SCIENTIFIC REPORTS

OPEN

Change points, memory and epidemic spreading in temporal networks

Received: 5 March 2018
Accepted: 17 September 2018
Published online: 19 October 2018

Tiago P. Peixoto[1,2] & Laetitia Gauvin[2]

Dynamic networks exhibit temporal patterns that vary across different time scales, all of which can potentially affect processes that take place on the network. However, most data-driven approaches used to model time-varying networks attempt to capture only a single characteristic time scale in isolation — typically associated with the short-time memory of a Markov chain or with long-time abrupt changes caused by external or systemic events. Here we propose a unified approach to model both aspects simultaneously, detecting short and long-time behaviors of temporal networks. We do so by developing an arbitrary-order mixed Markov model with change points, and using a nonparametric Bayesian formulation that allows the Markov order and the position of change points to be determined from data without overfitting. In addition, we evaluate the quality of the multiscale model in its capacity to reproduce the spreading of epidemics on the temporal network, and we show that describing multiple time scales simultaneously has a synergistic effect, where statistically significant features are uncovered that otherwise would remain hidden by treating each time scale independently.

Recent advances in the study of network systems — usually with social, technological and biological origins — have been moving beyond the more traditional approach of considering them as static or growing entities, and instead have been introducing more realistic descriptions that allow them to change arbitrarily in time[1,2]. This effort includes modeling of the time-varying network structure[3,4], as well as processes that take place on this dynamic environment, such as epidemic spreading[5–8]. Further recent works[9–11] have highlighted the role of memory, burstiness and time ordering as key features of empirical temporal networks that affect dynamical processes taking place on it.

Most approaches, however, rely on a characteristic time scale on which they describe the dynamics. These can be divided, roughly, into approaches that model temporal correlations via Markov chains relating short-time memory with future behavior[12,13], and those that model the dynamics at longer times, usually via network snapshots[14–19] or discrete change points[20–22]. For example, in Refs.[12,13] the time evolution of a network is represented as a static Markov chain where the placement of new edges is conditioned on the last few edges placed. Since the transition probabilities themselves do not change in time, the system eventually reaches equilibrium and cannot maintain any kind of long-term memory. Conversely, the approaches of Refs.[14–22] do not attempt to model any kind of short term memory, and simply divide the temporal evolution into discrete intervals, according to how large is the change in the network structure between these intervals. In so doing, these approaches focus only on a larger temporal scale, describing only abrupt changes in the large-scale network structure. In reality, however, most systems exhibit both kinds of dynamics, and focusing on a single aspect comes at the expense of ignoring the other. In this work, we introduce a data-driven modeling approach that includes both aspects simultaneously, and is capable of uncovering both the short-time Markov properties as well a the long-time abrupt changes.

We develop a Bayesian formulation that allows both the change points and the Markov order to be inferred from data in a principled manner, prevents overfitting and enables model selection. As an extraneous evaluation of our approach, we investigate the behavior of epidemic spreading both in the original data and in artificial ones generated from our inferred models. We show that the most plausible models tend to mix both short-time memory and many change points, and those tend to capture well the nontrivial epidemic behavior observed in the original data. Importantly, the inferred models with change points typically uncover higher-order memory

[1]Department of Mathematical Sciences and Centre for Networks and Collective Behaviour, University of Bath, Claverton Down, Bath, BA2 7AY, United Kingdom. [2]ISI Foundation, Via Chisola 5, 10126, Torino, Italy. Correspondence and requests for materials should be addressed to T.P.P. (email: t.peixoto@bath.ac.uk)

than the simpler stationary variants, demonstrating that the mixed approach is more powerful than considering individual ones in isolation.

This paper is divided as follows. In Sec. 2.1 we present the epidemic models that will be used for the model comparison. In Sec. 2.2 we describe our modeling and inferring approach, and apply it to empirical data. In Sec. 4 we finalize with a conclusion.

Results

Proximity networks and epidemic dynamics. In the interest of simplicity, we will consider a minimal model of temporal networks and epidemic dynamics that takes place on it. The most central simplification we will make is that the dynamics takes place in discrete time, so that the placement of edges forms a temporal sequence, where only one edge is placed at any given time. Real dynamical networks and epidemic spreading occur in continuous time, but our objective here is not to construct a detailed realistic model, but rather to illustrate how multiple time scales can be described simultaneously. More realistic features can then be added to the model at a later stage.

More specifically, we consider temporal networks composed of N nodes, where the placement of the edges occurs sequentially in time, i.e. they define a sequence $\boldsymbol{s} = \{x_t\}$, where $x_t = (u, v)_t$ is an edge between nodes u and v observed at time t, with $t = \{1, 2, \ldots, E\}$, where E is the total number of edge occurrences, and the number of nodes N remains constant. Although this formulation is general, we focus in particular on proximity networks, obtained by tracking volunteers with wearable sensors over a period of time[23–26], so that an edge $(u, v)_t$ is recorded if the respective people came closer than a given radius at time t. Data recorded in this manner possess enough time resolution for our analysis, and also serve as a plausible scenario for epidemic spreading[27].

In the above scenario, we assume that an infection can only occur at time t over the current "active" edge $(u, v)_t$. If the epidemics follows the Susceptible-Infected-Recovered (SIR) model, and $\sigma_u(t) \in \{\mathrm{S}, \mathrm{I}, \mathrm{R}\}$ is the state of node u at time t, we have at each time step t:

1. If $(u, v)_t$ is the current edge, with $(\sigma_u(t-1), \sigma_v(t-1)) = (S, I)$ or (I, S), the infection spreads with probability β, so that $(\sigma_u(t), \sigma_v(t)) = (I, I)$.
2. For every infected node u with $\sigma_u(t-1) = I$, it becomes recovered $\sigma_u(t) = R$ with probability γ.

