

PAPER • OPEN ACCESS

## How contact patterns destabilize and modulate epidemic outbreaks

To cite this article: Johannes Zierenberg *et al* 2023 *New J. Phys.* **25** 053033

View the [article online](#) for updates and enhancements.

You may also like

- [The Euler characteristic as a topological marker for outbreaks in vector-borne disease](#)  
Danillo Barros de Souza, Everlon Figueirôa dos Santos and Fernando A N Santos
- [Suppression of epidemic outbreaks with heavy-tailed contact dynamics](#)  
Byungjoon Min, K.-I. Goh and I.-M. Kim
- [Impact of temporal scales and recurrent mobility patterns on the unfolding of epidemics](#)  
David Soriano-Paños, Gourab Ghoshal, Alex Arenas et al.



## PAPER

# How contact patterns destabilize and modulate epidemic outbreaks

## OPEN ACCESS

## RECEIVED

17 February 2023

## REVISED

20 April 2023

## ACCEPTED FOR PUBLICATION

2 May 2023







## PUBLISHED

30 May 2023

Original Content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.



Johannes Zierenberg<sup>1,5,\*</sup> , F Paul Spitzner<sup>1,5</sup> , Jonas Dehning<sup>1</sup> , Viola Priesemann<sup>1,2</sup> ,  
Martin Weigel<sup>3</sup>  and Michael Wilczek<sup>1,4</sup> 

<sup>1</sup> Max Planck Institute for Dynamics and Self-Organization, 37077 Göttingen, Germany

<sup>2</sup> Institute for the Dynamics of Complex Systems, University of Göttingen, 37077 Göttingen, Germany

<sup>3</sup> Institut für Physik, Technische Universität Chemnitz, 09107 Chemnitz, Germany

<sup>4</sup> Theoretical Physics I, University of Bayreuth, 95440 Bayreuth, Germany

<sup>5</sup> J Z and F P S contributed equally.

\* Author to whom any correspondence should be addressed.

E-mail: [johannes.zierenberg@ds.mpg.de](mailto:johannes.zierenberg@ds.mpg.de)

**Keywords:** human contact patterns, non-Markovian dynamics, epidemic outbreak, latent period

Supplementary material for this article is available [online](#)

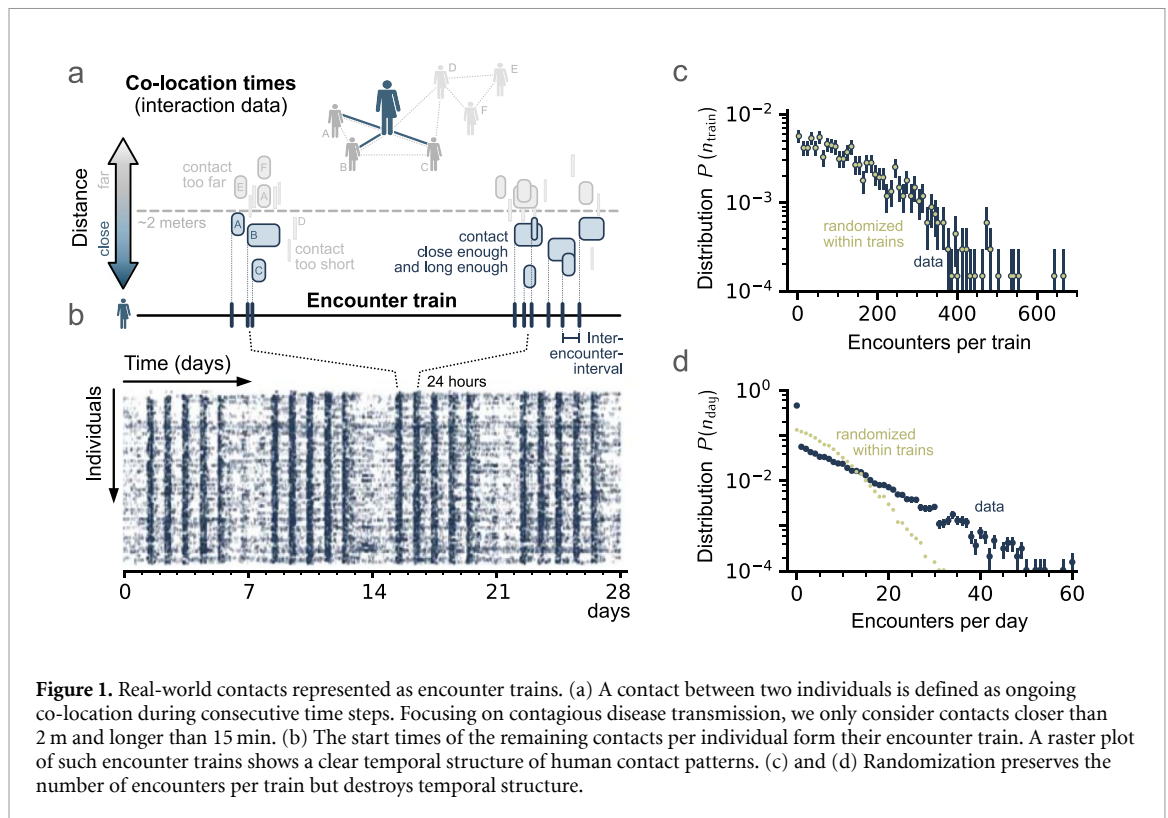
## Abstract

The spread of a contagious disease clearly depends on when infected individuals come into contact with susceptible ones. Such effects, however, have remained largely unexplored in the study of epidemic outbreaks. In particular, it remains unclear how the timing of contacts interacts with the latent and infectious stages of the disease. Here, we use real-world physical proximity data to study this interaction and find that the temporal statistics of actual human contact patterns (i) destabilize epidemic outbreaks and (ii) modulate the basic reproduction number  $R_0$ . We explain both observations by distinct aspects of the observed contact patterns. On the one hand, we find the destabilization of outbreaks to be caused by the temporal clustering of contacts leading to over-dispersed offspring distributions and increased probabilities of otherwise rare events (zero- and super-spreading). Notably, our analysis enables us to disentangle previously elusive sources of over-dispersion in empirical offspring distributions. On the other hand, we find the modulation of  $R_0$  to be caused by a periodically varying contact rate. Both mechanisms are a direct consequence of the memory in contact behavior, and we showcase a generative process that reproduces these non-Markovian statistics. Our results point to the importance of including non-Markovian contact timings into studies of epidemic outbreaks.

## 1. Introduction

As contagious diseases are passed on through contacts, the number of secondary infections depends crucially on the contact patterns of infectious individuals. These contact patterns encode relevant information such as the number of interaction partners and contact timing. However, the majority of prevailing models for disease spread prioritize simpler descriptions that neglect these aspects—despite evidence from studies that show the effects of contact patterns to be crucial for disease spread: structurally, when interaction partners are modeled by a static complex network [1], the network structure affects disease spread through the occurrence of hubs [2, 3], multiscale link communities [4] and influential spreaders [5]. Dynamically, real-world interaction times generally follow a non-Markovian process (in contrast to commonly assumed memoryless processes), which influences epidemics through the occurrence of bursts [6, 7] and daily and weekly variations in human interaction [8, 9].

