Temporal network structures controlling disease spreading

Petter Holme[*]

Department of Energy Science, Sungkyunkwan University, Suwon 440-746, Korea
(Received 3 May 2016; published 15 August 2016)

We investigate disease spreading on eight empirical data sets of human contacts (mostly proximity networks recording who is close to whom, at what time). We compare three levels of representations of these data sets: temporal networks, static networks, and a fully connected topology. We notice that the difference between the static and fully connected networks—with respect to time to extinction and average outbreak size—is smaller than between the temporal and static topologies. This suggests that, for these data sets, temporal structures influence disease spreading more than static-network structures. To explain the details in the differences between the representations, we use 32 network measures. This study concurs that long-time temporal structures, like the turnover of nodes and links, are the most important for the spreading dynamics.

I. INTRODUCTION

The spread of infectious disease continues to be one of the major challenges to global health. Despite advances in genomics and access to large data sets, predicting outbreaks is still disturbingly difficult [1]. Even knowing fairly much about an outbreak (at the time of writing the major concern regards the Zika virus [2]), nobody can be very certain about its future. The basic mathematics of infectious disease outbreaks as emergent phenomena is well studied. No paper, to our knowledge, has strong alternatives to compartmental models—models dividing individuals into classes with respect to the disease and assigning transition rules between the classes. Straightforward implementations of compartmental models do not, however, explain the difficulties in predicting emergent outbreaks in real populations [3]. There can be many reasons for this difficulty to predict the extinction time outbreaks. One obvious reason is that the data quality is still not good enough to make high-precision forecasting. There can, however, be other, more fundamental issues with how the compartmental models are integrated with models of contact patterns (describing how people meet in such a way that disease can spread). In this paper, we investigate different levels of representing contact structures: as temporal networks (including information both of the time of contact and the individuals involved), as static networks (including information of pairs of people between which the disease can spread), and as fully connected networks (which is the traditional contact structure of theoretical epidemiology [4]).

Many studies have pointed out that, to model disease spreading accurately, we need to understand both static-network structures [5,6] and temporal-network structures [7]. To make this point, a standard approach has been to first observe some structure in empirical data, then use models to prove this structure affects disease spreading, and finally conclude that this structure is important for epidemics. For example, Ref. [8] observed power-law distributions of degree (number of neighbors in the network) in sexual networks, and Ref. [9] showed that model networks with power-law degree distributions need not have an epidemic threshold, thus concluding the degree distribution is an important structure.

For another example, Ref. [10] found power laws in interevent time distributions, and Refs. [11,12] showed that outbreaks are slowed down by such fat-tailed distributions. Can we from this conclude that timing of contacts is important for disease spreading? Perhaps, but Ref. [13] argued that other, longer time-scale temporal structures are even more important. However, also Ref. [13] tested two *a priori* chosen structures. There could of course be other structures present affecting the spreading processes even more strongly. The idea of this paper is to scan the possible structures in a less restrictive way, so as to be open to the discovery of new important temporal-network structures. For the same reason—that it is hard to *a priori* reason about what the important temporal-network structures are—we use empirical networks as our starting point rather than models generating the contact structure.

In this paper, we run the susceptible-infectious-recovered (SIR) compartmental disease spreading model (a canonical model for diseases that give immunity upon recovery) on eight human contact networks. We use temporal-network, static-network, and fully connected representations of these data sets. Then, to explain the deviations between the three representations, we explore 32 quantities measuring temporal-network structure.

II. PRELIMINARIES

In this section, we clarify the methods and precise model definitions in common with the rest of the paper. We also mention some computational considerations. In general, we assume a temporal network H as input. It can be described as a list of C *contacts* (i,j,t) where $i,j \in V$ are individuals and t is the time of the contact (assuming a discretized time, as is common for most data sets). The order of i and j does not matter. We set the smallest time to zero and label it T.