The parameters β and γ control the infection and recovery probabilities, respectively. We also consider the Susceptible-Infected-Susceptible (SIS) model, which is a variation of the above, where in the second step the infected nodes become susceptible, $\sigma_u(t) = S$, instead of recovered. In both cases, we consider the total number of infected nodes at given time t, $X(t)$. For any positive recovery probability $\gamma > 0$, the long-time behavior of the SIR model is always $\lim_{t\to\infty} X(t) = 0$, as the outbreak invariably dies out, whereas in the SIS model it can persist for arbitrarily long times in large systems. In the following, we will use the behavior of $X(t)$ as a proxy for the comparison between data and model in capturing the underlying network dynamics.

When considering epidemics on dynamical networks, there are two properties that are believed to be crucial for the spreading process[10,11]: 1. The distribution of number of contacts per link, i.e. the frequency of token x in sequence $\boldsymbol{s}$, and 2. The distribution of waiting (or inter-event) times, i.e. the time between two occurrences of the same edge. Although a link that occurs frequently is likely to have shorter inter-event times, the latter tends to vary in ranges that cannot be explained fully by the former, and represents temporal correlations that go beyond the mere frequency of occurrence of edges[10,11]. We will have these two aspects in mind when elaborating our models.

Models for temporal networks. Our objective is to construct a generative model for temporal networks that includes both short-term memories and abrupt change points. We begin by formulating a stationary version, without change points, and show how it is insufficient to capture many features in the data. We then extend the model to include change points, and perform a comparison.

Stationary Markov chains. We consider sequences of discrete tokens, i.e. edges, $\boldsymbol{s} = \{x_t\}$ with $t \in \{1, \ldots, E\}$ being by definition both the time and the number of edges that have been placed, and $x_t \in \{1, \ldots, D\}$ the set of unique edges with cardinality D, which are generated from a stationary Markov chain of order n, i.e. they occur with probability

$$P(\boldsymbol{s}|\boldsymbol{p}, n) = \prod_t p_{x_t, \boldsymbol{x}_{t-1}} = \prod_{x,\boldsymbol{x}} p_{x,\boldsymbol{x}}^{a_{x,\boldsymbol{x}}}, \tag{1}$$

where $\boldsymbol{p}$ corresponds to the transition matrix and $p_{x_t, \boldsymbol{x}_{t-1}}$ is the probability of observing token x_t given the previous n tokens $\boldsymbol{x}_{t-1} = \{x_{t-1}, \ldots, x_{t-n}\}$ in the sequence, and $a_{x,\boldsymbol{x}}$ is the number of observed transitions from memory $\boldsymbol{x}$ to token x. This serves a simple model for temporal networks, where each possible token corresponds to an edge in the network, i.e. $x_t \equiv (i, j)_t$, as we considered previously. Despite its simplicity, this model is able to reproduce arbitrary edge frequencies, determined by the steady-state distribution of the tokens x, as well as causal temporal correlations between edges. This means that the model should be able to reproduce properties of the data that can be attributed to the distribution of number of contacts per link, which are believed to be important for epidemic spreading[10,11]. However, due to its Markovian nature, the dynamics will eventually forget past states, and converge to the limiting distribution (assuming the chain is ergodic and aperiodic). This latter property means that the model should be able to capture nontrivial statistics of waiting times only at a short time scale, comparable to the Markov order.

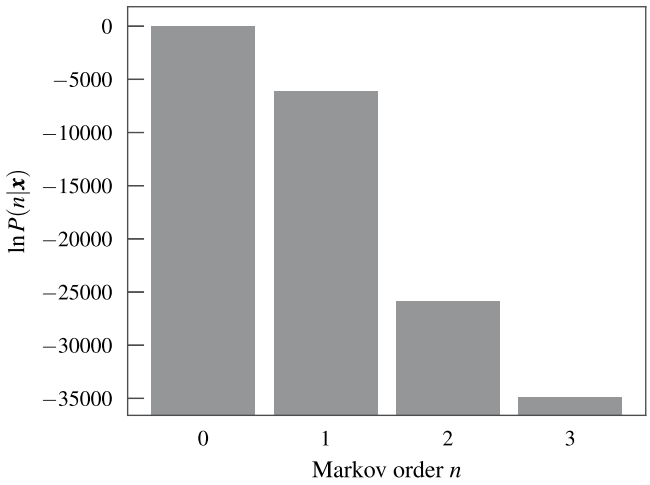

Figure 1. Posterior distribution of the Markov order $P(n|\boldsymbol{x})$ (Eq. 3) for a temporal network between students in a high school[36].

Given the above model, the simplest way to proceed would be to infer transition probabilities from data using maximum likelihood, i.e. maximizing Eq. 1 under the normalization constraint $\sum_x p_{x,\boldsymbol{x}} = 1$. This yields

$$\hat{p}_{x,\boldsymbol{x}} = \frac{a_{x,\boldsymbol{x}}}{k_{\boldsymbol{x}}}. \tag{2}$$

where $k_{\boldsymbol{x}} = \sum_x a_{x,\boldsymbol{x}}$ is the number of transitions originating from $\boldsymbol{x}$. However, if we want to determine the most appropriate Markov order n that fits the data, the maximum likelihood approach cannot be used, as it will *overfit*, i.e. the likelihood of Eq. 1 will increase monotonically with n, favoring the most complicated model possible, and thus confounding statistical fluctuations with actual structure. Instead, the most appropriate way to proceed is to consider the Bayesian posterior distribution

$$P(n|\boldsymbol{s}) = \frac{P(\boldsymbol{s}|n)P(n)}{P(\boldsymbol{s})}, \tag{3}$$

which involves the integrated marginal likelihood[28]

$$P(\boldsymbol{s}|n) = \int P(\boldsymbol{s}|\boldsymbol{p},\ n)P(\boldsymbol{p}|n)\,\mathrm{d}\boldsymbol{p}, \tag{4}$$

where the prior probability $P(\boldsymbol{p}|n)$ encodes the amount of knowledge we have on the transitions $\boldsymbol{p}$ before we observe the data. If we possess no information, we can be agnostic by choosing a uniform prior

$$P(\boldsymbol{p}|n) = \prod_{\boldsymbol{x}} (D-1)!\delta\left(1 - \sum_x p_{x,\boldsymbol{x}}\right), \tag{5}$$

which assumes that all transition probabilities are equally likely. Inserting Eqs. 1 and 5 in Eq. 4, and calculating the integral we obtain

$$P(\boldsymbol{s}|n) = \prod_{\boldsymbol{x}} \frac{(D-1)!}{(k_{\boldsymbol{x}} + D - 1)!} \prod_x a_{x,\boldsymbol{x}}!. \tag{6}$$

