EPJ Data Science

https://doi.org/10.1140/epjds/s13688-026-00663-0

Article in Press

Modeling memory in time-respecting paths on temporal networks

Received: 21 November 2025

Accepted: 3 May 2026

Published online: 12 May 2026

Cite this article as: Guerrini, S., Cattuto, C., Dall'Amico, L. Modeling memory in time-respecting paths on temporal networks. *EPJ Data Science* (2026). https://doi.org/10.1140/epjds/s13688-026-00663-0

Silvia Guerrini, Ciro Cattuto, Lorenzo Dall'Amico

Modeling memory in time-respecting paths on temporal networks

Silvia Guerrini, Ciro Cattuto, Lorenzo Dall'Amico*

ISI Foundation, Via Chisola, 5, Turin, 10126, Italy.

*Corresponding author(s). E-mail(s): lorenzo.dallamico@isi.it;
Contributing authors: silvia.guerrini@isi.it; ciro.cattuto@isi.it;

Abstract

Human close-range proximity interactions are the key determinant for spreading processes like knowledge diffusion, norm adoption, and infectious disease transmission. These dynamical processes can be modeled with time-respecting paths on temporal networks. Here, we propose a framework to quantify memory in time-respecting paths and evaluate it on several empirical datasets encoding proximity between humans collected in different settings. Our results show strong memory effects, robust across settings, model parameters, and statistically significant when compared to memoryless null models. We further propose a generative model to create synthetic temporal graphs with memory and use it to show that memory in time-respecting paths decreases the diffusion speed, affecting the dynamics of spreading processes on temporal networks.

Keywords: Temporal networks, Human proximity, Memory, Time respecting paths, Generative models

1 Introduction

Human encounters at short distances are a key driver of social interactions between individuals [1]. They enable diffusive processes such as knowledge diffusion, norms adoption, and infectious disease propagation [2]. The unfolding of these diffusive processes depends on the evolution of proximity interactions over time, and calls for appropriate modeling using temporal networks [3]. In these networks, individuals are modeled as *nodes* and their interactions are represented by *temporal edges* formed by node pairs and a time index indicating when the interaction occurred. Temporal interaction networks of humans display recurrent properties, observed in all collection

contexts, such as the bursty behavior and high heterogeneity of temporal patterns [4–8]. The ubiquity of these observations has attracted the efforts of several researchers who proposed generative models of temporal networks capable of reproducing the observed properties [9–15].

Here, we focus on describing and modeling long memory patterns, another commonly observed feature of empirical social networks [16, 17]. While in the remainder we will only consider memory in diadic interactions, *i.e.,* between pairs of nodes, it is worth mentioning that memory effects have also been recently observed in higher-order structures composed of groups of nodes [18]. Memory in temporal networks influences epidemic spreading [19, 20], diffusion speed [21, 22], and the evolution of the network itself [23]. Recent works have also shown that accounting for memory leads to improved performance in machine learning tasks such as change-point detection [24], identification of causal temporal timescales [25], and community detection [26, 27]. As such, memory is a key ingredient of temporal network modeling.

Memory in temporal graphs is not uniquely defined, and different approaches have been explored to quantify it. Some works considered self-reinforcement mechanisms in which the existence of a temporal edge is influenced by its past [28], or by the past of other edges [29]. Using a self-reinforcement mechanism, Ref. [16] showed a relation between memory and the observed high heterogeneity of temporal patterns in human proximity networks. Another approach to modeling memory relies on the identification of significant temporal motifs, *i.e.,* of temporal patterns among nodes [30, 31]. Similarly to Refs. [25–27], we define memory by relying on *time respecting paths*, which describe the navigability of temporal networks, *i.e.,* how information flows from one node to another. Time respecting paths drive the unfolding of dynamical processes [32] and have been recently used to define a notion of distance between temporal graphs [33]. We quantify memory as the likelihood that a time-respecting path returns to a previously visited node. Our definition of memory allows us to detect memory effects in Markovian processes, such as diffusion, induced by a non-Markovian evolution of the graph topology. We propose two models for time respecting paths: one that only includes memory, the other that also accounts for the presence of a community structure. Long memory effects lead to increased model complexity to account for temporal correlations, and some works [24, 27] reduce the complexity by assuming memory intervenes only at the mesoscale, *i.e.,* on groups of nodes forming communities. Unlike these approaches, both our models have a single parameter encoding memory and do not need the mesoscale assumption for a reduced model complexity.

We evaluate our models on eight empirical temporal networks describing human proximity interactions [5, 6, 34–38], collected in various contexts by the `SocioPatterns` collaboration ([4], sociopatterns.org). The results show significant memory effects across all networks, as demonstrated by the comparison with null models. Interestingly, we observe comparable memory effects in similar contexts (like schools and workplaces), suggesting a robust behavior of memory across data collections. We further develop a generative model to create synthetic graphs with memory, and demonstrate that increased memory decreases the diffusion speed. Our results

enable the numerical quantification of memory, which is, together with the network density and the community structure, a key property in temporal graph modeling.

2 Results

Temporal graphs model complex systems with time-dependent interactions [39]. They are formed by a set of temporal edges $\mathcal{E}$ with elements (i, j, t, w) that denote interactions between node pairs $i, j \in \mathcal{V}$ at time t with weight w. This definition accounts time as a discrete variable with resolution t_{res}. All interactions within t_{res} are considered simultaneous, and we account for the frequency of interaction by letting the weight w be the total interaction time between two nodes in the time resolution interval.

We model non-backtracking time-respecting paths (TRPs) on temporal graphs. In the static setting, a path $\mathcal{P} = \{v_1, v_2, \ldots, v_a\}$ is an ordered sequence of nodes that represents the trajectory of a walker on the graph. A *transition* of the walker from one node v_x to the next, v_{x+1}, is possible only if (v_x, v_{x+1}) is a graph edge. Non-backtracking paths exclude self-loops ($v_{x+1} \neq v_x$) and backtracks ($v_{x+1} \neq v_{x-1}$). TRPs extend this concept by incorporating time and respecting the chronological order of interactions. In a TRP, transitions between nodes at time t are possible only if there is a connection between those nodes at time t. Unlike paths on static graphs, they are non-symmetric: a path from i to j does not imply the existence of a path from j to i. The size of a TRP can be defined in two alternative ways: one is the *path length* that equals the number of nodes in the path; the other is the *duration*, *i.e.,* the time elapsed between the last and first steps in the path [32]. Both definitions are relevant as they describe properties that directly influence the unfolding of dynamical processing on temporal networks, and crucially determine their outcome [40]. For a more detailed definition of TRPs and their implementation, we refer the reader to Section 4.1. We remark that, according to our definition, TRPs preserve the ordering in which path transitions occur and do not constrain the actual time elapsed between consecutive steps.

We model the likelihood that a TRP returns to a previously visited node within a *memory horizon*. We observe that TRPs on time-resolved human proximity data display statistically significant memory effects and return to already visited nodes with high likelihood, even if TRPs' steps do not depend on their history. This is possible because TRPs are constrained on the graph's temporal edges. This effect weakens by increasing the aggregation values t_{res}, which impose milder constraints on the path dynamics because the number of edges at each time step can only increase with t_{res}. In the Appendix, we evaluate the robustness of our results across different resolution values t_{res} and we expand our analysis to a memory definition based on the TRP duration instead of length, which also provides statistically significant results.