A. SIR simulation

We use the constant-duration version of the SIR model [14]. In this model, a contact between susceptible and infectious individuals infects the susceptible (instantaneously) with a probability λ. Then the infectious recovers a time δ later, and stays recovered for the rest of the simulation. The infection seed is chosen randomly and taken to become infectious immediately prior to its first contact. When there are no

---
[*]holme@skku.edu

infectious individuals left, the infection is extinct. The time from when the infection is introduced until the last recovery is the extinction time, τ. The fraction of recovered individuals when the outbreak is extinct is the outbreak size, Ω.

For the static networks, we consider a disease spreading on a graph $G = (V, E)$ where $(i, j) \in E$ if there is a contact $(i, j, t) \in H$. We generate C contacts between individuals connected by a link in E at randomly chosen times between zero and T. Thus we use as closely as possible the original data (assuming the maximum entropy principle—to maximize the randomness of the unknown structures). Analogously, for the fully connected case, we also generate C contacts at times in the interval $[0, T]$, but this time it can be between any pair of nodes.

Each data point of the 20×20 parameter combinations is averaged over 2×10^5 independent runs. We let the sequences of λ and δ grow exponentially, as will be evident later, and exponential growth is needed to separate the data sets. For the same reason it is convenient to use the logarithm (we use the base-ten logarithm) of these values for discussion.

B. Data sets

As mentioned above, this study is based on empirical data sets of human proximity. In other words, they record two persons in close proximity at a certain time. For obvious reasons, these are interesting for disease spreading. We list the basic statistics—sizes, sampling durations, etc.—of the data sets in Table I.

Our first data set (*Prostitution*) comes from self-reported sexual contacts between female prostitutes and male sex buyers [15]. This is a special form of proximity network in that a contact is sexual. Perhaps it should be classified as a separate type of network, but it is relevant for disease spreading. Several other data sets come from the Sociopatterns project (sociopatterns.org). These data sets are created by radio-frequency identification sensors that record a contact when two sensors are within 1–1.5 m. One of these data sets comes from a conference [16] (*Conference*), another from a school (*School*) [17], a third from a hospital (*Hospital*) [18], and a fourth from an art gallery (*Gallery*) [19]. The *Gallery*

TABLE I. Basic statistics of the data sets. N is the number of individuals, C is the number of contacts, T is the total sampling time, t-res. is the time resolution of the data set, and M is the number of links in the projected static networks.

Data set	N	C	T	t-res.	M
Prostitution	16 730	50 632	6.00 yr	1 d	39 044
Conference	113	20 818	2.50 d	20 s	2 196
Hospital	75	32 424	96.5 h	20 s	1 139
Reality	64	26 260	8.63 h	5 s	722
School 1	236	60 623	8.64 h	20 s	5 901
School 2	238	65 150	8.58 h	20 s	5 541
Gallery 1	200	5 943	7.80 h	20 s	714
Gallery 2	204	6 709	8.05 h	20 s	739

data set comprises 69 days where we use the first two. *School* consists of 2 days and we use both.

A similar data set to the Sociopatterns data sets comes from the Reality mining study [20] (*Reality*). Here contacts within a cohort of university students were recorded by the Bluetooth channel of smartphones. The range of such connections is between 10 and 15 m. We use the same subset of the data set as in Ref. [21].

C. Temporal network descriptors

To characterize the temporal-network structures of the data sets, we use 32 different quantities, which we call network descriptors. We choose these both to be relatively simple and straightforward to interpret and to cover as wide a spectrum of structures as possible.