The remaining prior, $P(n)$, that represents our *a priori* preference to the Markov order, can also be chosen in an agnostic fashion in a range $[0, N]$, i.e.

$$P(n) = \frac{1}{N+1}. \tag{7}$$

Since this prior is a constant, the upper bound N has no effect on the posterior of Eq. 3, provided it is sufficiently large to include most of the distribution.

Differently from the maximum-likelihood approach described previously, the posterior distribution of Eq. 3 will select the size of the model to match the statistical significance available, and will favor a more complicated model only if the data cannot be suitably explained by a simpler one, i.e. it corresponds to an implementation of Occam's razor that prevents overfitting.

When applying this approach to empirical data, we observe that it favors $n = 0$ for all datasets we considered (not shown), indicating that a higher-order model is not statistically justified, as can be seen in Fig. 1. However, if we generate temporal networks from the fitted models, i.e. sequence of edges using the transition probabilities

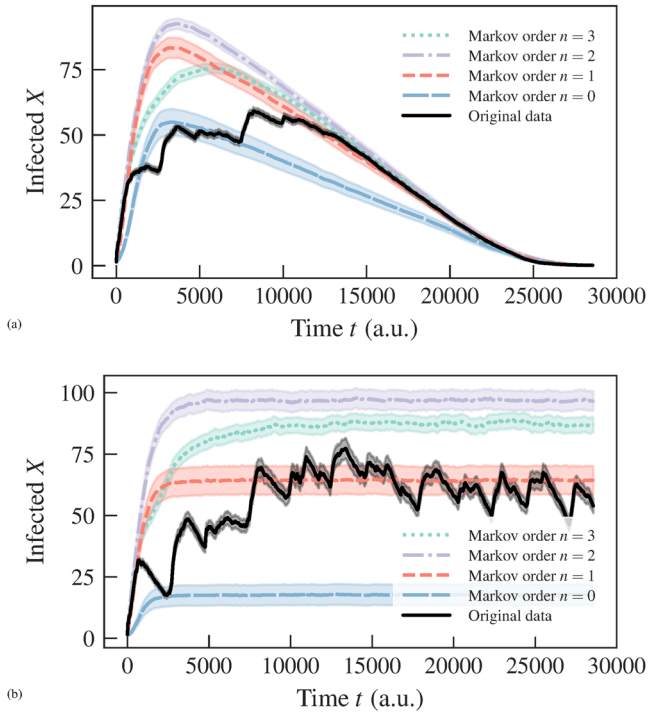

Figure 2. Number of infected nodes over time $X(t)$ for a temporal network between students in a high-school[36] ($N = 126$), considering both the original data and artificial time-series generated from the fitted Markov model of a given order n, using (**a**) SIR ($\beta = 0.41$, $\gamma = 0.005$) and (**b**) SIS ($\beta = 0.61$, $\gamma = 0.03$) epidemic models. In all cases, the values were averaged over 100 independent realizations of the network model (for the artificial datasets) and dynamics. The shaded areas indicate the standard deviation of the mean. The values of the infection and recovery rates were chosen so that the spreading dynamics spans the entire time range of the dataset.

$\hat{p}_{x,\boldsymbol{x}} = a_{x,\boldsymbol{x}}/k_{\boldsymbol{x}}$, they exhibit epidemic dynamics that are very different from what we observe on the empirical time-series, as can be seen in Fig. 2: for the original data, the epidemic spreading is marked by abrupt changes in the infection rate, which are not reproduced by the model for any value of Markov order n — even those that overfit. Therefore, these patterns in the epidemic dynamics seem to stem from changes in the underlying structure of the temporal network that are not captured by the above Markov model. Among other things, this means that the behavior cannot be explained by a heterogeneous distribution of edge frequencies, as this is well described by the model. As we show in the next section, the situation changes considerably once we generalize the model to incorporate heterogeneous Markov chains with change points.

Markov chains with change points. We attempt to model the abrupt changes observed in the previous section by non-stationary transition probabilities $p_{x,\boldsymbol{x}}$ that change abruptly at a given "change point," but otherwise remain constant between change points. The occurrence of change points is governed by the probability q that one is inserted at any given time. The existence of M change points divide the time series into $M+1$ temporal segments indexed by $l \in \{0, \ldots, M\}$. The variable l_t indicates to which temporal segment a given time t belongs among the M segments. Thus, the conditional probability of observing a token x at time t in segment l_t is given by

$$P(x_t, l_t | \boldsymbol{x}_{t-1}, l_{t-1}) = p^{l_t}_{x,\boldsymbol{x}} \left[q\left(1 - \delta_{l_t, l_{t-1}}\right) + (1 - q)\delta_{l_t, l_{t-1}} \right], \tag{8}$$

where $p^{l_t}_{x,\boldsymbol{x}}$ is the transition probability inside segment l_t and q is the probability to transit from segment l to $l+1$. The probability of a whole sequence $\boldsymbol{s} = \{x_t\}$ and $\boldsymbol{l} = \{l_t\}$ being generated is then