2.1 Modeling memory

To get our definition of memory, we start from *non-backtracking* TRPs, in which a walker cannot move to the node it came from, *i.e.,* $v_{x+1} \neq v_{x-1}$ for all x. This choice allows us to focus on long-range memory effect, avoiding short-term back-and-forth

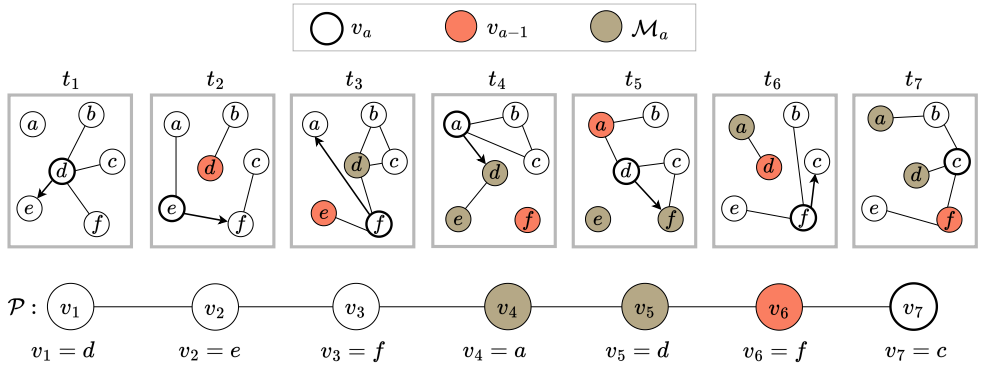

Fig. 1 Memory in time-respecting paths. *Top row.* A temporal graph with $n = 6$ nodes. We depict a non-backtracking time-respecting path $\mathcal{P}$ on this graph. The walker is located in the white dot marked with the thick line. An arrow points from this node to the node occupied at the following time step. The red dots indicates the node the path comes from. This node cannot be reached by a path starting from the node marked with the thick line, as it would create a backtrack. The brown dots denote the memory set $\mathcal{M}_a$ at each time step for a memory horizon $m = 4$. Note that the size of the memory set is not constant. *Bottom row.* The path $\mathcal{P}$ at time t_7 with the colors and markers obtained at this time.

transitions between node pairs that may generate fictitious memory effects in the presence of bursty dynamics [41]. Considering a path $\mathcal{P} = \{v_1, v_2, \ldots, v_a\}$, we let its *memory* $\mathcal{M}_a$ be the set of nodes appearing in the last m path positions, excluding v_a and v_{a-1}. Formally, given $\mathcal{P} = \{v_1, v_2, \ldots, v_a\}$, and m, we let $\mathcal{M}_a(\mathcal{P}, a) = \{v_x \in \mathcal{P} \ : \ 2 \leq a - x < m\} = \{v_{a-m+1}, v_{a-m+2}, \ldots, v_{a-2}\}$. Figure 1 provides a schematic representation of this definition. The parameter m sets the memory horizon in terms of the path length. In the Appendix, we explore a definition of $\mathcal{M}_a$ in which the memory horizon is expressed in time units.

We propose two models for non-backtracking TRPs. The first (**MEM**) only accounts for memory and, given $p \in [0, 1]$ and $n = |\mathcal{V}|$, the model chooses the next node of $\mathcal{P}$ according to the following rule for $v_{a+1} \notin \{v_a, v_{a-1}\}$:

$$P\left(v_{a+1}|\mathcal{M}_a\right) = p \cdot \frac{\delta\left(v_{a+1} \in \mathcal{M}_a\right)}{|\mathcal{M}_a|} + \frac{1-p}{n-2} \ , \tag{1}$$

and $P(v_{a+1}|\mathcal{M}_a) = 0$ if $v_{a+1} \in \{v_a, v_{a-1}\}$. According to this model, with a probability p, a node is chosen among those in $\mathcal{M}_a$, and, with probability $1 - p$, is randomly sampled from all nodes that can form a non-backtracking path, including those in $\mathcal{M}_a$.

Several social networks display a community structure that impacts the statistical properties of TRPs [33, 41, 42]. We thus propose a generalization of Eq. (1) to account for this effect inspired by the stochastic block model [43] (**MEM + SBM**). The stochastic block model generates random graphs with a community structure, by assigning a larger probability of connection between nodes in the same community. Let $\ell : \mathcal{V} \to \{1, \ldots, k\}$ be a known labeling function that assigns each node to one of k communities. We let $C \in \mathbb{R}^{k \times k}$ be a symmetric non-negative matrix whose entry

Name	Description	n	k	Duration
Primary [5]	Students and teachers from a primary school	242	11	2 days
High School 1 [6]	Students and teachers from a high school	126	4	4 days
High School 2 [6]	Students from a high school	180	5	7 days
High School 3 [36]	Students from a high school	327	9	5 days
Conference [34]	Participants to a scientific conference	405	/	2 days
Office [34]	Workers in an office	232	12	10 days
Hospital [37]	Patients and health-care workers in a hospital	75	4	5 days
Malawi [38]	Individuals in a village in rural Malawi	86	/	26 days

Table 1 Summary description of the `Sociopatterns` social networks. The column "Name" reports the name used in the text to refer the dataset. "Description" provides concise information on the context. The subsequent columns indicate the number of nodes (n), the number of communities (k), and the temporal span of the dataset.

$C_{\alpha,\beta}$ describes the affinity between the communities α and β. Then, according to the stochastic block model, an edge between nodes u and v exists with a probability proportional to $L(u,v) := C_{\ell_u,\ell_v}$. Similarly, the **MEM+SBM** selects the next node of $\mathcal{P}$ according to the following rule for $v_{a+1} \neq v_a, v_{a-1}$:

$$P\left(v_{a+1}|\mathcal{M}_a\right) = p \cdot \frac{\delta\left(v_{a+1} \in \mathcal{M}_a\right)}{|\mathcal{M}_a|} + (1-p) \cdot \frac{L(v_a, v_{a+1})}{Z_a} \ , \tag{2}$$

where $Z_a = \sum_{u \in \mathcal{V} \setminus \{v_a, v_{a-1}\}} L(v_a, u)$ is the normalization and $P(v_{a+1}|\mathcal{M}_a) = 0$ if $v_{a+1} \in \{v_a, v_{a-1}\}$. We remark that for both Eqs. (1, 2), $\sum_{u \in \mathcal{V}} P(u|\mathcal{M}_a) = 1$. Eq. (1) is a particular case of Eq. (2), obtained for $L_{u,v} = 1$ for all u, v, *i.e.,* in the absence of a community assignment.

2.2 Inference on empirical data

We evaluate our models on eight empirical temporal networks collected by the `SocioPatterns` collaboration ([4], sociopatterns.org) describing time-resolved proximity interactions between humans, with a temporal resolution $t_{\text{res}} = 20\ s$. These datasets were collected in schools [5, 6, 36], an office [34], a hospital [37], a scientific conference [34], and a rural village in Malawi [38]. Some datasets provide additional node categorical attributes: in the school datasets, each node is a student in a known school class; in the hospital and office datasets, nodes have a role attribute. We provide a summary description of these datasets in Table 1.

We first compare the two models to assess whether the higher complexity of the **MEM + SBM** model leads to a better goodness-of-fit of the empirical data. We then proceed by describing the maximum likelihood estimator of the memory parameter p for different values of the memory horizon m. To assess the significance of these results, we compare the inferred memory p with the one obtained on temporal graphs generated from memoryless null models.

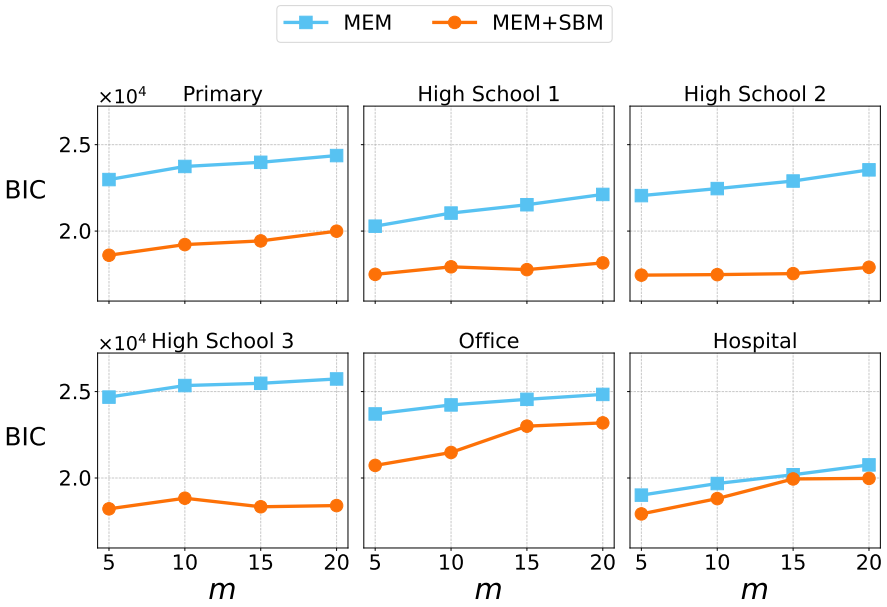

Fig. 2 Comparison of the goodness-of-fit between MEM and MEM + SBM models as a function of the memory horizon m**.** Each plot refers to one of the six datasets of Table 1 with a known node label assignment. These dataset have a known node partition in communities. In *Primary*, *High school 1, 2, 3*, each node is assigned to a school class. In the *Office* and in the *Hospital* datasets labels are a role attribute. We show the BIC (lower is better) of the **MEM** model (cyan squares) and of the **MEM + SBM** model (orange circles) as a function of the memory length m for $t_{\text{res}} = 20$ s. While the plots share the same y-axis, the results cannot be compared across datasets.