Thus, for a better understanding of disease spread through human contacts, a complete description of time-varying interactions in the form of so-called temporal networks [10, 11] seems necessary. However, constructing detailed temporal networks from real-world contacts requires extensive amounts of recorded data, which in principle can be collected in field studies [12–14] but such data are notoriously limited in



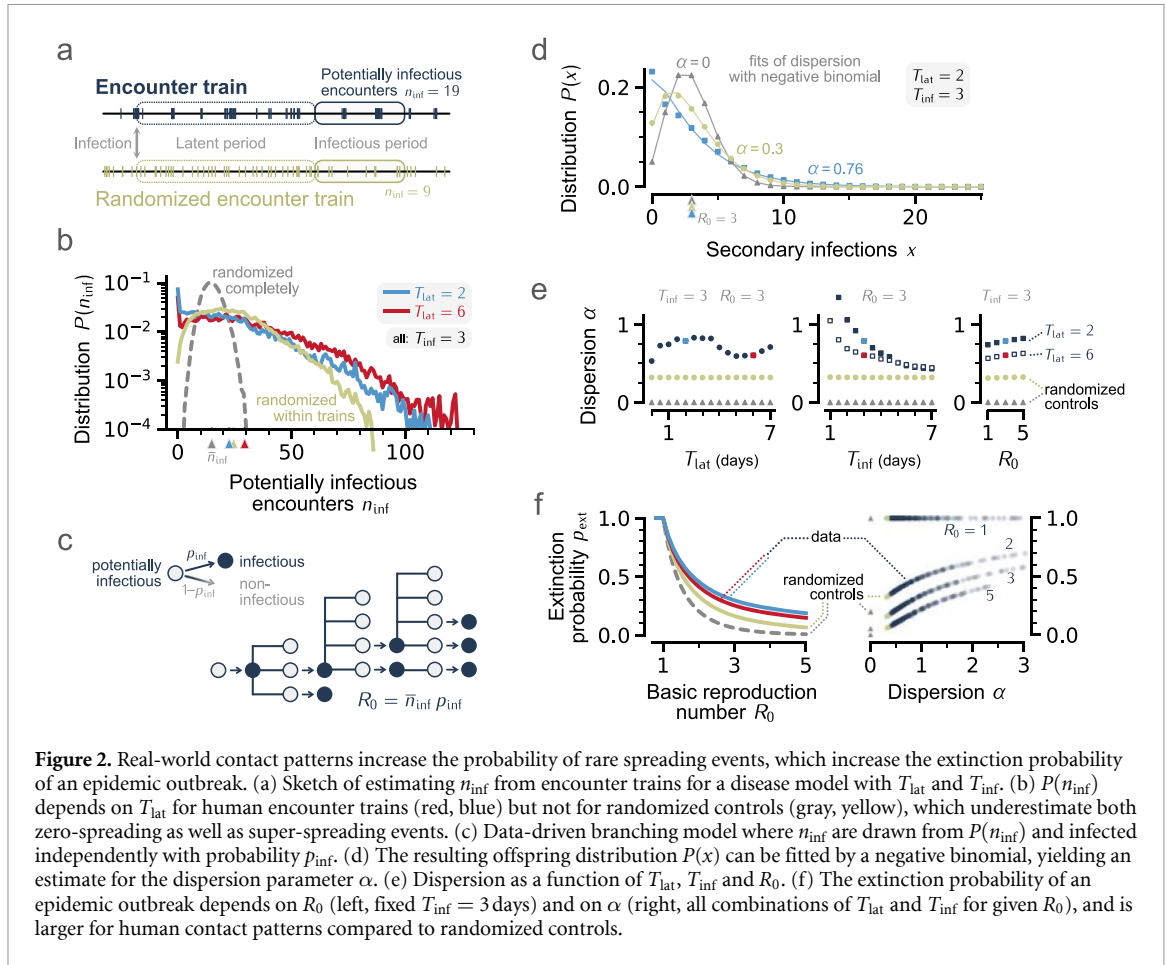
duration and system size. Although such limitations of real-world data can be partly remedied by generating surrogate data [15, 16], it is often unclear to which extent they represent the real system. An unsolved task is thus to generate surrogate data that mimics temporal statistics and individual variations of actual human contact data.

Here, we address this gap by identifying and isolating features of contact behavior that affect epidemic outbreaks using a novel analysis of real-world contact data. Instead of characterizing full epidemic outbreaks on a large (likely under-determined) temporal network, we develop an effective description through *potentially infectious encounters* that propagates statistics of contacts to statistics of disease spread. This approach avoids treating microscopic (non-linear) network effects [17–19] and allows us to focus on how contact patterns statistically affect epidemic outbreaks. Our analysis reveals two main mechanisms: (i) contact clustering destabilizes outbreaks by increasing the dispersion of offspring distributions and the probability of zero-spreading events, and (ii) temporal variation of the contact rate modulates the mean basic reproduction number,  $R_0$ , due to an interference between contact patterns and disease progression. Finally, we showcase a non-Markovian process that faithfully reproduces the temporal statistics and their effect on disease spread as a proof of principle for a new class of generative models for surrogate data that mimic human contact patterns.

## 2. Results

To shed light on the interplay of contact patterns and epidemic outbreaks, we analyse proximity data from the Copenhagen Networks Study [20] and, in the supplementary material, from SocioPatterns [13, 21]. We filter each individual's contacts by distance and duration, and define encounters as their starting times (see methods). The resulting *encounter trains* are a point-process-like representation that captures the non-Markovian statistics of the underlying contact patterns (figure 1). The importance of these non-Markovian statistics can be seen when comparing to randomized encounter trains. In these surrogate data, encounter times are uniformly reassigned within the duration of the experiment—which preserves the number of encounters per train, i.e. the inter-individual variability (figure 1(c)), but destroys any temporal structure (figure 1(d)).

In order to quantify the effect of contact patterns on epidemic outbreaks, we focus on potential secondary infections and assume a contagious disease that can be transmitted only during infectious encounters. Further, if  $\tau$  is the time elapsed since infection, infected individuals undergo a (non-infectious) latent period  $\tau \in [0, T_{\text{lat}})$  and an infectious period  $\tau \in [T_{\text{lat}}, T_{\text{lat}} + T_{\text{inf}})$  during which all  $n_{\text{inf}}$  encounters are



**Figure 2.** Real-world contact patterns increase the probability of rare spreading events, which increase the extinction probability of an epidemic outbreak. (a) Sketch of estimating  $n_{inf}$  from encounter trains for a disease model with  $T_{lat}$  and  $T_{inf}$ . (b)  $P(n_{inf})$  depends on  $T_{lat}$  for human encounter trains (red, blue) but not for randomized controls (gray, yellow), which underestimate both zero-spreading as well as super-spreading events. (c) Data-driven branching model where  $n_{inf}$  are drawn from  $P(n_{inf})$  and infected independently with probability  $p_{inf}$ . (d) The resulting offspring distribution  $P(x)$  can be fitted by a negative binomial, yielding an estimate for the dispersion parameter  $\alpha$ . (e) Dispersion as a function of  $T_{lat}$ ,  $T_{inf}$  and  $R_0$ . (f) The extinction probability of an epidemic outbreak depends on  $R_0$  (left, fixed  $T_{inf} = 3$  days) and on  $\alpha$  (right, all combinations of  $T_{lat}$  and  $T_{inf}$  for given  $R_0$ ), and is larger for human contact patterns compared to randomized controls.