$$P(\boldsymbol{s}, \boldsymbol{l} | \boldsymbol{p}, q) = q^M (1 - q)^{E-M} \prod_{l,x,\boldsymbol{x}} \left(p^l_{x,\boldsymbol{x}}\right)^{a^l_{x,\boldsymbol{x}}} \tag{9}$$

where $a^l_{x,\boldsymbol{x}}$ is the number of transitions from memory $\boldsymbol{x}$ to token x in the segment l. Note that we recover the stationary model of Eq. 1 by setting $q = 0$. The maximum-likelihood estimates of the parameters are

$$\hat{p}^l_{x,\boldsymbol{x}} = \frac{a^l_{x,\boldsymbol{x}}}{k^l_{\boldsymbol{x}}}, \quad \hat{q} = \frac{M}{E} \tag{10}$$

where $k_{\boldsymbol{x}}^{l} = \sum_{x} a_{x,\boldsymbol{x}}^{l}$ is the number of transitions originating from $\boldsymbol{x}$ in a segment l. But once more, we want to infer the model the segments $\boldsymbol{l}$ in a Bayesian way, via the posterior distribution

$$P(\boldsymbol{l}|\boldsymbol{s},\, n) = \frac{P(\boldsymbol{s},\, \boldsymbol{l}|n)}{P(\boldsymbol{s}|n)}, \tag{11}$$

where the numerator is the integrated likelihood

$$P(\boldsymbol{s},\, \boldsymbol{l}|n) = \int P(\boldsymbol{s},\, \boldsymbol{l}|\boldsymbol{p},\, q,\, n)P(\boldsymbol{p}|n)P(q)\ \mathrm{d}\,\boldsymbol{p}\ \mathrm{d}q \tag{12}$$

using uniform priors $P(q) = 1$, and

$$P(\boldsymbol{p}|n) = \prod_{l} P(\boldsymbol{p}_l|d_l,\, n)P(d_l), \tag{13}$$

with the uniform prior

$$P(\boldsymbol{p}_l|d_l,\, n) = \prod_{x} (D_l - 1)!\delta\left(1 - \sum_{x} p_{x,x}^{l}\right). \tag{14}$$

and

$$P(d_l) = 2^{-D} \tag{15}$$

being the prior for the alphabet d_l of size D_l inside segment l, sampled uniformly from all possible subsets of the overall alphabet of size D. Performing the above integral, we obtain

$$P(\boldsymbol{x},\, \boldsymbol{l}|n) = 2^{-D(M+1)} \frac{M!(E-M)!}{(E+1)!} \prod_{l} \prod_{\boldsymbol{x}} \frac{(D_l - 1)!}{(k_{\boldsymbol{x}}^{l} + D_l - 1)!} \prod_{x} a_{x,\boldsymbol{x}}^{l}!. \tag{16}$$

Like with the previous stationary model, both the order and the positions of the change points can be inferred from the joint posterior distribution

$$P(\boldsymbol{l},\, n|\boldsymbol{x}) = \frac{P(\boldsymbol{x},\, \boldsymbol{l}|n)P(n)}{P(\boldsymbol{x})}, \tag{17}$$

in a manner that intrinsically prevents overfitting. This constitutes a robust and elegant way of extracting this information from data, that contrasts with non-Bayesian methods of detecting change points using Markov chains that tend to be more cumbersome[29], and is more versatile than approaches that have a fixed Markov order[30].

The exact computation of the posterior of Eq. 11 would require the marginalization of the above distribution for all possible segments $\boldsymbol{l}$, yielding the denominator $P(\boldsymbol{x}|n)$, which is unfeasible for all but the smallest time series. However, it is not necessary to compute this value if we sample $\boldsymbol{l}$ from the posterior using Monte Carlo. We do so by making move proposals $\boldsymbol{l} \rightarrow \boldsymbol{l}'$ with probability $P(\boldsymbol{l}'|\boldsymbol{l})$, and accepting it with probability a according to the Metropolis-Hastings criterion[31,32]

$$a = \min\left(1,\, \frac{P(\boldsymbol{l}'|\boldsymbol{x},\, n)P(\boldsymbol{l}|\boldsymbol{l}')}{P(\boldsymbol{l}|\boldsymbol{x},\, n)P(\boldsymbol{l}'|\boldsymbol{l})}\right), \tag{18}$$

which does not require the computation of $P(\boldsymbol{x}|n)$ as it cancels out in the ratio. If the move proposals are ergodic, i.e. they allow every possible partition $\boldsymbol{l}$ to be visited eventually, this algorithm will asymptotically sample from the desired posterior. Here we use the following move proposal scheme, choosing between one the following actions with equal probability:

1. We select a segment randomly and split it in a random point in the middle.
2. We merge two adjacent segments.
3. We move a randomly chosen boundary to a random position between the two enclosing ones.

We perform this algorithm many times, starting from a single segment, and waiting sufficiently long for equilibration — determined by observing if the likelihood value no longer changes significantly — and we choose the partition with the largest probability across runs. For all datasets we investigated, we observed a fast convergence of this algorithm, which typically shows very little variation between runs.

Note that it is also possible to change the Markov order during the algorithm, by proposing moves $n \rightarrow n'$, and using the Metropolis-Hastings criterion to accept or reject them. However, we found that Markov order typically settles very early in the algorithm, and no longer changes during the remaining run, as it incurs a macroscopic change in the likelihood. Since changing the Markov order is an expensive operation of order $O(E)$, we have found it is best to leave it fixed during the MCMC, and select it later according to the likelihood value.