1. Time evolution

We measure nine network descriptors characterizing the long-term behavior of the contact dynamics—briefly speaking, how the contact process differs from a stationary process. Some of these data sets (e.g., *Prostitution*, *Gallery 1* and *Gallery 2*) are growing throughout the sampling period, and this has been argued to influence the spreading dynamics
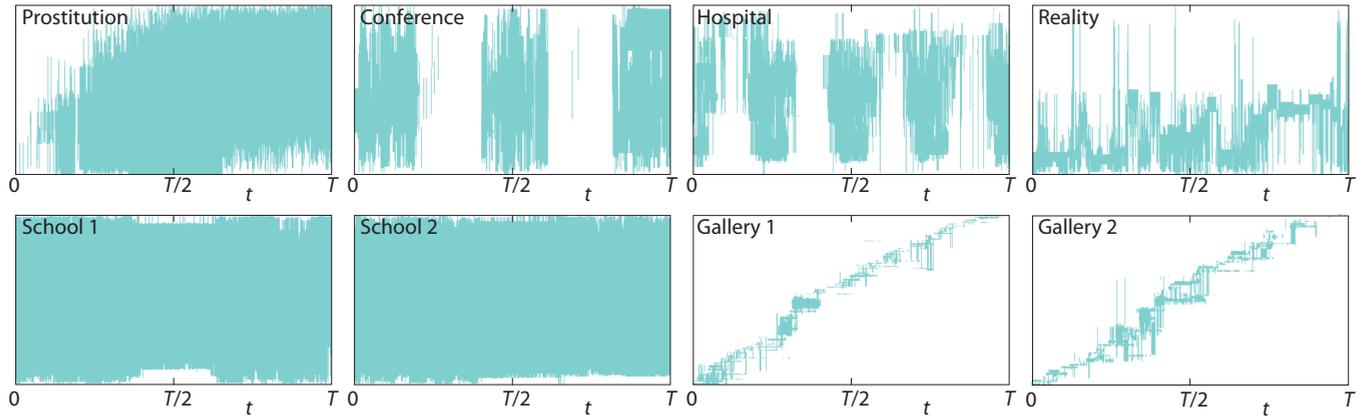

FIG. 1. A visualization of the temporal structures of the data sets. The nodes are represented by a coordinate of the vertical axis. A contact at a certain time is displayed as a horizontal line between the coordinates of the nodes involved. The assignment of coordinates is optimized to reduce the total vertical distance of the lines.
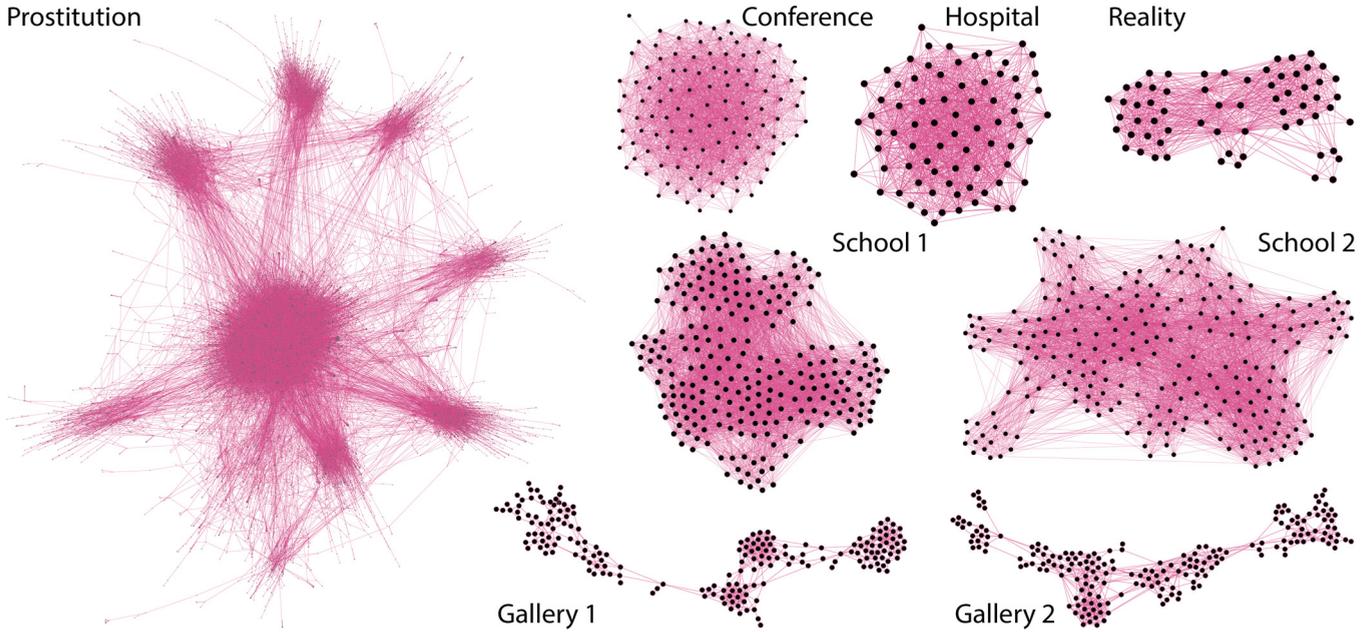
FIG. 2. The network structure of the networks of aggregate contacts displayed using the "Force Atlas 2" method of the software package GEPHI (gephi.org).