*potentially infectious*. We estimate  $n_{inf}$  by considering every encounter in the data set as a potential start for an infection (figure 2(a), see methods). As we show in figure 2(b), empirical contact patterns increase both the probability of very few  $n_{inf}$  (related to zero-spreading events) as well as very many  $n_{inf}$  (related to super-spreading events) when compared to randomized controls. This increase in variability influences whether a single infection results in an epidemic outbreak or not.

**2.1. Human contact patterns destabilize epidemic outbreaks**

To demonstrate the effect of empirical contact patterns on epidemic outbreaks, we map the probability distribution of  $n_{inf}$  to an offspring distribution using a two-step, data-driven branching process (figure 2(c)): Each infected individual first generates encounters according to the empirical distribution  $P(n_{inf})$ , and then infects each of them independently with probability  $p_{inf}$  resulting in binomial-distributed secondary infections. Taking the expectation value yields the offspring distribution

$$P(x) = \sum_{n_{inf}=x}^{\infty} P(n_{inf}) \binom{n_{inf}}{x} p_{inf}^x (1 - p_{inf})^{n_{inf}-x}. \tag{1}$$

Similar to empirical distributions from contact tracing [22],  $P(x)$  can be well described by a negative binomial distribution (figure 2(d)) with mean  $\bar{x} = R_0 = p_{inf} \bar{n}_{inf}$  and variance  $(x - R_0)^2 = R_0 + \alpha R_0^2$ , where  $\alpha$  is the dispersion parameter that characterizes the increase in variance relative to a Poisson distribution.

Our data analysis provides a systematic approach to identifying sources of the dispersion observed in empirical offspring distributions [23–25]. We analyze step by step the dispersion occurring because of human contact patterns, and how it depends on  $T_{lat}$ ,  $T_{inf}$  and  $R_0$  (figure 2(e), left to right). For a completely randomized control, where encounters are uniformly distributed across trains and time, we consistently find Poissonian offspring distributions (figure 2(d), gray) with vanishing dispersion ( $\alpha = 0$ ), independent of the three disease parameters (figure 2(e), gray triangles). When including variability of contact rates into the control, while still randomizing within trains, the dispersion of the offspring distribution increases ( $\alpha \approx 0.3$ ) but remains mostly independent of disease parameters (yellow circles). Lastly, when also including the precise timing of human contact patterns, offspring distributions show a large dispersion that depends on

$T_{\text{lat}}$  and  $T_{\text{inf}}$  (blue symbols). In particular, dispersion is strongest for short  $T_{\text{inf}}$  but decays as  $T_{\text{inf}}$  increases. Hence, part of the empirical dispersion can be attributed to variability of contact rates between individuals, but the non-Markovian timing of human contact patterns causes a further increase—for realistic parameters roughly by a factor of two.

From  $P(x)$  we derive the extinction probability  $p_{\text{ext}}$ , defined as the fraction of outbreaks that asymptotically end up in the absorbing state of zero infections (figure 2(f)). It can be calculated using the probability generating function,  $\pi(\theta) = \sum_{x=0}^{\infty} P(x)\theta^x$ , as the smallest  $\theta^*$  that solves  $\theta^* = \pi(\theta^*)$  [26]. In addition to the anticipated monotonic decrease of  $p_{\text{ext}} = \theta^*$  with increasing  $R_0$ , we find that extinction is more likely for actual human contact patterns (red, blue) than for randomized controls (gray, yellow). Moreover, for fixed  $R_0$ , we find that an increased dispersion  $\alpha$  due to human contact patterns non-linearly increases  $p_{\text{ext}}$  (figure 2(f), right).

Summarizing, we find that the non-Markovian timing of human contact patterns can be a strong source of variability, relevant to explaining the over-dispersion of empirical offspring distributions. In particular, increasing the dispersion for a fixed  $R_0$  increases the probability of zero-spreading events (figure 2(d), blue vs gray), and results in a higher extinction probability (figure 2(f)) — in other words, the non-Markovian temporal structure of human contact patterns destabilizes epidemic outbreaks.

## 2.2. Interplay between contact pattern and disease progression modulates basic reproduction number

As highlighted in figure 2(b),  $\bar{n}_{\text{inf}}$  depends on  $T_{\text{lat}}$  for human encounter trains. This might be at first glance surprising, because for memory-less processes,  $\bar{n}_{\text{inf}}$  is proportional to  $T_{\text{inf}}$  but independent of  $T_{\text{lat}}$ . Hence, in the following, we systematically vary  $T_{\text{lat}}$  and  $T_{\text{inf}}$  to study how the interplay between human contact patterns and disease progression affects  $\bar{n}_{\text{inf}}$  (figure 3).

Considering a fixed  $T_{\text{inf}} = 3$  days (figure 3(a)) and scanning  $T_{\text{lat}}$  leads to a periodic modulation of  $\bar{n}_{\text{inf}}$  from human encounter trains (black, dashed) around the constant estimate from randomized trains (yellow). Thus, we consider  $\bar{n}_{\text{inf}}$  relative to randomized (figure 3(b)), which accounts for the trivial increase of  $\bar{n}_{\text{inf}}$  with increasing  $T_{\text{inf}}$ . For small  $T_{\text{inf}} < 1$  day, we find daily modulations as a function of  $T_{\text{lat}}$ , with regions below-randomized (blue) and above-randomized (red). This effect diminishes for larger  $T_{\text{inf}}$ , where we find extended, triangular regions with interfaces located at  $T_{\text{lat}} + T_{\text{inf}} = 7$  days and  $T_{\text{lat}} = 7$  days. We thus find periodic modulations of  $\bar{n}_{\text{inf}}$  on the scale of days (small  $T_{\text{lat}}$ ) and weeks (large  $T_{\text{lat}}$ ).

In the following, we uncover the origin of these periodic modulations using what we call *conditional encounter rate*  $\Psi(\tau)$ , see methods. In short,  $\Psi(\tau)$  describes the average rate of encounters conditioned on an initial encounter (figure 3(c)). Considering the initial encounter as an infection,  $\Psi(\tau)$  measures the rate of potentially infectious encounters but neglects variability and dispersion. We find that  $\Psi(\tau)$  features a peak at 0 (which implies strong clustering [6]) and the anticipated periodic modulations between high and low encounter rates (which cause a time-dependent secondary attack rate). Both, the initial peak and modulations are again lost for randomized controls (yellow line).