2.2.1 Goodness-of-fit

For each temporal graph in Table 1, we run a collection of non-backtracking TRPs, introduced in Section 2. We then use a maximum likelihood estimator to infer the probability of the TRP returning to an already visited node and the community affinity matrix. Section 4.1 provides the details on the TRP implementation, and Section 4.2 the expression of the maximum likelihood estimators for the **MEM** and **MEM + SBM** models. We evaluate the goodness-of-fit using the Bayesian information criterion (BIC) [44]. Small BIC values imply a better fit.

For the six datasets of Table 1 with known node attributes, we compare the BIC for both **MEM** and **MEM + SBM** models to evaluate whether the more complex model is required to explain the TRPs statistics. Figure 2 shows the BIC values for the two models as a function of the memory size, m. In all datasets, the **MEM + SBM** model achieves better results than the **MEM** model, showing that the community structure needs to be accounted for to explain the TRPs statistic. This is especially evident in the four school datasets that display a highly assortative community structure. Similar results are obtained defining the *memory horizon* in terms of time duration, as reported in the Appendix.

2.2.2 Inference of the memory parameter

Figure 3 shows the inferred memory parameter p as a function of the memory size m across datasets. For all datasets with known node labels, we observe the **MEM** model provides larger memory values p than those observed in the **MEM+SBM** model. Accounting for the community structure is thus necessary to disentangle homophily from memory. The largest mismatch between the models is observed in the school datasets, coherently with the results shown in Figure 2.

Comparing the results across datasets, we observe similar inferred values of p for all schools and in particular, all high schools. Also, workplaces – *Office* and *Hospital* – lead to comparable results, with values of p slightly larger than those observed in schools. The high memory observed in *Malawi* and the low one in *Conference* are understood from the experimental context. *Malawi* describes interactions between family members in an African rural village. These interactions are known to be frequent and prolonged [45], thus explaining the high memory effects observed. *Conference* describes the interactions among researchers at a scientific conference. This can be explained by several characteristics of scientific conferences, like the presence of large gatherings, the schedule of the conference, or the fact that attendees commonly browse around to meet new people.

To assess the statistical significance of our results, we compare them with two null models [46] that do not have memory by construction. In the first null model, we obtain a temporal graph by generating a sequence of snapshots drawn from the Erdős-Rényi model in which each snapshot has the same density as the corresponding one in the empirical dataset. We obtain p as the maximum likelihood estimator of the **MEM** model on TRPs obtained from this temporal graph. This temporal graph does not have correlations between its snapshots, and any inferred memory effect can be ascribed to randomness. In the second model, we independently generate graphs with a community structure from the *stochastic block model* and estimate p from the **MEM + SBM** model. The community structure introduces a correlation between the snapshots due to homophily and not memory. Also in this case, each snapshot has the same graph density as the corresponding empirical dataset. Coherently with the results discussed above, we observe that in the school datasets, the inferred p on the null model with communities is non-zero, which is caused by a highly homophilic community structure. On the contrary, the p inferred from the Erdős Rényi model is null in all cases. On all datasets, the inferred memory values are much larger than the values inferred from the null models, evidencing a non-trivial memory effect in empirical pathway data.

2.3 Memory in the TRPs decreases the diffusion speed

Several works have shown that memory affects the speed of diffusion on temporal networks [19, 22, 29, 47, 48]. We investigate how memory in time-respecting paths (TRPs) impacts the diffusion speed. To do so, we first introduce a model to generate synthetic graphs with varying memory. This allows us to control memory and span different regimes in a controlled setting. We introduce two parameters $\hat{m} \in \mathbb{N}$, and $\alpha \in [0, 1]$, describing memory. The parameter $\hat{m}$, similarly to m in the previous sections,

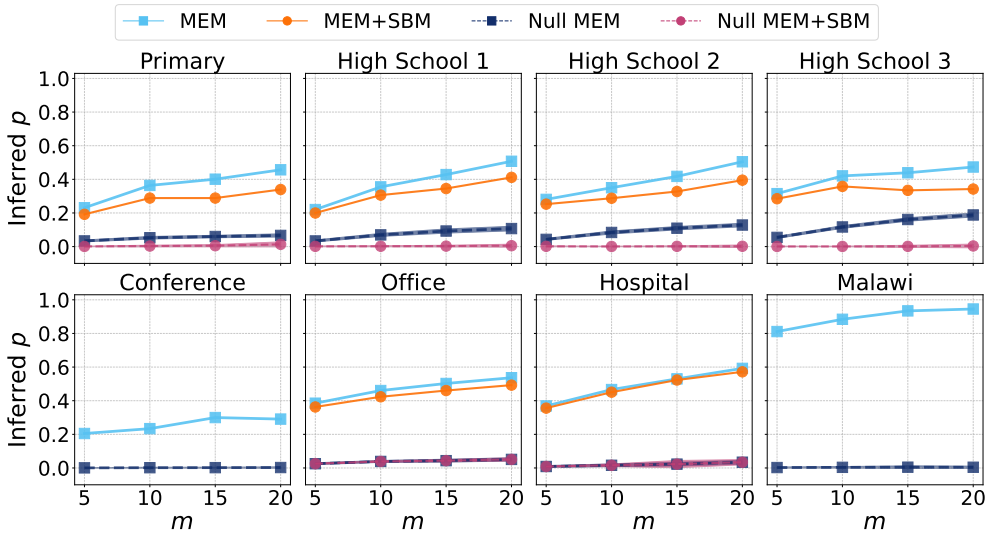

Fig. 3 Maximum likelihood estimates of the memory parameter p and comparison with the null models. Each plot refers to one of the datasets described in Table 1 and shows the inferred value of the probability p of the time-respecting paths to return to a node, as a function of the memory length m. The curve "MEM" (solid line with squares) is obtained from the **MEM** model of Eq. (1) on the empirical data. The curve "MEM+SBM" (solid line with dots) is obtained from the **MEM+SBM** model of Eq. (2) on the empirical data. The curve "Null MEM" (dashed line with squares) is obtained from the **MEM** model (Eq. (1)) on the synthetic data, generated from the Erdős Rényi null model. The curve "Null MEM + SBM" (dashed line with dots) is obtained from the **MEM+SBM** model Eq. (2) on the synthetic data, generated from the SBM null model. For the *Conference* and *Malawi* datasets, we only consider the null model based on Erdős-Rényi random graphs, while for all other datasets, we consider the one generated from stochastic block model graphs. For these datasets, the labels are known attributes that indicate school classes (*Primary*, *High school 1, 2, 3*), and role attributes (*Office*, *Hospital*). The curves referring to the null models show the mean over 50 realizations, and the shaded areas represent the standard deviation. For all graphs, we consider the temporal resolution $t_{\text{res}} = 20$ *s*.

is a memory horizon used to introduce correlation between the last $\hat{m}$ snapshot. The parameter α weights the correlation term: by design, we expect the memory p to be an increasing function of α. By letting D_t, A_t be the degree and adjacency matrices at time t respectively, we introduce $L_t = (D_t + I_n)^{-1}(A_t + I_n)$, where I_n denotes the identity matrix of size n. The entries of L_t indicate the probability of moving between two nodes at time t, including self-edges for every node. Consequently, the entries of $M(t;\hat{m}) = \prod_{t'=t-\hat{m}}^{t-1} L_{t'}$ express the likelihood of two nodes to be connected by a TRP in the last $\hat{m}$ steps.