strongly [13]. In such a system, the disease could burn out in the population even before some individuals have entered it.

The first of these measures focuses on the time when nodes and links first appear in the data. First, we measure the fraction of nodes (links) present at half the sampling time relative to the final number of nodes, f_{TN} (links, f_{TL}). Some studies argue that the order of events is a more natural measure of time than the actual time. Thus we also measure the corresponding quantities f_{CN} and f_{CL}, where half the sampling time is replaced by half the contacts.

The second class of network descriptors focuses on the persistence nodes or links. Let F_{TN} (F_{TL}) be the fraction of nodes (links) present in the first and last 5% of the time. The corresponding quantities for the sequence of contacts are F_{CN} and F_{CL}.

Yet a measure related to the time evolution is the largest gap g on the contact sequence. (During a gap, the disease cannot spread, and for long enough gaps, the disease could die out.)

2. Node and link activity

The node- and link-activity descriptors capture the bursty nature of human behavior, i.e., intense periods of activity separated by long periods of inactivity [22]. One can imagine many ways to measure burstiness. The common starting point is interevent times—the time gap between consecutive contacts of a node or link. We measure four descriptors characterizing this kind of time series—the mean μ, standard deviation σ, coefficient of variation, c (i.e., the standard deviation divided by the mean), and the skewness,

$$\gamma = \frac{(n^2 - n)^{1/2}}{n - 2} \frac{\mu_3}{\mu_2^{3/2}},\qquad(1)$$

where μ_2 and μ_3 are the second and third moment of the distribution, respectively.

Some studies have pointed out that the duration (time from the first to the last observation) of nodes or links can be important for spreading phenomena [13]. Therefore, we also study the distribution of node and link durations by the
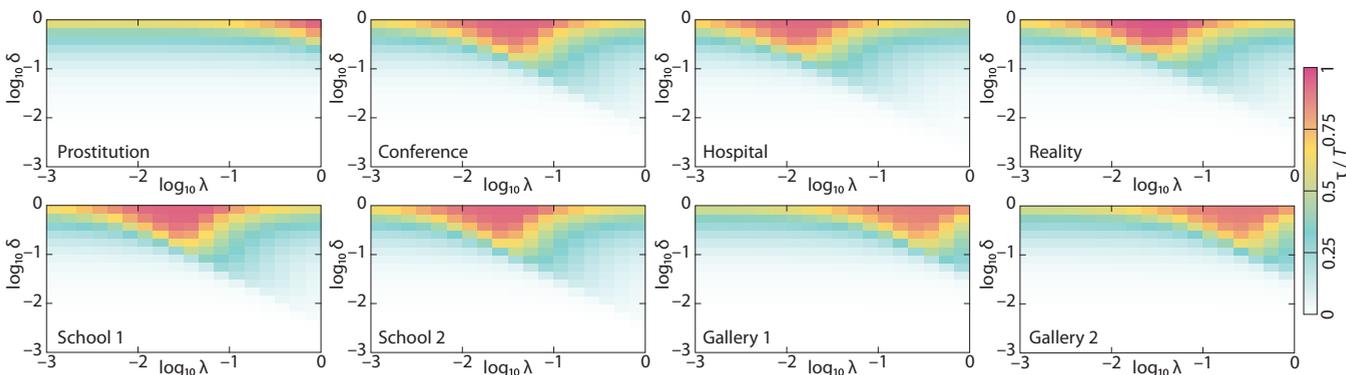