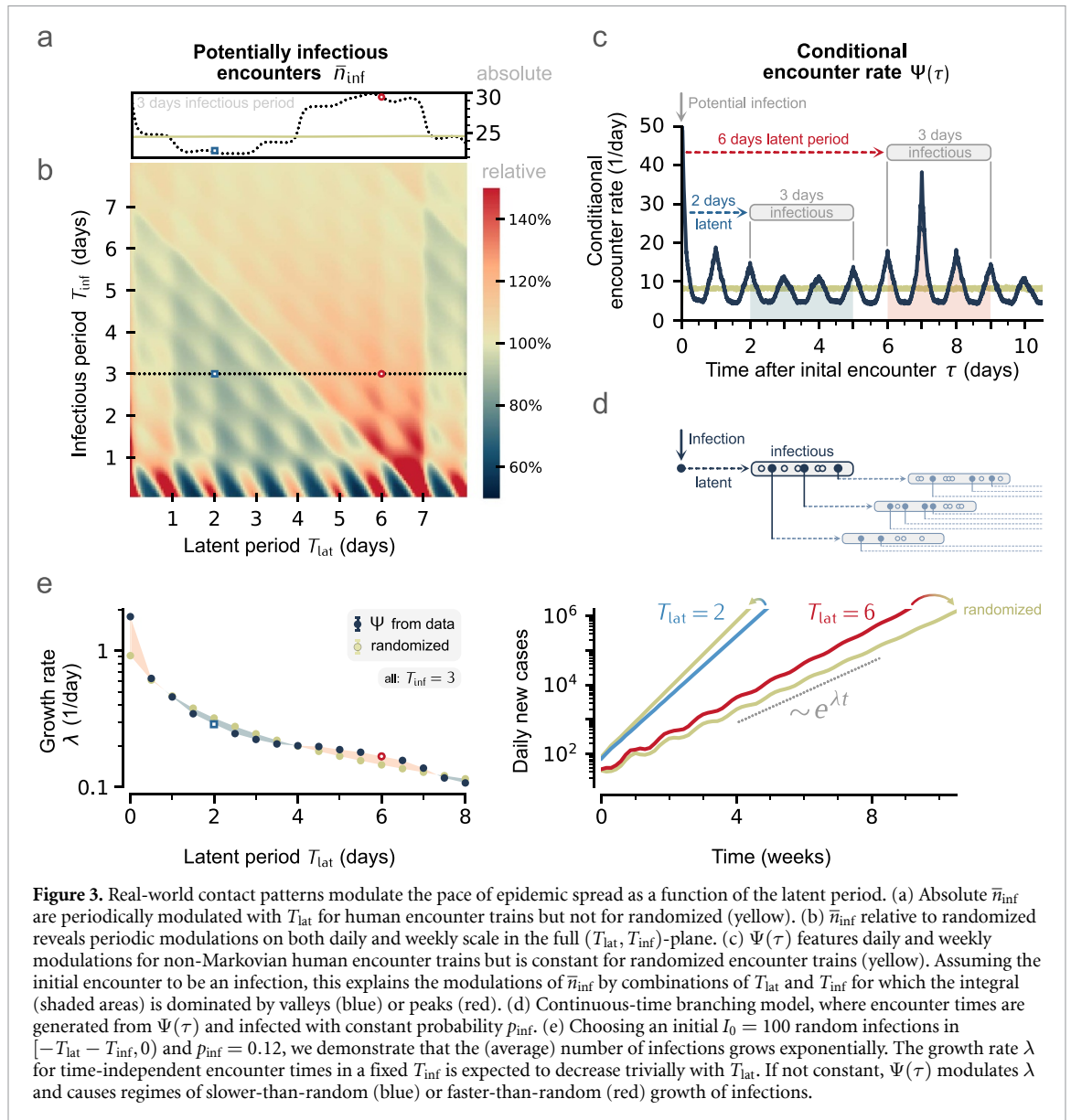
Note that we can directly obtain an estimate of  $\bar{n}_{\text{inf}}$  for a particular disease progression ( $T_{\text{lat}}, T_{\text{inf}}$ ) by integrating  $\Psi(\tau)$  over the infectious period (figure 3(c), shaded areas). Reconsidering the previous examples, this explains the lower  $\bar{n}_{\text{inf}}$  for  $T_{\text{lat}} = 2$  days (blue area covering the valley) and the larger  $\bar{n}_{\text{inf}}$  for  $T_{\text{lat}} = 6$  days (red area covering the 7 day peak). The examples showcase that  $\bar{n}_{\text{inf}}$  is determined by the alignment of the infectious period with regions of low or high  $\Psi(\tau)$ .

Consequently, since  $\bar{n}_{\text{inf}}$  is related to  $R_0$ , the interplay between contact patterns and disease progression modulates the pace of epidemic spread. To illustrate this, we construct a continuous-time branching process, where each exposed individual generates encounters according to  $\Psi(\tau)$ . During the infectious period, encounters again have a probability  $p_{\text{inf}}$  to become infected (figure 3(d)). Assuming an outbreak that survived the initial generations, we prepare the system with 100 random initial infections in the interval  $[-T_{\text{lat}} - T_{\text{inf}}, 0)$ . The resulting time evolution of daily new cases shows clear exponential growth, where the growth rate  $\lambda$  trivially decreases with the generation time and, thus,  $T_{\text{lat}}$  (figure 3(e)). However, this expected decrease of  $\lambda$  for memoryless encounter timings (yellow) is modulated in the model due to variations in  $\Psi(\tau)$ , which results in slower-than-random (blue) or faster-than-random (red) growth.

Summarizing, human contact patterns cause a dependence of  $\bar{n}_{\text{inf}}$  on  $T_{\text{lat}}$  that modulates  $R_0$  and thereby the growth rate of an epidemic outbreak.