Once a fit is obtained, we can compare the above model with the stationary one by computing the posterior odds ratio

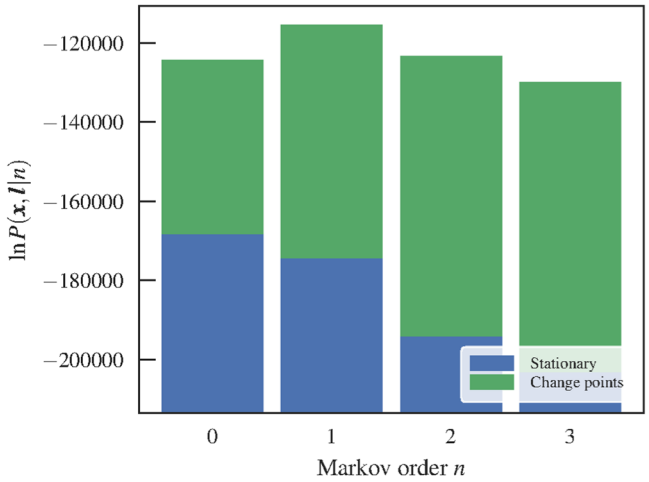

Figure 3. Integrated joint likelihood $P(\boldsymbol{x}, \boldsymbol{l}|n)$ (Eq. 16) for a temporal network between students in a high school[36], for the stationary (i.e. zero change points) and nonstationary models. For all values of n the likelihoods are higher for the nonstationary model (yielding a posterior odds ratio $\Lambda > 1$).

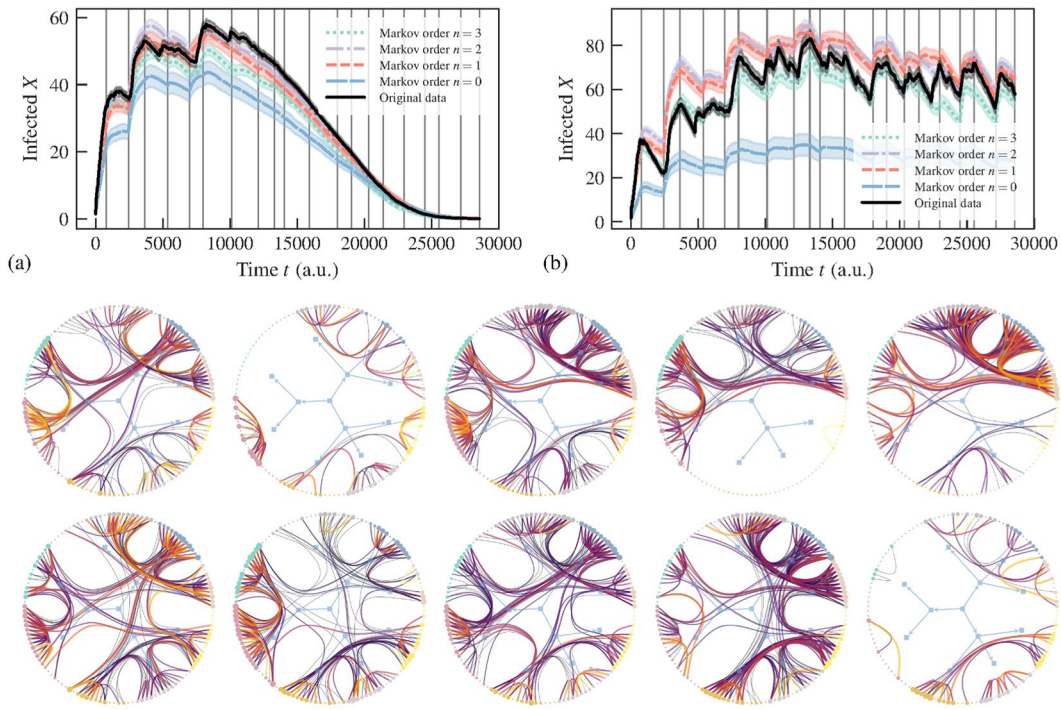

Figure 4. (Above) Number of infected nodes over time $X(t)$ for a temporal network between students in a high-school[36] ($N = 126$), considering both the original data and artificial time-series generated from the fitted nonstationary Markov model of a given order n, using (**a**) SIR ($\beta = 0.41$, $\gamma = 0.005$) and (**b**) SIS ($\beta = 0.61$, $\gamma = 0.03$) epidemic models. The vertical lines mark the position of the inferred change points. In all cases, the values were averaged over 100 independent realizations of the network model (for the artificial datasets) and dynamics. The shaded areas indicate the standard deviation of the mean. (Below) Network structure inside the first ten segments, as captured by a layered hierarchical degree-corrected stochastic block model[16] using the frequency of interactions as edge covariates[33] (indicated by colors), where each segment is considered as a different layer. The values of the infection and recovery rates were chosen so that the spreading dynamics spans the entire time range of the dataset.

$$\Lambda = \frac{P(\boldsymbol{l}, n|\boldsymbol{x})}{P(\boldsymbol{l}_0, n_0|\boldsymbol{x})} = \frac{P(\boldsymbol{x}, \boldsymbol{l}|n)}{P(\boldsymbol{x}, \boldsymbol{l}_0|n_0)}, \tag{19}$$