FIG. 3. Average extinction times for the SIR model on eight temporal-network data sets of human proximity.
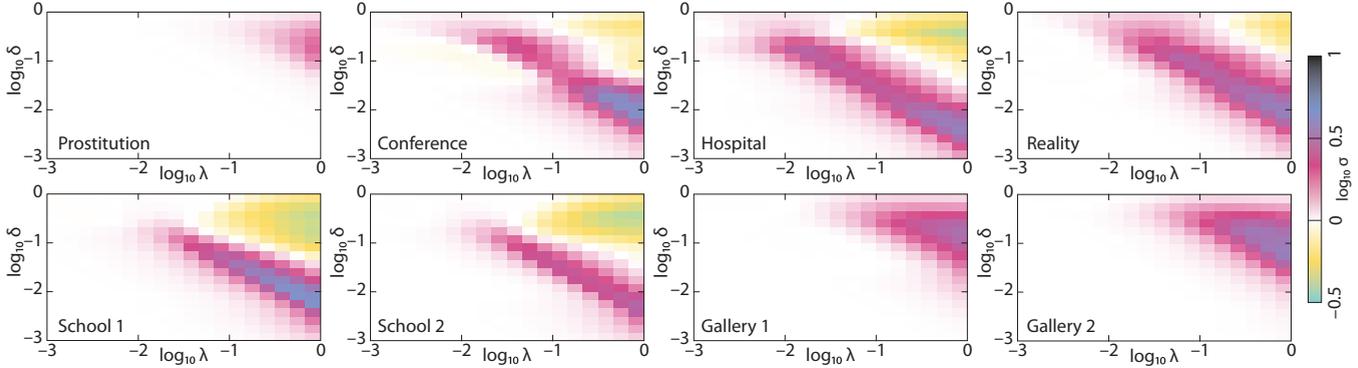
FIG. 4. The difference between the average time to extinction of static- and temporal-network representations of the contact patterns.

same four descriptors as the interevent times. In total, for this category, we define 16 network descriptors—μ, σ, c, and γ, for both interevent-time and duration distributions and for both nodes and links. See Fig. 1 for visualizations of the time structure of the data sets.

3. Measures of static-network structure

How contact structures affect dynamic processes, such as epidemic spreading, is more established for static-network structures than for temporal structures. We measure the static-network structure for the networks of accumulated contacts; i.e., if there has been at least one contact between two nodes we consider them connected by an link.

Arguably, the most important static-network structure is the degree distribution describing how frequent it is to observe a node of a particular degree. Essentially a broad, right-skewed degree distribution (such as that frequently observed in real systems) speeds up spreading phenomena [6]. Usually researchers are interested in inferring the functional form of the degree distribution. For our purpose, we need to summarize the structures to numbers, no matter the functional forms. Therefore, we measure the same four quantities—μ, σ, c, and γ—as for the interevent time and duration distributions.

In addition to the degree distribution, we also measure other static-network descriptors. First, and simplest, is the number of nodes, N (but not the number of links since it is equal to $N\mu_{\text{deg}}/2$). The next static-network descriptor is the assortativity r. This is, in essence, the Pearson correlation of

the degrees at either side of a link. One only has to symmetrize the arguments of the correlation coefficient (since the first and second arguments are different, but links are unordered with respect to the nodes—see Ref. [23] for details). The assortativity captures the tendency for nodes of similar degree to connect to each other. A large assortativity means that high-degree nodes connect to other high-degree nodes, and low-degree nodes connect to other low-degree nodes. It has been shown to have an influence on disease dynamics—assortative networks have lower epidemic thresholds [24]. Finally, we study the clustering coefficient—the number of triangles in the network normalized to the unit interval [23]. A high clustering coefficient is known to slow down disease spreading [25].