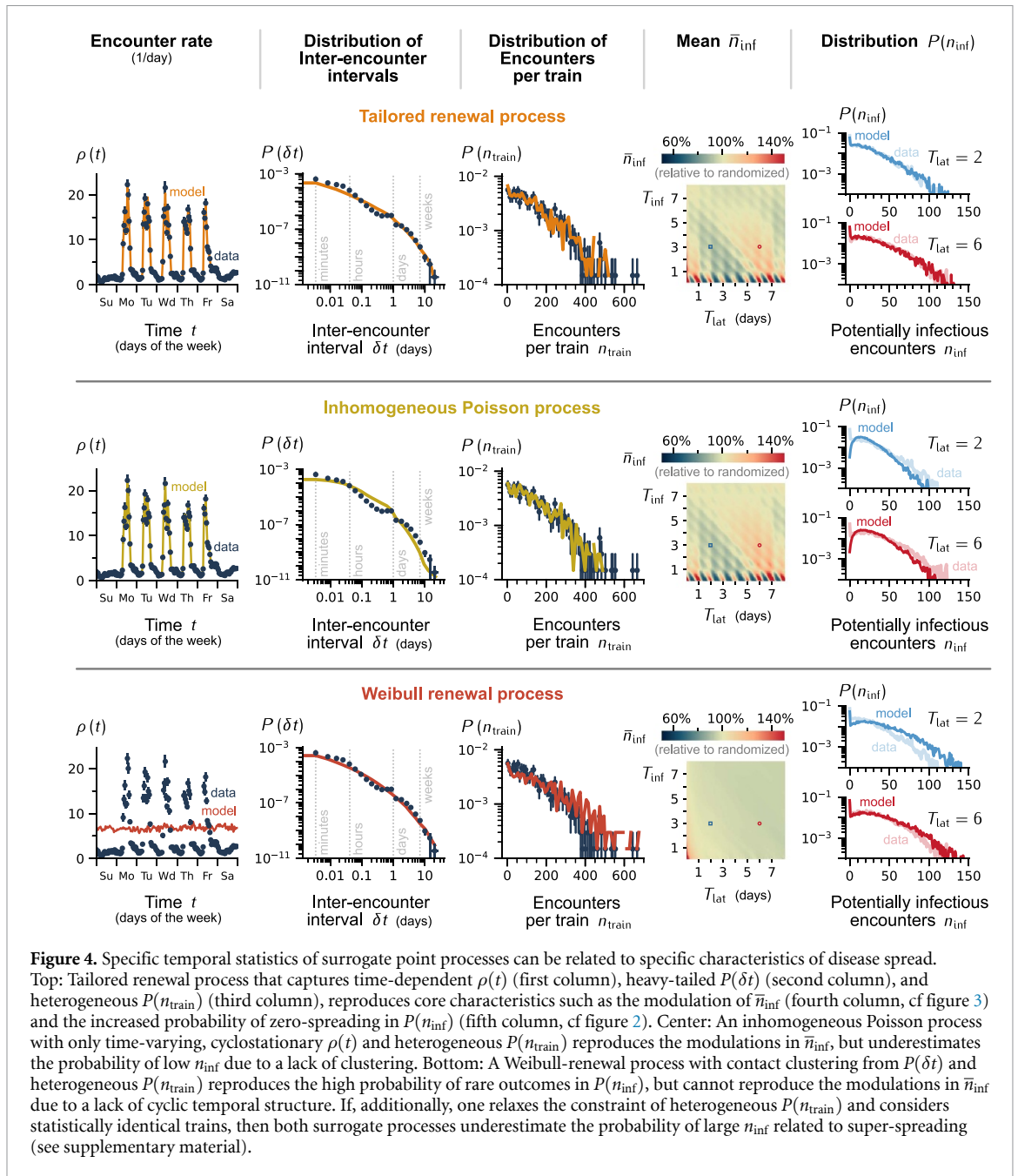
## 2.3. Destabilization and modulation of epidemic spread can be attributed to specific temporal statistics of contact patterns

After illustrating that non-Markovian statistics can destabilize and modulate epidemic outbreaks, it seems natural to ask how they can be included in models of disease spread. In such models it is common to approximate encounter times between individuals as memoryless (Poisson) processes [1]. Assuming independence, these processes can be merged to result in encounter trains with Poisson statistics—the same



statistics as our randomized encounter trains. In the following, we construct encounter trains with non-Markovian statistics and identify three specific features of contact patterns that are necessary to reproduce the relevant statistics of encounters. As a proof of principle, we showcase a novel *tailored renewal process* that is constrained by data and reproduces all salient features (figure 4, top row):

- i) Focusing on temporal statistics, the encounter rate  $\rho(t)$  averaged across individuals and weeks is time-dependent but cyclostationary;  $\rho(t)$  repeats in a weekly cycle with differences between day and night, and between weekdays and weekends (figure 4, first column). This can be captured by an inhomogeneous Poisson process (figure 4, middle row), which reproduces the periodic modulation of  $\bar{n}_{inf}$  (fourth column) and  $\Psi(\tau)$  (see supplemental figure S3).
- ii) The distribution of inter-encounter intervals  $P(\delta t)$  has high probability for small  $\delta t$  and a heavy tail of non-vanishing probability for large  $\delta t$  (second column). Because this tail corresponds to long periods without any encounter, it causes the high probability of  $n_{inf} \approx 0$  (last column) that strongly contributes to the destabilization of epidemic outbreaks.  $P(\delta t)$  is dominantly shaped by the clustering of human contacts and can be well approximated by a Weibull distribution [7, 27]. Accordingly, a Weibull-renewal processes (last row) reproduces  $P(\delta t)$  and  $P(n_{inf})$  well, but it does not have a time-varying  $\rho(t)$  and cannot reproduce the period modulations of  $\Psi(\tau)$  and  $\bar{n}_{inf}$ .
- iii) Encounter rates vary between individuals (third column). This variability can be attributed to intrinsic differences in contact behavior (cf figure 2, gray vs yellow) and is partly captured by the degree distribution of the contact network [28]. Recall that such across-individual variability is crucial to



reproduce the heavy tails of  $P(n_{inf})$  and offspring distributions (see also supplemental figure S3 for generative processes where individuals share a common rate).

Clearly, models of disease spread can benefit from a generative process that reproduces those relevant features of human contact patterns, such as the tailored Weibull-renewal process showcased here. However, although our process reproduces all discussed features, it is built heuristically, and future work is needed to construct microscopic models that give rise to cyclostationary rates with clustering in a principled way, while remaining mathematically tractable.

### 3. Discussion

We analyzed real-world human contact patterns and found that their non-Markovian timings shape epidemic spread in two important ways. Firstly, they increase the over-dispersion of offspring distributions, compared to random (Poisson) contact patterns, which (a) leads to more zero- and super-spreading events, and (b) decreases the probability of an epidemic outbreak from an initial infection. While clustering is typically associated with super-spreading events, it inevitably causes periods of low contact rate that increase the probability of zero-spreading events. The resulting increase in extinction probability (despite

super-spreading) is consistent with previous results, where individual variation of  $R_0$  captured the over-dispersion of empirical offspring distributions [23]. Still, the sources of this variation remained poorly understood, with candidates ranging from environmental factors (behavior, seasonality) to intrinsic ones (viral load, susceptibility) [29]. Here, we disentangled two sources based on contact patterns and identified heterogeneous contact rates and the non-Markovian timing of contacts as relevant factors for over-dispersion in disease transmissibility.

Secondly, human contact patterns non-trivially modulate the pace of epidemic spread depending on the latent period, which we attribute to time-dependent but cyclostationary encounter rates. A cyclostationary rate leads to periods of statistically high and low encounter rates conditioned on a potential infection. How these periods typically align with the infectious period is affected by the latent period and determines whether the number of potential secondary infections, and in turn  $R_0$ , increases or decreases. This modulation of  $R_0$  can thus be understood as a resonance following either a constructive or destructive interference between a periodically changing contact rate and the disease progression. This resonance is a new mechanism to explain the previously observed slow-down or speed-up of diffusion processes on temporal networks due to non-Markovian characteristics [30].

In the main manuscript we focus on deterministic disease progression with fixed periods ( $T_{\text{lat}}$ ,  $T_{\text{inf}}$ ), but we also considered non-deterministic disease progression with gamma-distributed periods [31–33]; the results are summarized in the supplementary material (figure S1). We find our main conclusion verified for non-deterministic disease progression: the probability of zero-spreading events is reliably higher for human contact patterns compared to randomized; however, the modulation of  $\bar{n}_{\text{inf}}$  with  $T_{\text{lat}}$  is smeared out with increasing variability in the period durations. Thus, in the unrealistic (but commonly adopted) limit of exponentially distributed periods, human contact patterns still reduce the robustness of outbreaks but no longer modulate the pace of epidemic spread.