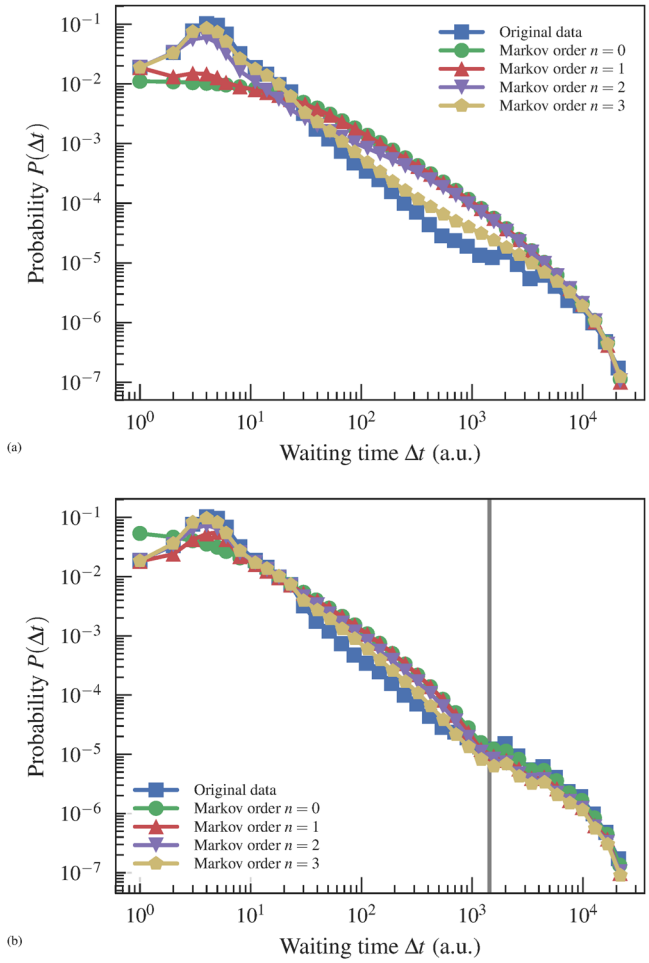

Figure 5. Distribution of waiting times Δt between the same edge for the empirical dataset and fitted (**a**) stationary and (**b**) nonstationary models (a single instance of each), for a temporal network between students in a high school[36]. The vertical line shows the average length of inferred stationary segments between change points.

where $\boldsymbol{l}_0$ is the partition into a single interval (which is equivalent to the stationary model). A value $\Lambda > 1$ [i.e. $P(\boldsymbol{x}, \boldsymbol{l}|n) > P(\boldsymbol{x}, \boldsymbol{l}_0|n_0)$] indicates a larger evidence for the nonstationary model. As can be seen in Fig. 3, we observe indeed a larger evidence for the nonstationary model for all Markov orders. In addition to this, using this general model we identify $n = 1$ as the most plausible Markov order, in contrast to the $n = 0$ obtained with the stationary model. Therefore, identifying change points allows us not only to uncover patterns at longer time scales, but the separation into temporal segments enables the identification of statistically significant patterns at short time scales as well, which would otherwise remain obscured with the stationary model — even though it is designed to capture only these kinds of correlations.

The improved quality of this model is also evident when we investigate the epidemic dynamics, as shown in Fig. 4. In order to obtain an estimate of the number of infected based on the model, we generated different sequences of edges using the fitted segments and transition probabilities $\hat{p}^l_{x,\boldsymbol{x}} = a^l_{x,\boldsymbol{x}}/k^l_{\boldsymbol{x}}$ in each of the segments estimated with Markov orders going from 0 to 3. We simulated SIR and SIS processes on top of the networks generated and averaged the number of infected over many instances. Looking at Fig. 4, we see that the inferred positions of the change-points tend to coincide with the abrupt changes in infection rates, which show very good agreement between the empirical and generated time-series. For higher Markov order, the agreement improves, although the improvement seen for $n > 1$ is probably due to overfitting, given the results of Fig. 3. We note also that the fact that $n = 0$ provides the worse fit and agreement with epidemic dynamics shows that it is not only the existence of change points, but also the inferred Markov dynamics that contribute to the quality of the model in reproducing the epidemic spreading.

In order to examine the link between the structure of the network and the change points, we fitted a layered hierarchical degree-corrected stochastic block model[16,33] to the data, considering each segment as a separate edge layer. From the figure Fig. 4) we can see that the density of connections between node groups vary in a substantial manner, suggesting that change point marks an abrupt transition in the typical kind of encounters between students — representing breaks between classes, meal time, etc (see Fig. 4). This yields an insight as to why these changes in pattern may slow down or speed up an epidemic spreading: if students are confined to their classrooms, contagion across classrooms is inhibited, but as soon they are free to move around the school grounds, so can the epidemic.