We generate the first graph snapshot from the Erdős-Rényi model and then, for all $t > 1$ and $i \neq j$, the (i, j) entry of the adjacency matrix A_t are set to one independently at random with probability

$$\mathbb{P}\Big((A_t)_{i,j} = 1\Big) = d\left[\frac{1-\alpha}{n} + \frac{\alpha}{Z}\Big(M_{ij}(t;\hat{m}) + M_{ji}(t;\hat{m})\Big)\right] , \tag{3}$$

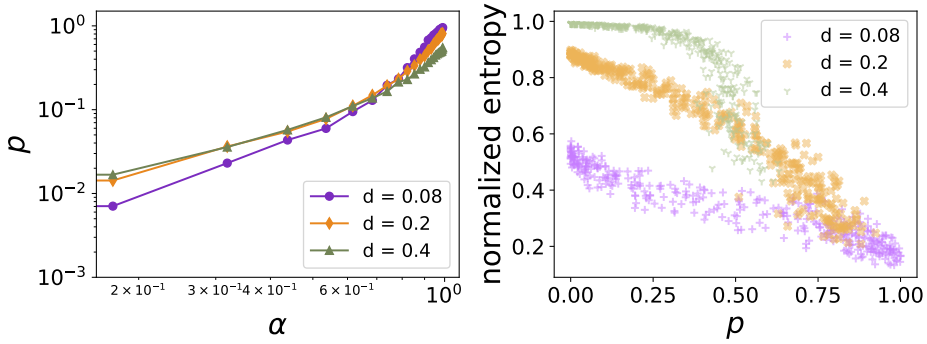

Fig. 4 The relation between memory and diffusion speed for varying average degree per snapshot, d. *Left panel.* Relation between the weight α appearing in Eq. (3) and the estimated p from the **MEM** model of Eq. (1) with $m = \hat{m} = 5$. The color- and marker-coded lines refer to three values of the graph density, reported in the legend. The solid line is obtained by averaging over 10 realizations with fixed parameters, while the shaded line is the $5-95$ confidence interval. Each graph has $n = 250$ nodes, and $T = 300$ snapshots. *Right panel.* For each of the graphs of the left panel, we run 15 simulations of a diffusive process of a signal $\boldsymbol{u}$ starting from a randomly selected node at $t = 0$. The figure shows the relation between the inferred p (the same as in the left panel) and the normalized Shannon entropy of the signal density at the end of the process.

where d denotes the average expected degree of each snapshot. The first summand creates edges between random nodes, while the second enhances the likelihood that two nodes are connected if there is a TRP between them in the last $\hat{m}$ steps. The normalization constant $Z = \frac{1}{n}\sum_{i\neq j} M_{ij}(t;\hat{m})$ balances the two terms. We generate synthetic temporal graphs for varying values of α and infer p using the **MEM** model introduced in Section 2.1. The left panel of Figure 4 shows the relation between p and α, which is monotonically increasing as expected. We also observe that, while the memory p depends on the weight α, it also depends on other graph properties, such as the density.

For each graph, we run 15 instances of a diffusion process of a scalar quantity on the graph's nodes. We let u_i be the abundance of the diffusing quantity at node i and we let $\boldsymbol{u} \in \mathbb{R}^n$ be a vector with entries u_i. We consider the following diffusive process:

$$\boldsymbol{u}_{t+1} = (I_n - \beta\,(D_t - A_t))\,\boldsymbol{u}_t \ ,$$

where β is the diffusion coefficient which we set to $\beta = 0.03$ in the experiments. In each run, we randomly select an active node s at $t = 0$ and let $(u_t)_i = \delta_{is}$. We then compute the normalized Shannon entropy $H(\boldsymbol{u}) = -\sum_{i\in\mathcal{V}} u_i \log(u_i)$, normalized by $\log(n)$ as a measure of the diffusion speed. A low entropy means that the final configuration of $\boldsymbol{u}$ is close to the initial one, while when the normalized entropy tends to one, it means that the system has reached equilibrium. The right panel of Figure 4 evidences the determinant role of memory in slowing down diffusion.

3 Discussion

Time-respecting paths (TRPs) drive dynamical processes on temporal graphs [32]. Here, we focused on high-resolution temporal networks encoding face-to-face proximity between humans, and we introduced a model to quantify memory in TRPs, defined as the probability that a TRP returns to an already visited node. We evaluated our model on several empirical datasets collected by the `SocioPatterns` collaboration [4] and inferred memory effects with comparable results across datasets collected in similar settings. We showed that the inferred memory is statistically significant when compared against memoryless null models. TRPs are constrained to the graph topology, which introduces long correlations in the dynamics and generates long memory effects even in Markovian dynamical processes. Memory in TRPs is observed for small aggregation values t_{res}. In this case, each graph snapshot tends to be sparse and imposes tight topological constraints on the nodes the TRP can visit. For large aggregations, instead, the snapshot density increases allowing a larger variance in the TRP realization. By introducing a simple model to generate graphs with memory, we explicitly showed that memory in the TRPs slows down the speed of diffusion.

Our model provides robust and statistically significant results across datasets, model parameters, and temporal aggregations, but it entails some limitations worth discussing. First, we assumed that a model with time-independent parameters would well describe TRPs, thus that the temporal graph is in a stationary regime. For most of the datasets used in this study – and in general for most datasets – this condition cannot be given for granted. For instance, in the school datasets [5, 6, 36], lesson hours are interspersed by breaks with a different dynamics. To refine our results, one can identify change-points in the data [24] and add a time-dependent notion of memory. Second, unlike other models [24, 25], we could describe memory with a low-complexity model, using a single parameter. Its simplicity, however, does not allow the model to capture heterogeneity patterns across nodes that could be included, for instance, by introducing a memory p depending on the node's community class.

We foresee two main directions that follow from our work. One is the design of generative models for temporal networks with memory in the TRPs. The model introduced in Eq. (3) takes a first step in this direction, but its objective was primarily to evaluate the relation between diffusion speed and memory in a controlled setting, while a more complex model could generate more realistic synthetic data. Such a model would have applications to better simulate infectious disease spread [35, 49, 50] or to provide an anonymized version of sensitive data encoding proximity between humans [51, 52]. The second direction is the study and design of accurate aggregation strategies of temporal data. Compared to a time-agnostic framework, modeling dynamical processes at high temporal resolution requires additional efforts in terms of model complexity and data to collect. However, as shown also in this paper, the properties of the temporal network affect the dynamics of dynamical processes unfolding on top of it. Our work calls for the design of efficient procedures to aggregate temporal data that are aware of properties – such as memory in the TRPs – that are only defined in the temporal setting.

Algorithm 1 `TRP generation`

Input: temporal graph $\{A_t\}_{t=1,\dots,T}$; path-length a
Output: a TRP $\mathcal{P}$
$t_1 \leftarrow$ randomly sample a time with at least one interaction
$v_1 \leftarrow$ sample a node with at least one neighbor at t_1
$v_2 \leftarrow$ sample v_2 w.p. $\propto A_{v_1,v_2}$
$\mathcal{P} \leftarrow \{v_1, v_2\}$, initialize the path
for $2 \leq x \leq a$ **do**
  $t_x \leftarrow$ end of the interaction (v_{x-1}, v_x)
  if $\mathcal{N}_{v_x}(t_x) \setminus \{v_{x-1}\} = \emptyset$ **then**
    $t_x \leftarrow \min_{t>t_x} |\mathcal{N}_{v_x}(t)| > 0$
  end if
  $v_{x+1} \leftarrow$ sample v_{x+1} from $\mathcal{N}_{v_x}(t_x)$ w.p. $\propto A_{v_x,v_{x+1}}$
  $\mathcal{P} \leftarrow$ add v_{x+1}
end for

4 Methods

4.1 Time-Respecting Paths Generative Procedure.

The models proposed in Eqs. (1, 2) describe memory on non-backtracking time-respecting paths (TRPs) on temporal networks. A TRP is a sequence of node-time pairs $\mathcal{P} = \{(v_1, t_1), (v_2, t_2), \dots, (v_a, t_a)\}$ representing the steps of a path on a temporal graph. We require nodes to appear in a chronological order (*i.e.,* that $t_{a+1} > t_a$ for all values of a) and (v_a, v_{a+1}, t_a) to be a temporal edge for all values of a. On top of these conditions, we detail additional definitions of the TRPs that account for known properties of the empirical data under analysis.