To reproduce the relevant temporal features of human contact patterns, we introduced non-Markovian contact dynamics in the form of Weibull-distributed inter-encounter intervals (clustering) or inhomogeneous encounter rates (cyclostationarity). Previous studies of non-Markovian disease spread [7] found that clustering drastically affects the epidemic threshold for  $T_{\text{lat}} = 0$ , which is caused by the high frequency of small inter-encounter intervals [34] that, in our context, manifests as a near-zero peak in the conditional encounter rate (figure 3). Although it was shown that some non-Markovian models can be mapped onto effective Markov models [35, 36], our results suggest that the non-Markovian and cyclostationary features of human contact patterns make a similar general mapping elusive. This highlights the necessity for generative models that are non-Markovian, yet well understood and simple enough to find broad use in epidemic modeling and beyond.

Our work is a first step towards providing such models. We identified temporal statistics of real-world contact data that affect disease spread, and faithfully reproduced them with our tailored Weibull-renewal process. Thereby, our work provides an accessible pathway towards including non-Markovian statistics into spreading processes, in general, and paves the way to systematically study their non-equilibrium physics.

## 4. Methods

**Extracting contacts from real-world physical proximity data:** Consider data composed of a list of co-locations (physical proximity) described by the tuple (timestamp, user id A, user id B). We first sort the co-location times into unique lists for all id pairs (A, B) and (B, A). For each valid A, we then iterate over its list of (A, B) and merge co-location times that span consecutive time steps to construct pairwise contacts with starting time  $s$  and duration  $D$ . Combining these contacts yields a list of contacts  $\{(s_i, D_i)\}_A$  for each participant A.

From the lists of contacts, we construct a point-process-like representation for each participant that we call encounter train (see figure 1). Throughout the manuscript, an encounter refers to the starting time of a contact. The encounter train of participant A is the time-sorted list of all contact starting times  $s_i$  and can formally be written as

$$T(t) = \sum_i \delta(t - s_i). \quad (2)$$

The main data set from the Copenhagen Networks Study [12, 20] is based on Bluetooth signals between phones of individuals that participated in the study. The published data is a list of interactions described by the tuple (timestamp, A, B, RSSI), where user id B can be negative if the interaction is with a device outside of the study or an empty scan, and RSSI is the received (Bluetooth) signal strength indicator. The RSSI can be considered as a proxy for interaction distance, especially since all participants used the same device [37], with an RSSI  $\approx -80$  dBm corresponding to a distance of about  $2 \text{ m} \pm 0.5 \text{ m}$ . Since the data provides a maximal



RSSI per time window, we consider  $\text{RSSI} < -80$  dBm to indicate interactions to be further apart than 2 m throughout the full time window [37], and exclude them. Consequently, we filter the raw data to only include those interactions that are within the study (user id  $B \geq 0$ ) and have  $\text{RSSI} \geq -80$  dBm. See supplementary material for various controls. The data set covers a duration of  $t_{\max} = 28$  days, with a time step of 5 min, for 675 encounter trains.

**Average time-dependent encounter rate  $\rho(t)$ :** Because encounter trains are a point-process-like representation, we can define an encounter rate as the number of encounters in a window of size  $\Delta t$ . Assuming statistical independence between weeks and between participants, we determine the average time-dependent encounter rate  $\rho(t)$  by averaging the number of encounters in a time windows of size  $\Delta t = 1$  h throughout the week (i.e. first hour of a Sunday until last hour of a Saturday) across weeks of the experiment and across participants. Statistical errors are calculated on the level of participants using delete- $m$  jackknife error analysis.

**Inter-encounter interval  $\delta t$ :** To study temporal clustering and contact bursts, we measure the interval  $\delta t$  between consecutive encounter times. Since we are interested in the encounter statistics, each encounter has the same statistical weight independent of its encounter train origin. Consequently, the distribution  $P(\delta t)$  is simply the distribution over all observed intervals. To estimate statistical errors, we take into account that the number of encounters  $n_j$  differs between individual trains (hence also the number of inter-encounter intervals  $n_j - 1$ ), and evaluate statistical errors on the level of observed intervals using delete- $m_j$  jackknife error analysis with  $m_j = n_j - 1$ .

**Conditional encounter rate  $\Psi(\tau)$ :** To investigate how contact patterns interact with disease spread, we measure the encounter rate  $\Psi(\tau)$  upon a hypothetical infection from an encounter at  $\tau = 0$ . To construct  $\Psi(\tau)$ , we iterate over all encounters to measure the time-dependent encounter rate with temporal resolution of the experiment, starting from the encounter time, i.e.  $\tau = t - s_i = 0$ , until  $\tau = \tau_{\max}$  (we typically chose  $\tau_{\max} = 10$  days) or, if  $t_{\max} - s_i < \tau_{\max}$ , until  $\tau = t_{\max} - s_i$ . We then average over all these time-dependent encounter rates taking into account their different lengths. To estimate statistical errors, we take into account that the number of encounters  $n_j$  differs between individual trains by using delete- $m_j$  jackknife error analysis with  $m_j = n_j$ .

**Disease model:** We consider a disease that progresses in three discrete states upon infection: exposed-infectious-recovered. The duration  $T_{\text{lat}}$  within the exposed state is called latent period and the duration  $T_{\text{inf}}$  within the infectious state is called infectious period. For our main results, we consider the simple and intuitive case of a deterministic disease progression, where these periods are always the same. This corresponds to drawing the periods from delta distributions, which is quite different to commonly employed approximations that draw periods from exponential distributions (as expected for Poisson processes that describe many state transitions, from radioactive decay to chemical reactions). To confirm that our results also apply to non-deterministic disease progression, we repeated our analysis for the more realistic case of gamma-distributed periods [31–33] and obtained consistent results (supplemental material).

**Potentially infectious encounters  $n_{\text{inf}}$ :** To avoid assumptions on the probability of infection upon encounter, we introduce potentially infectious encounters as the number of encounters that occur during the infectious period of a hypothetical disease. For the deterministic disease progression, we can enumerate the statistics by iterating over all encounters of the data set. For each encounter  $s_j$ , we check whether the disease progression still fits into the experimental duration ( $s_j + T_{\text{lat}} + T_{\text{inf}} \leq t_{\max}$ ), and if true, estimate  $n_{\text{inf}}$  as the number of subsequent encounter  $s_j$  for which  $T_{\text{lat}} < s_j - s_i < T_{\text{lat}} + T_{\text{inf}}$ . For the non-deterministic disease progression, we need to sample disease realizations (see supplemental material). Statistical errors are calculated again on the level of encounters using the delete- $m_j$  jackknife analysis with  $m_j = n_j$ .

**Branching process with empirical distribution:** To estimate the survival probability from the empirical distribution of potentially infectious encounters,  $P(n_{\text{inf}})$ , we construct a discrete-time data-driven branching process (figure 2(c)). In a first step, each infection causes  $X \sim P(n_{\text{inf}})$  potentially infectious encounters. In a second step, each of these encounters can cause a secondary infection with probability  $p_{\text{inf}}$ , such that the number of secondary infections is binomial,  $Y \sim \mathcal{B}(X, p_{\text{inf}})$ . From  $Z_t$  infections in generation  $t$ , we thus obtain  $Z_{t+1} = \sum_{i=1}^{Z_t} Y_i$  infections in the next generation.