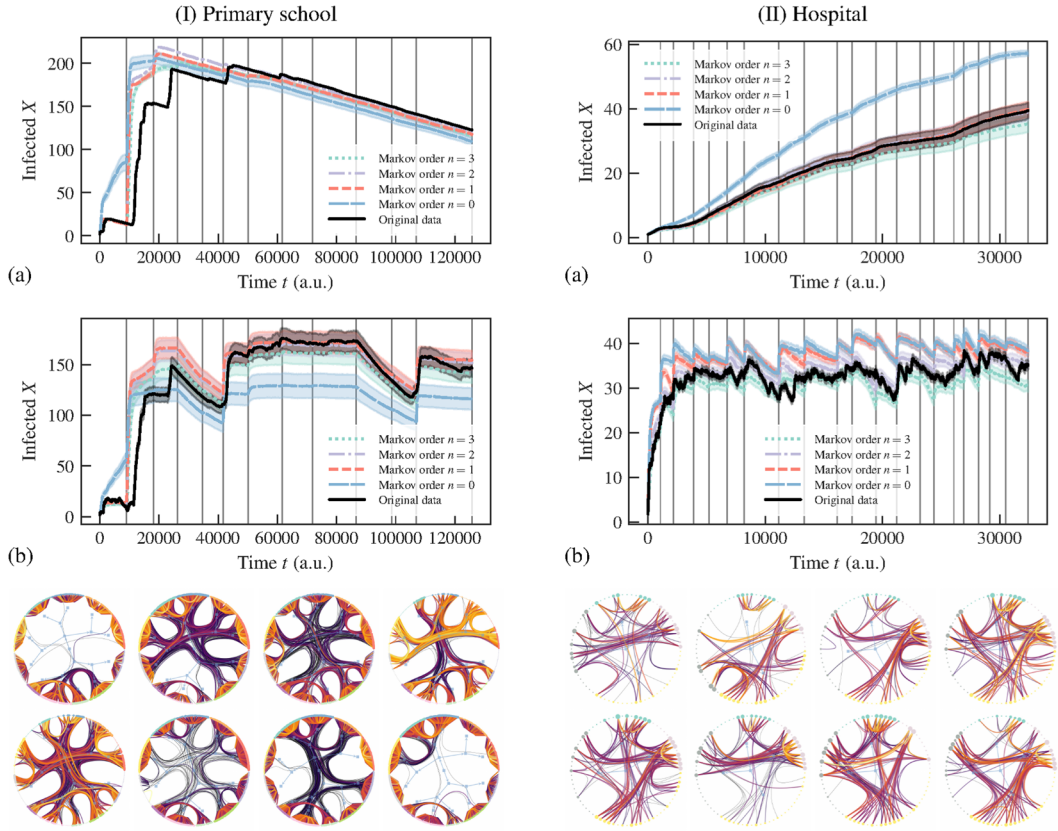

Figure 6. (Above) Number of infected nodes over time $X(t)$ for temporal networks between (I) students in a primary school[24] ($N=242$) and (II) patients and staff of a hospital[25] ($N=75$), considering both the original data and artificial time-series generated from the fitted nonstationary Markov model of a given order n, using (**a**) SIR [(I) $\beta=0.9$, $\gamma=0.001$; (II) $\beta=0.001$, $\gamma=0$] and (**b**) SIS [(I) $\beta=0.84$, $\gamma=0.01$; (II) $\beta=0.81$, $\gamma=0.015$] epidemic models. The vertical lines mark the position of the inferred change points. In all cases, the values were averaged over 100 independent realizations of the network model (for the artificial datasets) and dynamics. The shaded areas indicate the standard deviation of the mean. (Below) Network structure inside the first eight segments, as captured by a layered hierarchical degree-corrected stochastic block model[16] using the frequency of interactions as edge covariates[33] (indicated by colors), where each segment is considered as a different layer. The values of the infection and recovery rates were chosen so that the spreading dynamics spans the entire time range of the dataset.

We explore further the match between data and model by measuring the distribution of waiting times between temporal edges, i.e. the time interval between the occurrence in the time series of the same edge in the network, shown in Fig. 5 for both Markov models. For the empirical dataset, the waiting time distribution shows a characteristic peak at short times, and a broad decay for longer ones. For the stationary model, the distributions obtained with the fitted models show significant discrepancy — for both long and short times — except when the Markov order is increased to $n=3$, which, according to our Bayesian analysis cannot be used as an explanation for the data, as it represents an overfit. However, for the nonstationary model with change points, we observe a fair agreement between data and model for the most-likely model with $n=1$, across all time scales. The nonstationary model also provides an explanation to the shape of the distribution at longer times, which shows a separation of time scales inside individual stationary segments, from larger ones across change points (marked as vertical line in Fig. 5). In addition to this, the fact that the $n=0$ model does not reproduce the short time behavior of the distribution shows that the Markov property inside each stationary segment is indeed a necessary ingredient of the model. The model that best fits the data is able to reproduce with a quite good degree of approximation the distribution of waiting times, across all time scales. This point is in agreement with previous results highlighting the importance of the heterogeneity of inter-event times for dynamical processes[34], but here we see how two different time scales are sufficient to reproduce a large fraction of the observed behavior.

In Sec. 3 we show that the same behavior is obtained for a variety of different datasets.

Other datasets

Here we show that very similar results to those described above are also encountered for other proximity datasets. In Fig. 6(I) we show the analysis for the temporal behavior of students in a primary school[24], which shows a very clear correlation of the change in infection rate and the inferred change points. If we inspect the network structure inside each temporal segment, we see that amounts to periods of time where the students are either confined into classes, or mingling in larger groups. A similar behavior is seen if Fig. 6(II) for people (staff and patients) in a hospital ward[25].

Discussion

In this work we presented a data-driven approach to model temporal networks that is based on the simultaneous description of the network dynamics in two time scales: 1. The occurrence of the edges according to an arbitrary-order Markov chain, 2. The abrupt transition of the Markov transition probabilities at specific change-points. We developed a Bayesian framework that allows the inference of the change points and Markov order from data in manner that prevents overfitting, and enables the selection of competing models.

We have applied our approach to a variety of empirical proximity networks, and we have evaluated the inferred models based on their capacity to reproduce the epidemic spreading observed with the original data. We have seen that the nonstationary model accurately reproduces the highly-variable nature of the infection rate, with changes correlating strongly with the inferred change points. Furthermore, we showed that the inferred model also accurately reproduces the waiting time statistics in the empirical data, both at small and large time scales, neither of which are accurately captured if the different time scales are analyzed in isolation.

We argue that, ultimately, the incorporation of such temporal heterogeneity is indispensable for the evaluation of the speeding up or slowing down of processes taking place on dynamic networks[12,35], and the development of mitigating strategies against epidemics[27].

Although our model successfully captures key properties of real dynamic networks, it can still be made more realistic in a variety of ways. For instance, it can be extended to continuous time via the incorporation of waiting time distributions between events, as done in ref.[13]. Furthermore, it remains also to be seen how the approach presented here can be extended to scenarios where edges are allowed both to appear and disappear from the network, so that its dynamics can no longer be represented simply by a sequence of edges. And lastly, it would be desirable to provide a more direct connection between the edge probabilities and change points with large-scale network descriptors, such as community structure.

Data Availability

The datasets generated during analysed during the current study are available in the sociopatterns website, at http://www.sociopatterns.org.