First, all datasets of Table 1 span multiple measurement days. We treat each sampling day independently and consider every dataset a collection of temporal graphs. The only exception is the *Hospital* dataset, which we treat as a single temporal graph due to the irregular hours and schedules, with contact occurring almost continuously throughout day and night. We aggregate time with a resolution t_{res} and consider all events within t_{res} as simultaneous. The edge weight equals the number of edge occurrences within t_{res} before the aggregation, ranging from 0 to the number of timesteps within the resolution. To initialize the TRP, we first randomly sample an active time t_1, *i.e.,* a snapshot with at least one temporal edge, and then randomly sample an active node v_1, *i.e.,* a node with at least one neighbor at t_1.

We define the TRP evolution from the first node to account for the interaction and inter-event distributions, which are typically observed to be fat-tailed in temporal social interaction networks [4, 53]. The broadness of this distribution implies that the time two nodes spend in contact does not have a typical scale. With a non-negligible probability, we can hence observe prolonged interactions between node pairs. By performing a step in the TRP at each time index, *i.e.,* letting $t_{a+1} = t_a + 1$, a long interaction between a node pair remains active for several path steps and can be crossed multiple times, confounding a recurrent interaction with a prolonged one. To

counter this effect, we let t_{a+1} be the time at which the interaction between (v_a, v_{a+1}) is concluded. Consequently, the TRPs we obtain have varying durations.

The bursty dynamics typically observed in human activity imply that the time distribution between the end of an interaction and the beginning of the next one is also fat-tailed. Burstiness generates TRPs with sequences of node pairs that alternate, introducing fictitious memory effects [41]. We thus focus on non-backtracking paths in which $v_{a+1} \notin \{v_a, v_{a-1}\}$, thus alleviating this issue.

Algorithm 1 summarizes the TRP generation strategy. The algorithm takes a sequence of time-stamped weighted adjacency matrices and path length a and returns the TRP $\mathcal{P}$. The matrices A_t have size $n \times n$, and the neighborhood of a node u at time t is denoted with $\mathcal{N}_u(t)$. We recall that a TRP has a varying duration, and the path may reach the largest available time before having the desired length a. In that case, we regenerate the TRP.

4.2 Maximum likelihood estimators

In this section, we derive the maximum likelihood estimators for the parameters of the **MEM** and **MEM + SBM** models. We consider a collection of R TRPs generated independently following the procedure detailed in Section 4.1. For each TRP, we attempt to predict the next node in the sequence. By letting $v_{a;r}$ denote the a-th step of the r-th path, the likelihood of the realizations of the $a+1$ nodes over the R TRPs for the **MEM** (Eq. (4)) and **MEM + SBM** (Eq. (5)) models respectively read

$$\mathcal{L}_{\text{MEM}} = \prod_{r=1}^{R} P\left(v_{a+1;r} | \mathcal{M}_{a;r}\right) = \prod_{r=1}^{R} \left[\frac{p \cdot \delta\left(v_{a+1;r} \in \mathcal{M}_{a;r}\right)}{|\mathcal{M}_{a;r}|} + \frac{1-p}{n-2}\right] \tag{4}$$

$$\mathcal{L}_{\text{MEM+SBM}} = \prod_{r=1}^{R} P\left(v_{a+1;r} | \mathcal{M}_{a;r}\right) = \prod_{r=1}^{R} \left[\frac{p \cdot \delta\left(v_{a+1;r} \in \mathcal{M}_{a;r}\right)}{|\mathcal{M}_{a;r}|} + \frac{(1-p) \cdot L(v_{a;r}, v_{a+1;r})}{Z_{a;r}}\right] . \tag{5}$$

In Eq. (5) we used the notation introduced in Section 2.1 and denoted with $L \in \mathbb{R}^{n \times n}$ the symmetric non-negative matrix encoding the community affinity between nodes, *i.e.,* $L(u,v) = C_{\ell(u),\ell(v)}$. The normalization constant reads $Z_{a;r} = \sum_{u \in \mathcal{V} \setminus \{v_{a;r}, v_{a-1;r}\}} L_{v_{a;r},u}$. We derive the maximum likelihood estimators of p and C from Eq. (4), noting that Eq. (4) is obtained from Eq. (5) by letting $L(u,v) = 1$.

Parameter p

We compute the derivative of the total log-likelihood with respect to p and set it to zero. We let $\mathcal{R}_{\text{in}}$ denote the set of paths where the last node belongs to the memory set, *i.e.,* those satisfying $v_{a+1;r} \in \mathcal{M}_{a;r}$. All other paths form the set $\mathcal{R}_{\text{out}}$.

$$0 = \frac{\partial \log(\mathcal{L}_{\text{MEM+SBM}})}{\partial p} = \sum_{r \in \mathcal{R}_{\text{out}}} \frac{1}{p-1} + \sum_{r \in \mathcal{R}_{\text{in}}} \left[\frac{Z_{a;r} - |\mathcal{M}_{a;r}| \cdot L(v_{a;r}, v_{a+1;r})}{p \cdot Z_{a;r} + (1-p) \cdot |\mathcal{M}_{a;r}| \cdot L(v_{a;r}, v_{a+1;r})}\right] .$$

From this equation, we obtain the following implicit expression of the maximum likelihood estimator of p for the **MEM + SBM** model:

$$p = 1 - \frac{|\mathcal{R}_{\text{out}}|}{\sum_{r\in\mathcal{R}_{\text{in}}}\left[p + \left(\frac{Z_{a;r}}{|\mathcal{M}_{a;r}|\cdot L(v_{a;r}, v_{a+1;r})} - 1\right)^{-1}\right]^{-1}}\,. \tag{6}$$

The expression for the **MEM** model is obtained from Eq. (6) by letting $L(v_{a;r}, v_{a+1;r}) = 1$ and $Z_{a;r} = n - 2$.

Community Matrix C

Following the same procedure, we obtain the maximum likelihood estimator of the matrix C, recalling that $L(u,v) = C_{\ell(u),\ell(v)}$. We denote with $q_{\beta;r} = \sum_{u\in\mathcal{V}\setminus\{v_{a;r}, v_{a-1;r}\}}\delta[\beta, l(u)]$ and obtain

$$\begin{aligned}
0 &= \frac{\partial \log(\mathcal{L})}{\partial C_{\alpha,\beta}} \\
&= \sum_{r\in\mathcal{R}_{\text{out}}}\delta\left[\alpha, \ell(v_{a;r})\right]\left(\frac{\delta\left[\beta, \ell(v_{a+1;r})\right]}{C_{\alpha,\beta}} - \frac{q_{\beta;r}}{Z_{a;r}}\right) + \\
&\qquad + (1-p)\sum_{r\in\mathcal{R}_{\text{in}}}\frac{|\mathcal{M}_{a;r}|\delta[\alpha, \ell(v_{a;r})]\Big(\delta[\beta, \ell(v_{a+1;r})]Z_{a;r} - q_{\beta;r}L(v_{a;r}, v_{a+1;r})\Big)}{p\cdot Z_{a;r}^2 + (1-p)\cdot L(v_{a;r}, v_{a+1;r})Z_{a;r}\cdot|\mathcal{M}_{a;r}|}
\end{aligned}$$

Again, we can express the maximum likelihood estimator of C with an implicit expression:

$$C_{\alpha,\beta} = \frac{\sum_{r\in\mathcal{R}_{\text{out}}}\delta\left[\alpha, \ell(v_{a;r})\right]\delta\left[\beta, \ell(v_{a+1;r})\right]}{\sum_{r\in\mathcal{R}_{\text{out}}}\frac{\delta[\alpha, \ell(v_{a;r})]q_{\beta;r}}{Z_{a;r}} - \sum_{r\in\mathcal{R}_{\text{in}}}\frac{(1-p)\delta[\alpha, \ell(v_{a;r})]\left(Z_{a;r} - L(v_{a;r}, v_{a+1;r})q_{\beta;r}\right)\cdot|\mathcal{M}_{a;r}|}{p\cdot Z_{a;r}^2 + (1-p)L(v_{a;r}, v_{a+1;r})Z_{a;r}\cdot|\mathcal{M}_{a;r}|}}\,, \tag{7}$$

where we recall once again the dependence of L on C.