**Continuous-time branching process with inhomogeneous contacts:** To study the pace of epidemic spread, we construct a continuous-time branching process that captures the conditional encounter rates but neglects interactions between infected individuals. Here, each infected individual generates an independent encounter train starting from their initial infection time as an inhomogeneous Poisson process with a time-dependent rate given by  $\Psi(\tau)$  (figure 3(d)). Only those encounters that occur during the infectious period cause secondary infections with a chosen probability  $p_{\text{inf}}$ . Every secondary infection then generates a new encounter train and so on. For simplicity, we restrict our example to deterministic diseases with fixed latent and infectious periods.

**Point process models to approximate human contact patterns:** To disentangle the effect of distinct features of human contact patterns on the statistics of encounters, we constructed point-process models that captured (i) the distribution of rates across individuals, (ii), a time-dependent average encounter rate, and (iii), the distribution of inter-encounter intervals, or a combination thereof (see supplementary material for comparison of combinations).

To reproduce the inter-individual variability, we consider the same number of encounter trains as present in the data and weight each train  $i$  with their relative rates, i.e,  $w_i = n_{\text{train},i}/\langle n_{\text{train}} \rangle$ , where  $\langle \cdot \rangle$  is the average across trains.

To reproduce a time-dependent encounter rate  $\rho(t)$ , we employ thinning [38]: From a hidden process with rate  $\max_t[\rho(t)]$  we accept events at time  $t$  with probability  $p(t) = \rho(t)/\max_t[\rho(t)]$ . This procedure can formally only be applied for memory-less hidden processes, i.e. Poisson processes, in which case it results in an inhomogeneous Poisson process. To further reproduce heterogeneous rates in the inhomogeneous Poisson process, we rescale the rates of the hidden processes,  $\rho_i(t) = w_i \rho(t)$ , which keeps  $p(t)$  fixed.

To reproduce the empirical distribution of inter-encounter intervals, we construct a Weibull-renewal process: inter-encounter intervals are drawn from a Weibull distribution with scale parameter  $\lambda$  and shape parameter  $k$ . The Weibull distribution was parameterized by a fit to the data yielding  $(k, \lambda) = (0.3690, 3030)$ . To further reproduce heterogeneous rates in the Weibull-renewal process, we notice that the mean rate of a Weibull-renewal process is given by  $\rho_i = [\lambda_i \Gamma(1 + 1/k_i)]^{-1}$ , such that we can simply choose  $k_i = k$  and  $\lambda_i = \lambda/w_i$ .

To combine all features in a single model, we construct a tailored renewal process: A Weibull-renewal process with heterogeneous rates and additional (heuristic) thinning. We start with a set of hidden Weibull-renewal processes with  $k_i = k$ ,  $\lambda_i = \lambda/w_i$ , and time-dependent acceptance probability  $p(t)$  with time-average  $\bar{p}(t)$ . The mean rate of each process is  $\rho_i = \bar{p}(t)w_i/\lambda\Gamma(1 + 1/k)$ . Since we cannot fit  $(k, \lambda)$  of the hidden process, we further constrain the parameters with the mean rate from data, i.e.  $\rho(t) = \langle \rho_i \rangle = \bar{p}(t)/\lambda\Gamma(1 + 1/k)$ , with  $\langle w_i \rangle = 1$  by construction. Since  $\rho(t)/\bar{p}(t) = \max_t[\rho(t)]$ , we thus find  $\lambda = [\max_t[\rho(t)]\Gamma(1 + 1/k)]^{-1}$ , such that  $k$  remains the only free parameter. We obtained our best estimate of  $k$  by minimizing the Kullback–Leibler divergence [39] between the distribution tails ( $\delta t \gtrsim 0.5$  days) of model and empirical  $P(\delta t)$ , finding  $k \approx 0.24$ .

**Jackknife error estimation:** To estimate statistical errors of our results, we use jackknife error estimation while carefully taking into account the size of the left-out data set. The basic idea of the jackknife method is to estimate from some data  $X = \{x_1, \dots, x_g\}$  the variance of an observable  $\hat{O} = f(X)$  using a resampling approach [40]. Jackknife samples  $O_j$  are generated by systematically leaving out data, e.g.  $\hat{O}_j = f(X_j)$  with  $X_j = \{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_g\}$ . Importantly, here each  $x_j$  can be a block of (differently many) data points. While typically these blocks have the same size  $m$  (delete- $m$  jackknife), they could have different sizes  $m_j$  (delete- $m_j$  jackknife), which is relevant for some of our cases. From the jackknife samples, one can show that bias-reduced estimators of the mean and variance are given by [41]

$$\begin{aligned}\hat{O}_J &= \sum_{j=1}^g \frac{1}{h_j} \left( h_j \hat{O} - (h_j - 1) \hat{O}_j \right), \\ \hat{\sigma}_J^2 &= \frac{1}{g} \sum_{j=1}^g \frac{1}{h_j - 1} \left( h_j \hat{O} - (h_j - 1) \hat{O}_j - \hat{O}_J \right)^2,\end{aligned}\quad (3)$$

where  $h_j = (\sum_{i=1}^g m_i)/m_j$ , and  $\hat{O} = f(X)$  is the naive estimate. For blocks of equal size,  $m_j = m$ , we have  $h_j = g$  and this simplifies to

$$\begin{aligned}\hat{O}_J &= g\hat{O} - \frac{g-1}{g} \sum_{j=1}^g \hat{O}_j, \\ \hat{\sigma}_J^2 &= \frac{g-1}{g} \sum_{j=1}^g \left( \hat{O}_j - \frac{1}{g} \sum_{j=1}^g \hat{O}_j \right)^2.\end{aligned}\quad (4)$$

In our case, the data  $X$  is the set of all encounter trains and in the resampling step we leave out individual encounter trains. Since trains include differently many encounters, this can result in removing blocks of different sizes. In particular, all observables that derive from the number of encounters, e.g.  $\bar{n}_{\text{inf}}$  or  $P(n_{\text{inf}})$ , require the delete- $m_j$  analysis, equation (3), to estimate the statistical error. On the other hand, for observables that depend on time-binned data, e.g. the time-dependent rate, each encounter train has the

same size given by the number of time bins during the recording such that the delete- $m$  analysis, equation (4), is sufficient to estimate the statistical error.

## Data availability statements

The data that support the findings of this study are openly available [13, 20] and our code is available at the following URL: [https://github.com/Priesemann-Group/contact\\_pattern\\_outbreak](https://github.com/Priesemann-Group/contact_pattern_outbreak) [42].

## Acknowledgments

We would like to thank Peter Sollich and Sune Lehmann for helpful discussions. J Z received financial support from the Joachim Herz Stiftung. F P S and V P acknowledge funding from the SFB 1528 *Cognition of Interaction*. V P received support from the Federal Ministry of Education and Research (BMBF) via the RESPINOW (031L0298) and infoXpand (031L0300A) projects. J Z, F P S, J D, V P, and M Wi acknowledge funding by the Max Planck Society.

