather that $p(\tau) = O[\exp(-(\tau-p)/\sigma)]$ when $\tau \gg p$, by following similar computations leading to Eq. (A70). This is confirmed numerically in the right panel of Fig. 12. We can conclude that the interevent time distribution of the CDARN(p) model is an *approximate* exponential with a timescale equal or smaller than σ.

Finally, in Fig. 13 we show the empirical distribution of the interevent time of four real-world temporal networks, compared with the corresponding distribution for the CDARN(1) model with LCC coupling specification, obtained numerically by means of simulations, and with the theoretical one (A71) for the DARN(1) model. In both cases, the parameters of the model has been obtained by maximum likelihood estimation. In all cases, the CDARN(1) can be seen as an approximation of the empirical distributions of the interevent time for small τ, while the fatter tails (excluding the football network) are not captured by the model.

[1] M. C. González, C. A. Hidalgo, and A.-L. Barabási, Nature (London) **453**, 779 (2008).
[2] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, Phys. Rev. Lett. **110**, 168701 (2013).
[3] E. Yoneki, D. Greenfield, and J. Crowcroft, in *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining, Athens, Greece* (IEEE, Piscataway, NJ, 2009), pp. 356–361.
[4] R. Murcio, A. P. Masucci, E. Arcaute, and M. Batty, Phys. Rev. E **92**, 062130 (2015).
[5] D. Li, B. Fu, Y. Wang, G. Lu, Y. Berezin, H. E. Stanley, and S. Havlin, Proc. Natl. Acad. Sci. USA **112**, 669 (2015).
[6] P. Mazzarisi, P. Barucca, F. Lillo, and D. Tantari, Eur. J. Oper. Res. **281**, 50 (2020).
[7] M. Valencia, J. Martinerie, S. Dupont, and M. Chavez, Phys. Rev. E **77**, 050905(R) (2008).
[8] F. D. V. Fallani, V. Latora, L. Astolfi, F. Cincotti, D. Mattia, M. G. Marciani, S. Salinari, A. Colosimo, and F. Babiloni, J. Phys. A: Math. Theor. **41**, 224014 (2008).
[9] A. P. Millán, J. Torres, S. Johnson, and J. Marro, Nat. Commun. **9**, 2236 (2018).
[10] D. R. Chialvo, Nat. Phys. **6**, 744 (2010).
[11] P. Grindrod and D. J. Higham, Proc. R. Soc. London A **466**, 753 (2010).
[12] P. Holme and J. Saramäki, Phys. Rep. **519**, 97 (2012).
[13] L. R. Lambiotte and M. N. Masuda, *A Guide to Temporal Networks* (World Scientific, Singapore, 2021).
[14] L. Gauvin, A. Panisson, and C. Cattuto, PLoS ONE **9**, e86028 (2014).
[15] M. Zanin, L. Lacasa, and M. Cea, Chaos **19**, 023111 (2009).
[16] V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, and V. Latora, Chaos **22**, 023101 (2012).
[17] T. Weng, J. Zhang, M. Small, R. Zheng, and P. Hui, Sci. Rep. **7**, 41951 (2017).
[18] T. P. Peixoto and M. Rosvall, Nat. Commun. **8**, 582 (2017).
[19] A. Buscarino, L. Fortuna, M. Frasca, and V. Latora, Europhys. Lett. **82**, 38002 (2008).
[20] M. Starnini and R. Pastor-Satorras, Phys. Rev. E **89**, 032807 (2014).
[21] M. Karsai, N. Perra, and A. Vespignani, Sci. Rep. **4**, 4001 (2014).
[22] L. Alessandretti, K. Sun, A. Baronchelli, and N. Perra, Phys. Rev. E **95**, 052318 (2017).
[23] P. Singer, D. Helic, B. Taraghi, and M. Strohmaier, PLoS ONE **9**, e114952 (2014).
[24] M. Szell, R. Sinatra, G. Petri, S. Thurner, and V. Latora, Sci. Rep. **2**, 457 (2012).
[25] R. Lambiotte, M. Rosvall, and I. Scholtes, Nat. Phys. **15**, 313 (2019).
[26] A. Moinet, A. Barrat, and R. Pastor-Satorras, Phys. Rev. E **98**, 022303 (2018).
[27] Y. Zhang, A. Garas, and I. Scholtes, J. Phys. Complex. **2**, 015007 (2021).
[28] V. Salnikov, M. T. Schaub, and R. Lambiotte, Sci. Rep. **6**, 23194 (2016).
[29] J. T. Matamalas, M. De Domenico, and A. Arenas, J. R. Soc. Interface **13**, 20160203 (2016).
[30] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte, Nat. Commun. **5**, 4630 (2014).
[31] T. Hiraoka and H.-H. Jo, Sci. Rep. **8**, 15321 (2018).
[32] A. Sapienza, A. Barrat, C. Cattuto, and L. Gauvin, Phys. Rev. E **98**, 012317 (2018).
[33] I. Z. Kiss, G. Röst, and Z. Vizi, Phys. Rev. Lett. **115**, 078701 (2015).
[34] T. P. Peixoto and L. Gauvin, Sci. Rep. **8**, 15511 (2018).
[35] O. E. Williams, F. Lillo, and V. Latora, New J. Phys. **21**, 043028 (2019).
[36] A. Moinet, R. Pastor-Satorras, and A. Barrat, Phys. Rev. E **97**, 012313 (2018).
[37] N. Masuda, K. Klemm, and V. M. Eguíluz, Phys. Rev. Lett. **111**, 188701 (2013).
[38] J.-C. Delvenne, R. Lambiotte, and L. E. Rocha, Nat. Commun. **6**, 7366 (2015).
[39] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone, and F. Schweitzer, Nat. Commun. **5**, 5024 (2014).
[40] X.-X. Zhan, A. Hanjalic, and H. Wang, Sci. Rep. **9**, 6798 (2019).
[41] H.-H. Jo, J. I. Perotti, K. Kaski, and J. Kertész, Phys. Rev. E **92**, 022814 (2015).
[42] R. Burioni, E. Ubaldi, and A. Vezzani, J. Stat. Mech. (2017) 054001.
[43] H. Kim, M. Ha, and H. Jeong, Eur. Phys. J. B **88**, 315 (2015).
[44] N. Georgiou, I. Z. Kiss, and E. Scalas, Phys. Rev. E **92**, 042801 (2015).
[45] I. Scholtes, in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM Press, New York, 2017), pp. 1037–1046.
[46] R. Lambiotte, V. Salnikov, and M. Rosvall, J Complex Netw. **3**, 177 (2015).
[47] C. L. Vestergaard, M. Génois, and A. Barrat, Phys. Rev. E **90**, 042805 (2014).
[48] E. R. Colman and D. Vukadinović Greetham, Phys. Rev. E **92**, 012817 (2015).
[49] P. Van Mieghem and R. van de Bovenkamp, Phys. Rev. Lett. **110**, 108701 (2013).
[50] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Sci. Rep. **2**, 469 (2012).
[51] P. D. Hoff, A. E. Raftery, and M. S. Handcock, J. Am. Stat. Assoc. **97**, 1090 (2002).
[52] P. Sarkar and A. W. Moore, ACM SIGKDD Explor. Newsl. **7**, 31 (2005).
[53] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, Soc. Netw. **47**, 130 (2016).
[54] S. Hanneke, W. Fu, and E. P. Xing, Electron. J. Statist. **4**, 585 (2010).
[55] P. A. Jacobs and P. A. Lewis, Discrete Time Series Generated by Mixtures. III. Autoregressive Processes (DAR (p)), Tech. Rep., Naval Postgraduate School, Monterey, CA (1978).
[56] C. Campajola, F. Lillo, P. Mazzarisi, and D. Tantari, J. Stat. Mech. (2021) 033412.
[57] P. A. Jacobs and P. A. Lewis, J. R. Stat. Soc. B **40**, 94 (1978).
[58] P. A. Jacobs and P. A. Lewis, J. R. Stat. Soc. B **40**, 222 (1978).
[59] P. A. Jacobs and P. A. Lewis, J. Time Series Anal. **4**, 19 (1983).
[60] P. Mazzarisi, S. Zaoli, C. Campajola, and F. Lillo, J. Econ. Dyn. Control **121**, 104022 (2020).
[61] T. Hastie, R. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning Data Mining, Inference, and Prediction* (Springer, New York, 2009).