Resolution of the implicit formulas

As we have already commented, Eqs. (6, 7) provide an implicit solution for the maximum likelihood estimators. We detail the procedure adopted to solve these equations. We initialize the parameters, letting $p = 0.5$ and C be the matrix encoding the total number of interactions between communities over the temporal graph duration. We remark that the matrix C (or equivalently of L) is defined up to an unknown multiplicative factor, which is handled by the normalization constant $Z_{a;r}$. With an iterative procedure, we then update $p_{\text{new}} = g(C, p)$, where the function g is the right hand-side

of Eq. (6) and $C_{\text{new}} = f(C, p)$, where f is the right hand-side of Eq. (7), looking for the solutions only in the interval $[0, 1]$.

List of abbreviations. TRP, time respecting paths; SBM, stochastic block model.

Declarations

Availability of data and materials. All data used in this analysis are public and were downloaded from the `SocioPatterns` website sociopatterns.org/datasets/. The codes to reproduce the results are publicly available in the following repository github.com/SilviaGuerrini/MemoryModel where, for convenience, we also share the data we used.

Competing interests. The authors declare no competing interests.

Funding. The authors acknowledge support from the Lagrange Project of the ISI Foundation, funded by Fondazione CRT.

Authors' contributions. SG performed the theoretical analysis, developed computer code, performed the simulations and wrote the first version of the manuscript. LD conceived the project. LD, and CC supervised the project. All authors interpreted the results, reviewed and approved the final version of the manuscript.

Acknowledgements. The authors thank Vincenzo Perri for fruitful discussions.

References

[1] Kroczek, L.O., Pfaller, M., Lange, B., Müller, M., Mühlberger, A.: Interpersonal distance during real-time social interaction: Insights from subjective experience, behavior, and physiology. Frontiers in psychiatry **11**, 561 (2020)

[2] Salathé, M., Kazandjieva, M., Lee, J.W., Levis, P., Feldman, M.W., Jones, J.H.: A high-resolution human contact network for infectious disease transmission. Proceedings of the National Academy of Sciences **107**(51), 22020–22025 (2010) https://doi.org/10.1073/pnas.1009094108 https://www.pnas.org/doi/pdf/10.1073/pnas.1009094108

[3] Barrat, A., Cattuto, C.: Temporal networks of face-to-face human interactions. Temporal Networks, Understanding Complex Systems. ISBN 978-3-642-36460-0. Springer-Verlag Berlin Heidelberg, 2013, p. 191 (2013) https://doi.org/10.1007/978-3-642-36461-7_10

[4] Cattuto, C., Broeck, W., Barrat, A., Colizza, V., Pinton, J.-F., Vespignani, A.: Dynamics of person-to-person interactions from distributed rfid sensor networks. PLoS ONE **5**(7), 11596 (2010) https://doi.org/10.1371/journal.pone.0011596

[5] Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., Quaggiotto, M., Broeck, W., Régis, C., Lina, B., Vanhems, P.: High-resolution measurements

of face-to-face contact patterns in a primary school. PLoS ONE **6**(8), 23176 (2011) https://doi.org/10.1371/journal.pone.0023176

[6] Fournet, J., Barrat, A.: Contact patterns among high school students. PLoS ONE **9**(9), 107878 (2014) https://doi.org/10.1371/journal.pone.0107878

[7] Génois, M., Vestergaard, C.L., Fournet, J., Panisson, A., Bonmarin, I., Barrat, A.: Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. Network Science **3**(3), 326–347 (2015)

[8] Dall'Amico, L., Kleynhans, J., Gauvin, L., Tizzoni, M., Ozella, L., Makhasi, M., Wolter, N., Language, B., Wagner, R.G., Cohen, C., Tempia, S., Cattuto, C.: Estimating household contact matrices structure from easily collectable metadata. PLOS ONE **19**(3), 1–13 (2024) https://doi.org/10.1371/journal.pone.0296810

[9] Perra, N., Gonçalves, B., Pastor-Satorras, R., Vespignani, A.: Activity driven modeling of time varying networks. Scientific Reports **2**(1) (2012) https://doi.org/10.1038/srep00469

[10] Starnini, M., Baronchelli, A., Pastor-Satorras, R.: Modeling human dynamics of face-to-face interaction networks. Physical review letters **110**(16), 168701 (2013)

[11] Ubaldi, E., Vezzani, A., Karsai, M., Perra, N., Burioni, R.: Burstiness and tie activation strategies in time-varying social networks. Scientific Reports **7**(1) (2017) https://doi.org/10.1038/srep46225

[12] Hiraoka, T., Masuda, N., Li, A., Jo, H.-H.: Modeling temporal networks with bursty activity patterns of nodes and links. Physical Review Research **2**(2), 023073 (2020)

[13] Le Bail, D., Génois, M., Barrat, A.: Modeling framework unifying contact and social networks. Physical Review E **107**(2), 024301 (2023)

[14] Sheng, A., Su, Q., Li, A., Wang, L., Plotkin, J.: Constructing temporal networks with bursty activity patterns. Nature Communications **14** (2023) https://doi.org/10.1038/s41467-023-42868-1

[15] Masoumi, R., Gambaudo, J., Génois, M.: Simple crowd dynamics to generate complex temporal contact networks. arXiv preprint arXiv:2405.06508 (2024)

[16] Vestergaard, C.L., Génois, M., Barrat, A.: How memory generates heterogeneous dynamics in temporal networks. Phys. Rev. E **90**, 042805 (2014) https://doi.org/10.1103/PhysRevE.90.042805

[17] Williams, O.E., Lacasa, L., Millán, A.P., Latora, V.: The shape of memory in temporal networks. Nature communications **13**(1), 499 (2022)

[18] Gallo, L., Lacasa, L., Latora, V., Battiston, F.: Higher-order correlations reveal

complex memory in temporal hypergraphs. Nature Communications **15**(1), 4754 (2024)

[19] Tizzani, M., Lenti, S., Ubaldi, E., Vezzani, A., Castellano, C., Burioni, R.: Epidemic spreading and aging in temporal networks with memory. Physical Review E **98**(6) (2018) https://doi.org/10.1103/physreve.98.062315

[20] Williams, O.E., Lillo, F., Latora, V.: Effects of memory on spreading processes in non-markovian temporal networks. New Journal of Physics **21**(4), 043028 (2019) https://doi.org/10.1088/1367-2630/ab13fb

[21] Scholtes, I., Wider, N., Pfitzner, R., Garas, A., Tessone, C., Schweitzer, F.: Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. Nature communications **5**, 5024 (2014) https://doi.org/10.1038/ncomms6024

[22] Karsai, M., Kivelä, M., Pan, R.K., Kaski, K., Kertész, J., Barabási, A.-L., Saramäki, J.: Small but slow world: How network topology and burstiness slow down spreading. Phys. Rev. E **83**, 025102 (2011) https://doi.org/10.1103/PhysRevE.83.025102

[23] Sekara, V., Stopczynski, A., Lehmann, S.: Fundamental structures of dynamic social networks. Proceedings of the National Academy of Sciences **113**(36), 9977–9982 (2016) https://doi.org/10.1073/pnas.1602803113 https://www.pnas.org/doi/pdf/10.1073/pnas.1602803113

[24] Peixoto, T.P., Gauvin, L.: Change points, memory and epidemic spreading in temporal networks. Scientific Reports **8** (2017)

[25] Petrović, L.V., Wegner, A.E., Scholtes, I.: Higher-order patterns reveal causal temporal scales in time series network data. In: The First Learning on Graphs Conference (2022). https://openreview.net/forum?id=al0-qQsFCHV

[26] Rosvall, M., Esquivel, A.V., Lancichinetti, A., West, J.D., Lambiotte, R.: Memory in network flows and its effects on spreading dynamics and community detection. Nature Communications **5**(1) (2014) https://doi.org/10.1038/ncomms5630