See Fig. 2 for visualizations of the networks of accumulated contacts. Just like Fig. 1, this figure does not tell us more than that there are rich structures in the network topology that can influence the outbreak dynamics.

D. Overlap statistics

We look at groups of data sets with different behavior of the SIR model with respect to the three levels of representations of the contacts. A good candidate network descriptor should separate the two groups well. With more samples, we could use, e.g., the mutual information or Kullback-Leibler divergence, but with only eight data points, we can use a simpler quantity relying on the extreme values of the quantity for the two groups. Let A be one subset of the data sets and B
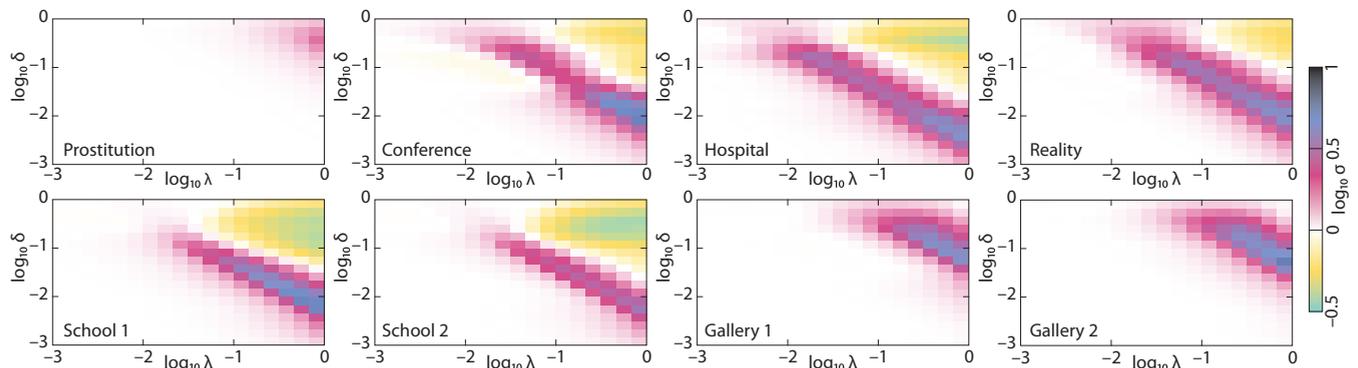

FIG. 5. The difference between the average time to extinction of fully mixed and temporal-network representations of the contact patterns.
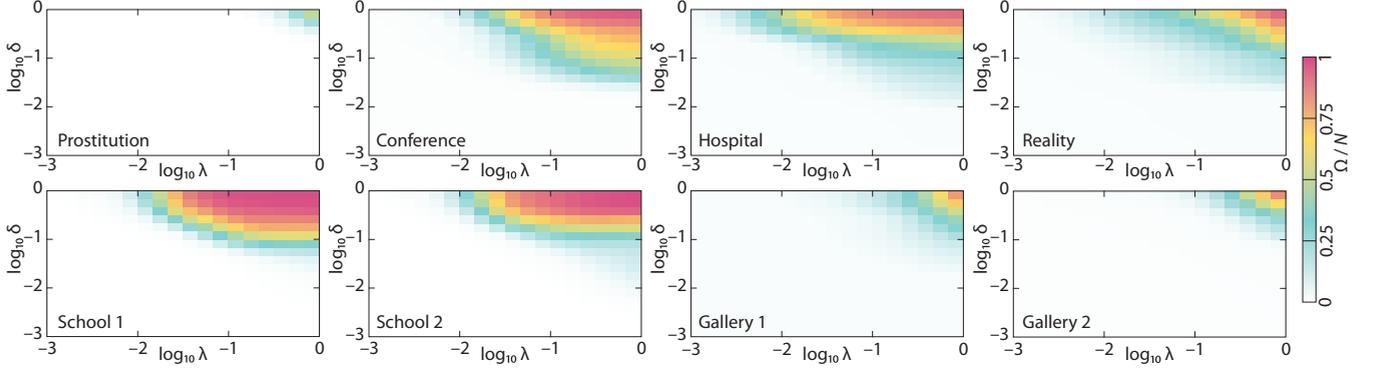
FIG. 6. Average outbreak sizes for the SIR model on eight temporal-network data sets of human proximity.