References

1. Holme, P. & Saramäki, J. Temporal networks. *Phys. Reports* **519**, 97–125, https://doi.org/10.1016/j.physrep.2012.03.001 (2012).
2. Holme, P. Modern temporal network theory: a colloquium. *The Eur. Phys. J. B* **88**, 234 (2015).
3. Ho, Q., Song, L. & Xing, E. P. Evolving cluster mixed-membership blockmodel for time-varying networks. *J. Mach. Learn. Res.: Work. Conf. Proc*. 342–350 (2011).
4. Perra, N., Gonçalves, B., Pastor-Satorras, R. & Vespignani, A. Activity driven modeling of time varying networks. *Sci. reports* 2 (2012).
5. Rocha, L. E. C., Liljeros, F. & Holme, P. Simulated Epidemics in an Empirical Spatiotemporal Network of 50,185 Sexual Contacts. *PLOS Comput. Biol.* **7**, e1001109 (2011).
6. Valdano, E., Ferreri, L., Poletto, C. & Colizza, V. Analytical Computation of the Epidemic Threshold on Temporal Networks. *Phys. Rev. X* **5**, 021005 (2015).
7. Génois, M., Vestergaard, C. L., Cattuto, C. & Barrat, A. Compensating for population sampling in simulations of epidemic spread on temporal contact networks. *Nat. Commun*. **6** (2015).
8. Ren, G. & Wang, X. Epidemic spreading in time-varying community networks. Chaos: An Interdiscip. *J. Nonlinear Sci.* **24**, 023116 (2014).
9. Karsai, M. *et al*. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E* **83**, 025102 (2011).
10. Gauvin, L., Panisson, A., Cattuto, C. & Barrat, A. Activity clocks: spreading dynamics on temporal networks of human contact. *Sci. reports* **3** (2013).
11. Vestergaard, C. L., Génois, M. & Barrat, A. How memory generates heterogeneous dynamics in temporal networks. *Phys. Rev. E* **90**, 042805 (2014).
12. Scholtes, I. *et al*. Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nat. Commun*. **5** (2014).
13. Peixoto, T. P. & Rosvall, M. Modelling sequences and temporal networks with dynamic community structures. *Nat. Commun*. **8**, 582 (2017).
14. Xu, K. S. & Iii, A. O. H. Dynamic Stochastic Blockmodels: Statistical Models for Time-Evolving Networks. In Greenberg, A. M., Kennedy, W. G. & Bos, N. D. (eds) *Social Computing, Behavioral-Cultural Modeling and Prediction*, no. 7812 in Lecture Notes in Computer Science, 201–210 (Springer Berlin Heidelberg, 2013).
15. Gauvin, L., Panisson, A. & Cattuto, C. Detecting the Community Structure and Activity Patterns of Temporal Networks: A Non-Negative Tensor Factorization Approach. *PLoS ONE* **9**, e86028 (2014).
16. Peixoto, T. P. Inferring the mesoscale structure of layered, edge-valued, and time-varying networks. *Phys. Rev. E* **92**, 042807 (2015).
17. Stanley, N., Shai, S., Taylor, D. & Mucha, P. J. Clustering Network Layers with the Strata Multilayer Stochastic Block Model. *IEEE Transactions on Netw. Sci. Eng.* **3**, 95–105 (2016).
18. Ghasemian, A., Zhang, P., Clauset, A., Moore, C. & Peel, L. Detectability Thresholds and Optimal Algorithms for Community Structure in Dynamic Networks. *Phys. Rev. X* **6**, 031005 (2016).
19. Zhang, X., Moore, C. & Newman, M. E. J. Random graph models for dynamic networks. *The Eur. Phys. J. B* **90**, 200 (2017).
20. Peel, L. & Clauset, A. Detecting Change Points in the Large-Scale Structure of Evolving Networks. In *Twenty-Ninth AAAI Conference on Artificial Intelligence* (2015).
21. De Ridder, S., Vandermarliere, B. & Ryckebusch, J. Detection and localization of change points in temporal networks with the aid of stochastic block models. *J. Stat. Mech. Theory Exp.* **2016**, 113302 (2016).
22. Corneli, M., Latouche, P. & Rossi, F. Multiple change points detection and clustering in dynamic network. *Stat. Comput* (2017).
23. Toroczkai, Z. & Guclu, H. Proximity networks and epidemics. *Phys. A: Stat. Mech. its Appl.* **378**, 68–75 (2007).
24. Stehlé, J. *et al*. High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School. *PLOS ONE* **6**, e23176 (2011).
25. Vanhems, P. *et al*. Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors. *PLoS ONE* **8**, e73970 (2013).
26. 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**, e0136497 (2015).
27. Gemmetto, V., Barrat, A. & Cattuto, C. Mitigation of infectious disease at school: targeted class closure vs school closure. *BMC Infect. Dis.* **14**, 695 (2014).

28. Strelioff, C. C., Crutchfield, J. P. & Hübler, A. W. Inferring Markov chains: Bayesian estimation, model comparison, entropy rate, and out-of-class modeling. *Phys. Rev. E* **76**, 011106 (2007).
29. Polansky, A. M. Detecting change-points in Markov chains. *Comput. Stat. & Data Analysis* **51**, 6013–6026 (2007).
30. Arnesen, P., Holsclaw, T. & Smyth, P. Bayesian Detection of Changepoints in Finite-State Markov Chains for Multiple Sequences. *Technometrics* **58**, 205–213 (2016).
31. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *The J. Chem. Phys.* **21**, 1087 (1953).
32. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biom.* **57**, 97–109 (1970).
33. Peixoto, T. P. Nonparametric Bayesian inference of the microcanonical stochastic block model. *Phys. Rev. E* **95**, 012317 (2017).
34. Karsai, M., Jo, H.-H. & Kaski, K. Bursty human dynamics (2017).
35. Masuda, N., Klemm, K. & Eguíluz, V. M. Temporal Networks: Slowing Down Diffusion by Long Lasting Interactions. *Phys. Rev. Lett.* **111**, 188701 (2013).
36. Fournet, J. & Barrat, A. Contact Patterns among High School Students. *PLoS ONE* **9**, e107878 (2014).

Acknowledgements

The authors thank Ciro Cattuto, André Panisson and Anna Sapienza for useful conversations. This research was supported by the Lagrange Project of the ISI Foundation funded by the CRT Foundation. The funding bodies had no role in study design, data collection and analysis, preparation of the manuscript, or the decision to publish.

Author Contributions

T.P.P. and L.G. conceived and conducted the research, and analysed the results. All authors reviewed the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018