[27] Petrović, L.V., Perri, V.: Bayesian detection of mesoscale structures in pathway data on graphs (2023) arXiv:2301.11120 [stat.ME]

[28] Karsai, M., Perra, N., Vespignani, A.: Time varying networks and the weakness of strong ties. Scientific Reports **4**(1) (2014) https://doi.org/10.1038/srep04001

[29] Williams, O.E., Mazzarisi, P., Lillo, F., Latora, V.: Non-markovian temporal networks with auto- and cross-correlated link dynamics. Phys. Rev. E **105**, 034301 (2022) https://doi.org/10.1103/PhysRevE.105.034301

[30] Longa, A., Cencetti, G., Lehmann, S., Passerini, A., Lepri, B.: Generating fine-grained surrogate temporal networks. Communications Physics **7**(1), 22 (2024) https://doi.org/10.1038/s42005-023-01517-1

[31] Girardini, N.A., Longa, A., Trebucchi, G., Cencetti, G., Passerini, A., Lepri, B.: Community aware temporal network generation. arXiv preprint arXiv:2501.07327 (2025)

[32] Pan, R.K., Saramäki, J.: Path lengths, correlations, and centrality in temporal networks. Phys. Rev. E **84**, 016105 (2011) https://doi.org/10.1103/PhysRevE.84.016105

[33] Dall'Amico, L., Barrat, A., Cattuto, C.: An embedding-based distance for temporal graphs. Nature Communications **15**(1) (2024) https://doi.org/10.1038/s41467-024-54280-4

[34] Génois, M., Barrat, A.: Can co-location be used as a proxy for face-to-face contacts? EPJ Data Science **7**(1), 11 (2018) https://doi.org/10.1140/epjds/s13688-018-0140-1

[35] Gemmetto, V., Barrat, A., Cattuto, C.: Mitigation of infectious disease at school: targeted class closure vs school closure. BMC infectious diseases **14**(1), 695 (2014) https://doi.org/10.1186/PREACCEPT-6851518521414365

[36] Mastrandrea, R., Fournet, J., Barrat, A.: Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. PLoS ONE **10**(9) (2015) https://doi.org/10.1371/journal.pone.0136497

[37] Vanhems, P., Barrat, A., Cattuto, C., Pinton, J.-F., Khanafer, N., Régis, C., Kim, B.-a., Comte, B., Voirin, N.: Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. PLoS ONE **8**(9), 73970 (2013) https://doi.org/10.1371/journal.pone.0073970

[38] Ozella, L., Paolotti, D., Lichand, G., Rodríguez, J.P., Haenni, S., Phuka, J., Leal-Neto, O.B., Cattuto, C.: Using wearable proximity sensors to characterize social contact patterns in a village of rural malawi. EPJ Data Science **10**(1), 46 (2021)

[39] Holme, P., Saramäki, J.: Temporal networks. Physics Reports **519**(3), 97–125 (2012) https://doi.org/10.1016/j.physrep.2012.03.001

[40] Sato, K., Oka, M., Barrat, A., Cattuto, C.: Predicting partially observed processes on temporal networks by dynamics-aware node embeddings (dyane). EPJ Data Science **10**(1), 22 (2021)

[41] Saramäki, J., Holme, P.: Exploring temporal networks with greedy walks. The European Physical Journal B **88**, 1–8 (2015)

[42] Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proceedings of the national academy of sciences **99**(12), 7821–7826 (2002)

[43] Holland, P.W., Laskey, K.B., Leinhardt, S.: Stochastic blockmodels: First steps. Social networks **5**(2), 109–137 (1983)

[44] Burnham, K.P., Anderson, D.R.: Model Selection and Multimodel Inference: a Practical Information-theoretic Approach. Springer, ??? (2002)

[45] Goeyvaerts, N., Santermans, E., Potter, G., Torneri, A., Van Kerckhove, K., Willem, L., Aerts, M., Beutels, P., Hens, N.: Household members do not contact each other at random: implications for infectious disease modelling. Proceedings of the Royal Society B **285**(1893), 20182201 (2018)

[46] Gauvin, L., Génois, M., Karsai, M., Kivelä, M., Takaguchi, T., Valdano, E., Vestergaard, C.L.: Randomized reference models for temporal networks. SIAM Review **64**(4), 763–830 (2022) https://doi.org/10.1137/19M1242252 https://doi.org/10.1137/19M1242252

[47] Rocha, L.E., Liljeros, F., Holme, P.: Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. PLoS computational biology **7**(3), 1001109 (2011)

[48] Artime, O., Ramasco, J.J., San Miguel, M.: Dynamics on networks: competition of temporal and topological correlations. Scientific reports **7**(1), 41627 (2017)

[49] Cencetti, G., Santin, G., Longa, A., Pigani, E., Barrat, A., Cattuto, C., Lehmann, S., Salathe, M., Lepri, B.: Digital proximity tracing on empirical contact networks for pandemic control. Nature communications **12**(1), 1655 (2021)

[50] Colosi, E., Bassignana, G., Contreras, D.A., Poirier, C., Boëlle, P.-Y., Cauchemez, S., Yazdanpanah, Y., Lina, B., Fontanet, A., Barrat, A., *et al.*: Screening and vaccination against covid-19 to minimise school closure: a modelling study. The Lancet Infectious Diseases **22**(7), 977–989 (2022)

[51] Creţu, A.-M., Monti, F., Marrone, S., Dong, X., Bronstein, M., Montjoye, Y.-A.: Interaction data are identifiable even across long periods of time. Nature Communications **13**(1), 313 (2022)

[52] Romanini, D., Lehmann, S., Kivelä, M.: Privacy and uniqueness of neighborhoods in social networks. Scientific reports **11**(1), 20104 (2021)

[53] Isella, L., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J., Van den Broeck, W.: What's in a crowd? analysis of face-to-face behavioral networks. Journal of Theoretical Biology **271**(1), 166–180 (2011) https://doi.org/10.1016/j.jtbi.2010.11.033

A Appendix

Inference results for different temporal resolution values

TRPs on time-resolved human proximity data exhibit statistically significant memory effects, even if TRPs' steps do not depend on their history. This is possible because TRPs are constrained on the graph's temporal edges. However, this effect deteriorates as the temporal resolution parameter t_{res} increases. Higher values of temporal resolution impose milder constraints on the path dynamics because the number of edges at each time step can only increase with t_{res}. In this section, we evaluate the two models across different resolution values t_{res}. Figure 5 shows the maximum likelihood estimates of the memory parameter p in the **MEM** model (Eq. 1), while figure 6 in the **MEM + SBM** model (Eq. 2), across various temporal resolution values. As expected, both figures show a decrease in the inferred memory parameter p as the temporal resolution value t_{res} increases. The effect is already evident between the first two resolution values.

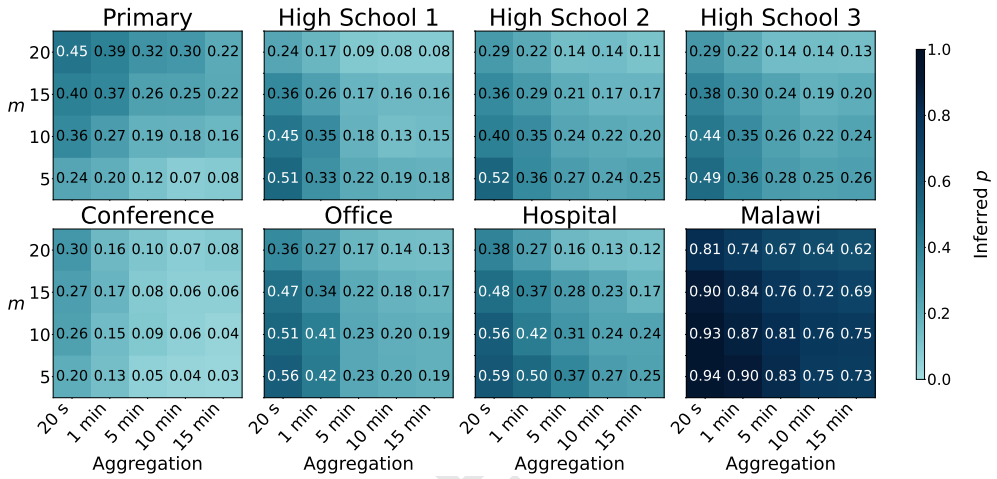

Fig. 5 Maximum likelihood estimates of the memory parameter p. Each plot refers to one of the datasets described in Table 1 and shows the inferred values of the probability p in the **MEM** model (Eq. 1), probability of the time-respecting paths to return to a node. Paths are generated as per Section 4 for different *memory horizon* lengths per path and temporal resolution values t_{res}.