its complement. Let $v(G)$ be the value for a quantity as a function of the data set G. Furthermore, assume (without loss of generality) that $\max_{G \in A} v(G) \geqslant \max_{G \in B} v(G)$. Then, more specifically, we measure

$$x_v(A,B) = \frac{\min_{G \in A} v(G) - \max_{G \in B} v(G)}{\max_{G \in A} v(G) - \min_{G \in A \cup B} v(G)}. \quad (2)$$

In other words, if $\{v(G) : G \in A\}$ and $\{v(G) : G \in B\}$ do not overlap, then x is the smallest difference between values in the two sets divided by the largest difference. If $x = 1$, the separation is maximal. If A and B do overlap, x will be negative, reaching a minimum of -1 if the range of $\{v(G) : G \in A\}$ and $\{v(G) : G \in B\}$ are the same.

III. RESULTS

A. Extinction time

One of our main quantities is the mean time to extinction, τ. Figure 3 shows the values for the SIR model simulated on temporal-network representations. τ is strictly increasing with the disease duration δ but has a maximum in the per-contact transmission probability λ. The maximum comes from two conflicting mechanisms [26]. For small λ, decreasing λ gives fewer chances for contagion and an increasing chance of the disease dying out. For large λ, the disease burns out fast in the population. The actual location of the peak varies much, from close to the maximum $\lambda = 1$ for the *Prostitution* data set to $\log_{10} \lambda \approx -1.8$ for the *Hospital* data.

The effect of removing the temporal information by aggregating the contacts to a static network is seen in Fig. 4. This figure shows the deviation $\Delta \tau$ between τ of the static and temporal networks (so negative values mean the outbreaks last longer in temporal networks). We see the *Prostitution*, *Gallery 1*, and *Gallery 2* data sets are different than the others in that they do not have regions of negative $\Delta \tau$—the static networks always give longer outbreaks. If we proceed, removing the network structure by making the network fully mixed (i.e., fully connected), then not much more happens (Fig. 5). $\Delta \tau$ becomes larger for some regions of, in particular, the *Gallery* data sets. The qualitative picture is, however, the same. Except for *Prostitution*, *Gallery 1*, and *Gallery 2*, extinction times are underestimated for the largest $\log_{10} \lambda$ and $\log_{10} \delta$ and overestimated for intermediate $\log_{10} \lambda$ and $\log_{10} \delta$. This seems to suggest that the extinction time is more dependent on temporal than on topological structures—a hypothesis in line with previous studies [27] but still open for future studies. For some data sets, this is somewhat trivial—the *Hospital* data, as the most extreme example, has 1,139 links (i.e., ~41% of the pairs of nodes are connected). On the one hand, this can certainly be true in many real situations, like a disease spreading at a hospital ward, farm, or some other closed community. On the other hand, the observation is still true even for the sparsest of data sets.