[62] W. Min and L. Wynter, Transp. Res. C **19**, 606 (2011).
[63] J. D. Hamilton, *Time Series Analysis* (Princeton University Press, Princeton, 2020), p. 625.
[64] J. Runge, V. Petoukhov, J. F. Donges, J. Hlinka, N. Jajcay, M. Vejmelka, D. Hartman, N. Marwan, M. Paluš, and J. Kurths, Nat. Commun. **6**, 8502 (2015).
[65] B. Bollobás, in *Random Graphs* (Cambridge University Press, Cambridge, 2001), p. 73.
[66] O. E. Williams, L. Lacasa, A. P. Millán, and V. Latora, Nat. Commun. **13**, 499 (2022).
[67] J. Fan, M. Farmen, and I. Gijbels, J. R. Stat. Soc. B **60**, 591 (1998).
[68] Here the number of nodes is corresponding to the number of stations, or stops, characterizing the specific public transport at each city.
[69] D. J. Watts and P. S. Dodds, J. Consum. Res. **34**, 441 (2007).
[70] F. Schweitzer and J. D. Farmer, *Brownian Agents and Active Particles Collective Dynamics in the Natural and Social Sciences* (Springer, Berlin, 2003).
[71] T. Di Matteo, T. Aste, and S. Hyde, arXiv:cond-mat/0310544.
[72] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Phys. Rep. **469**, 93 (2008).
[73] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[74] M. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
[75] J. Kunegis, The KONECT project, US airports network dataset (2016), http://konect.cc/.
[76] R. Michalski, S. Palus, and P. Kazienko, in *Proceedings of the 14th International Conference on Business Information Systems*, Lecture Notes in Business Information Processing LNBIP Vol. 87 (Springer, Berlin, Heidelberg, 2011), pp. 197–206.
[77] *TfL Rolling Origin and Destination Survey*, London Datastore (2014), https://data.london.gov.uk/dataset/tfl-rolling-origin-and-destination-survey.
[78] E. Cinlar, *Introduction to Stochastic Processes* (Dover Publications, New York, 2013).
[79] C. Ballester and M. Vorsatz, Rev. Econ. Stat. **96**, 383 (2014).
[80] I. L. MacDonald and W. Zucchini, *Hidden Markov and Other Models for Discrete-Valued Time Series* (CRC Press, Boca Raton, Florida, 1997), Vol. 110.
[81] L. Gauvin, M. Génois, M. Karsai, M. Kivelä, T. Takaguchi, E. Valdano, and C. L. Vestergaard, arXiv:1806.04032.
[82] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, Phys. Rev. E **83**, 025102(R) (2011).
[83] V.-P. Backlund, J. Saramäki, and R. K. Pan, Phys. Rev. E **89**, 062815 (2014).
[84] https://pypi.org/project/networksns/.
[85] https://pypi.org/project/networksns/docs.
[86] http://www.sobigdata.eu.