Results for memory expressed in time units

As mentioned in the main text, the TRP size can be defined in two ways: one is the *path length* that equals the number of nodes in the path; the other is the *duration*, *i.e.,* the time elapsed between the last and first steps in the path. In this section, we explore the results for the *memory horizon* expressed in actual time duration (in minutes).

Figure 7 shows the BIC values for the two models as a function of the memory time duration (in minutes). In all datasets the **MEM + SBM** model achieves better results

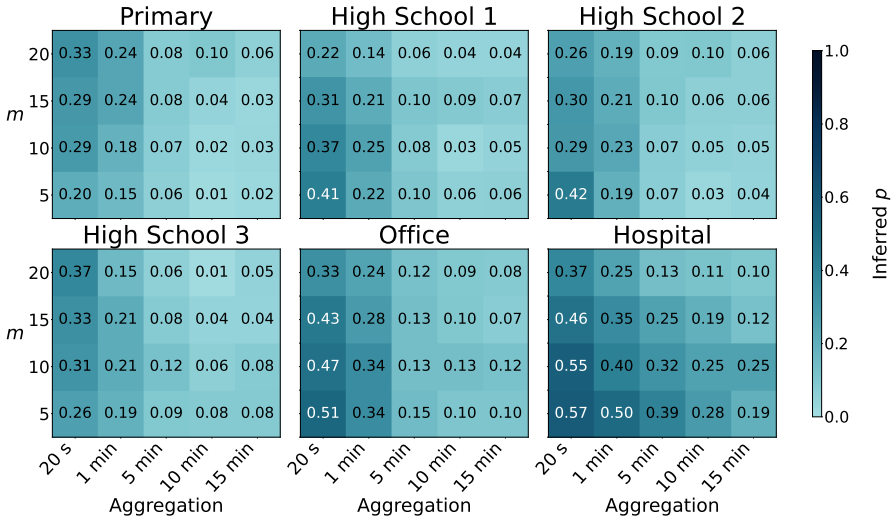

Fig. 6 Maximum likelihood estimates of the memory parameter p. Each plot refers to one of the datasets described in Table 1 and shows the inferred values of the probability p in the **MEM+SBM** model (Eq. 2), probability of the time-respecting paths to return to a node. Paths are generated as described in Section 4 for different *memory horizon* lengths per path and temporal resolution values t_{res}.

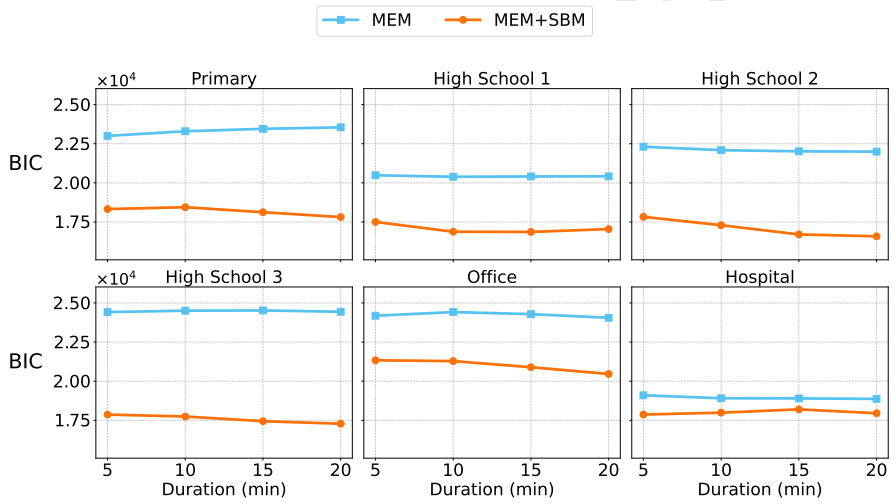

Fig. 7 Model selection between the MEM and MEM + SBM models. Each plot refers to one of the six datasets of Table 1 with a known node label assignment. We show the BIC_{MEM} (Equation (1)) and $\text{BIC}_{\text{MEM+SBM}}$ (Equation (2)) as a function of the memory time duration (min) for $t_{\text{res}} = 20$ *s*. Notice that the scale does not imply comparable results across the datasets.

than the **MEM** model showing the community structure needs to be accounted for to explain the TRPs statistic. This is especially evident in the four school datasets that display a highly assortative community structure. Figure 8 shows the inferred mem-

ory parameter p as a function of the memory time duration across datasets. For all datasets with known node labels, we compare the two models. For the *Conference* and *Malawi* datasets, we report only the inferred memory values from the **MEM** model. For all schools, **MEM** provides slightly larger memory values p than those observed in the **MEM+SBM** model. We observe similar inferred values of p for all schools and in particular all high schools. All workplaces – *Office* and *Hospital* – lead to comparable results, with values of p slightly larger than those observed in schools. The high memory observed in *Malawi* and the low one in *Conference* is understood from the experimental context. *Malawi* describes interactions between family members in an African rural village. These interactions are known to be frequent and prolonged [45], thus explaining the high memory effects observed. *Conference* describes the interactions among researchers at a scientific conference. Commonly, attendees browse around to meet new people, naturally decreasing the memory observed in the empirical data.

As we did for the path length configuration, we assess the statistical significance of our results by comparing them with two null models that do not have memory by construction. We observe that in all high school datasets, the inferred p on the null models is non-zero, which is probably caused by a highly homophilic community structure. Conversely, in all the other datasets, the memory parameter p inferred from both models is consistently zero. Across all datasets, the inferred memory values are much larger than the values inferred from the null models, evidencing a non-trivial memory effect in empirical pathway data.

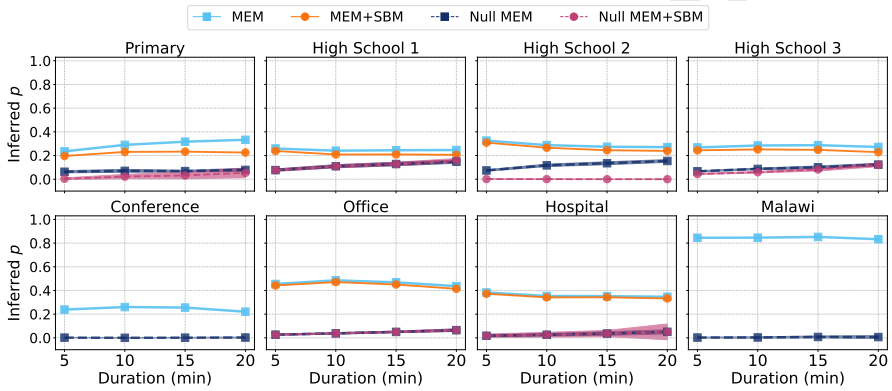

Fig. 8 Maximum likelihood estimates of the memory parameter p and comparison with the null models. Each plot refers to one of the datasets described in Table 1 and shows the inferred value of the probability p of the time-respecting paths to return to a node, as a function of the memory time duration (min). The curve "MEM" (solid line with squares) is obtained from the **MEM** model of Equation (1) on the empirical data. The curve "MEM+SBM" (solid line with dots) is obtained from the **MEM+SBM** model of Equation (2) on the empirical data. For the *Conference* and *Malawi* datasets, we only consider the null model based on Erdős-Rńyi random graphs, while for all other datasets, we consider the one generated from stochastic block model graphs. The curve "Null MEM" (dashed line with squares) is obtained from the **MEM** model (Equation (1)) on the synthetic data. The curve "Null MEM + SBM" (dashed line with dots) is obtained from the **MEM+SBM** model Equation (2)) on the synthetic data. The curves referring to the null models show the mean over 20 realizations, and the shaded areas represent the standard deviation. For all graphs, we consider the temporal resolution $t_{\text{res}} = 20\ s$.