## ORCID iDs

Johannes Zierenberg  <https://orcid.org/0000-0001-5840-3791>

F Paul Spitzner  <https://orcid.org/0000-0001-9774-4572>

Jonas Dehning  <https://orcid.org/0000-0002-1728-2505>

Viola Priesemann  <https://orcid.org/0000-0001-8905-5873>

Martin Weigel  <https://orcid.org/0000-0002-0914-1147>

Michael Wilczek  <https://orcid.org/0000-0002-1423-8285>

## References

- [1] Pastor-Satorras R, Castellano C, Van Mieghem P and Vespignani A 2015 Epidemic processes in complex networks *Rev. Mod. Phys.* **87** 925–79
- [2] Newman M E J and Park J 2003 Why social networks are different from other types of networks *Phys. Rev. E* **68** 036122
- [3] Goltsev A V, Dorogovtsev S N, Oliveira J G and Mendes J F F 2012 Localization and spreading of diseases in complex networks *Phys. Rev. Lett.* **109** 128702
- [4] Ahn Y-Y, Bagrow J P and Lehmann S 2010 Link communities reveal multiscale complexity in networks *Nature* **466** 761–4
- [5] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E and Makse H A 2010 Identification of influential spreaders in complex networks *Nat. Phys.* **6** 888–93
- [6] Barabási A-L 2005 The origin of bursts and heavy tails in human dynamics *Nature* **435** 207–11
- [7] Van Mieghem P and van de Bovenkamp R 2013 Non-Markovian infection spread dramatically alters the susceptible-infected-susceptible epidemic threshold in networks *Phys. Rev. Lett.* **110** 108701
- [8] Du Z, Fox S J, Holme P, Liu J, Galvani A P and Ancel Meyers L 2018 Periodicity in movement patterns shapes epidemic risk in urban environments (arXiv:1809.05203)
- [9] Towers S and Chowell G 2012 Impact of weekday social contact patterns on the modeling of influenza transmission and determination of the influenza latent period *J. Theor. Biol.* **312** 87–95
- [10] Masuda N and Holme P 2013 Predicting and controlling infectious disease epidemics using temporal networks *F1000Prime Rep* **5** 6
- [11] Holme P 2015 Modern temporal network theory: a colloquium *Eur. Phys. J. B* **88** 234
- [12] Sekara V, Stopczynski A and Lehmann S 2016 Fundamental structures of dynamic social networks *Proc. Natl. Acad. Sci.* **113** 9977–82
- [13] Génois M and Barrat A 2018 Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Sci.* **7** 1–18
- [14] Schlosser F, Maier B F, Jack O, Hinrichs D, Zachariae A and Brockmann D 2020 COVID-19 lockdown induces disease-mitigating structural changes in mobility networks *Proc. Natl. Acad. Sci.* **117** 32883–90
- [15] Leitch J, Alexander K A and Sengupta S 2019 Toward epidemic thresholds on temporal networks: a review and open questions *Appl. Netw. Sci.* **4** 105
- [16] Presigny C, Holme P and Barrat A 2021 Building surrogate temporal network data from observed backbones *Phys. Rev. E* **103** 052304
- [17] Kiss Ian Z, Miller J C and Simon Peter L 2017 *Mathematics of Epidemics on Networks: From Exact to Approximate Models (Interdisciplinary Applied Mathematics)* vol 46 (Cham: Springer)
- [18] Zierenberg J, Wilting J, Priesemann V and Levina A 2020 Description of spreading dynamics by microscopic network models and macroscopic branching processes can differ due to coalescence *Phys. Rev. E* **101** 022301
- [19] Nie Y, Zhong X, Tao W, Liu Y, Lin T and Wang W 2022 Effects of network temporality on coevolution spread epidemics in higher-order network *J. King Saud Univ. - Comput. Inf. Sci.* **34** 2871–82
- [20] Sapiezynski P, Stopczynski A, Dreyer Lassen D and Lehmann S 2019 Interaction data from the Copenhagen networks study *Sci. Data* **6** 1–10
- [21] SocioPatterns (available at: <http://www.sociopatterns.org>)
- [22] Lloyd-Smith J O and Rees M 2007 Maximum likelihood estimation of the negative binomial dispersion parameter for highly overdispersed data, with applications to infectious diseases *PLoS One* **2** e180
- [23] Lloyd-Smith J O, Schreiber S J, Kopp P E and Getz W M 2005 Superspreading and the effect of individual variation on disease emergence *Nature* **438** 355–9
- [24] Peak C M, Childs L M, Grad Y H and Buckee C O 2017 Comparing nonpharmaceutical interventions for containing emerging epidemics *Proc. Natl. Acad. Sci.* **114** 4023–8

- [25] Hellewell J et al 2020 Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts *Lancet Glob. Health* **8** e488–96
- [26] Harris T E 1963 *The Theory of Branching Processes* (Berlin: Springer)
- [27] Jiang Z-Q, Xie W-J, Li M-X, Podobnik B, Zhou W-X and Eugene Stanley H 2013 Calling patterns in human communication dynamics *Proc. Natl Acad. Sci.* **110** 1600–5
- [28] Newman M E J 2010 *Networks: An Introduction* (Oxford: Oxford University Press)
- [29] Chen P Z, Koopmans M, Fisman D N and Gu F X 2021 Understanding why superspreading drives the COVID-19 pandemic but not the H1N1 pandemic *Lancet Infect. Dis.* **21** 1203–4
- [30] Scholtes I, Wider N, Pfizner Re, Garas A, Tessone C J and Schweitzer F 2014 Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks *Nat. Commun.* **5** 5024
- [31] Bailey N T J 1964 Some stochastic models for small epidemics in large populations *Appl. Stat.* **13** 9
- [32] Anderson D and Watson R 1980 On the spread of a disease with gamma distributed latent and infectious periods *Biometrika* **67** 191–8
- [33] Lloyd A L 2001 Destabilization of epidemic models with the inclusion of realistic distributions of infectious periods *Proc. R. Soc. B* **268** 985–93
- [34] Masuda N and Holme P 2020 Small inter-event times govern epidemic spreading on networks *Phys. Rev. Res.* **2** 023163
- [35] Starnini M, Gleeson J P and Boguñá M 2017 Equivalence between Non-Markovian and Markovian dynamics in epidemic spreading processes *Phys. Rev. Lett.* **118** 128301
- [36] Feng Mi, Cai S-M, Tang M and Lai Y-C 2019 Equivalence and its invalidation between non-Markovian and Markovian spreading dynamics on complex networks *Nat. Commun.* **10** 3748
- [37] Sekara V and Lehmann S 2014 The strength of friendship ties in proximity sensor data *PLoS One* **9** e100915
- [38] Lewis P A W and Shedler G S 1979 Simulation of nonhomogeneous Poisson processes by thinning *Nav. Res. Logist. Q.* **26** 403–13
- [39] Kullback S and Leibler R A 1951 On information and sufficiency *Ann. Math. Stat.* **22** 79–86
- [40] Efron B 1982 *The Jackknife, the Bootstrap and Other Resampling Plans* (Philadelphia: Society of Industrial and Applied Mathematics)
- [41] Busing F M T A, Meijer E and Van Der Leeden R 1999 Delete-m jackknife for unequal m *Stat. Comput.* **9** 3–8
- [42] Analysis code (available at: [https://github.com/Priesemann-Group/contact\\_pattern\\_outbreak](https://github.com/Priesemann-Group/contact_pattern_outbreak))