B. Outbreak size

The average expected outbreak size Ω is perhaps a more common quantity than τ to characterize outbreaks in
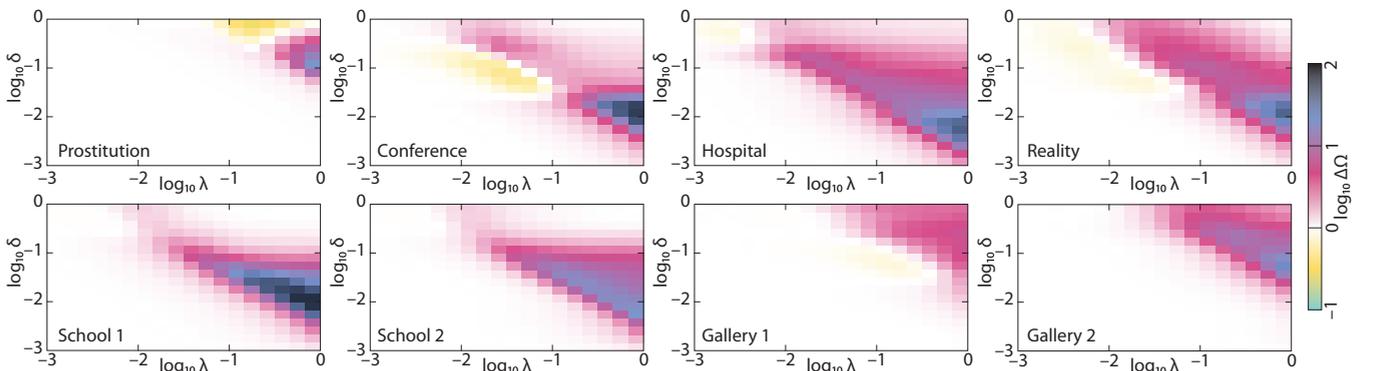

FIG. 7. The difference between the average outbreak size of static- and temporal-network representations of the contact patterns.
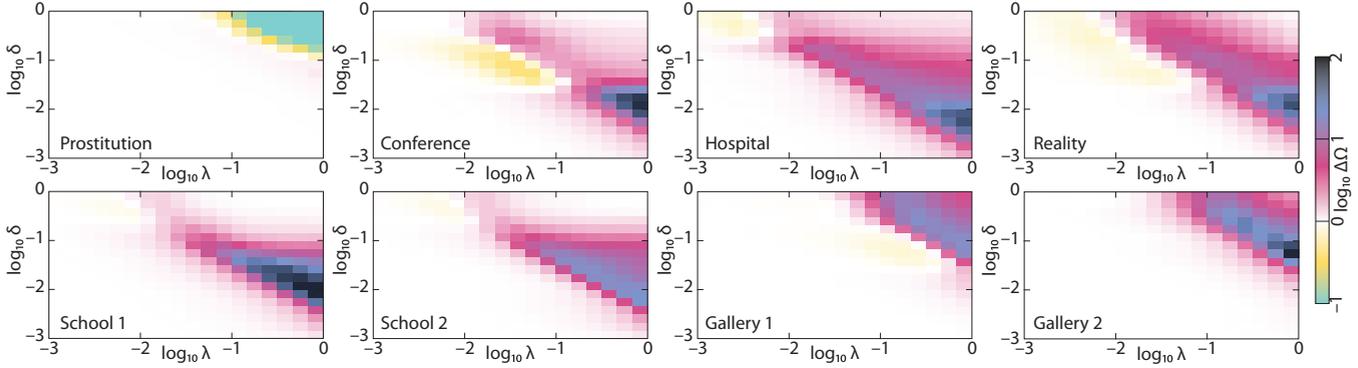
FIG. 8. The difference between the average outbreak size of fully mixed and temporal-network representations of the contact patterns.

computational studies of disease spreading. Figure 6 shows the values of Ω throughout the parameter space (for most data sets, these values were also presented in our Ref. [28]). Ω is monotonically increasing with $\log_{10} \lambda$ and $\log_{10} \delta$ which probably is inevitable on average (even though, for specific seeds i, a larger $\log_{10} \delta$ can lead to the disease burning out so fast around i that it is already extinct when a contact leading away from i's vicinity appears).

Figures 7 and 8 show, respectively, the deviation when the temporal and both temporal and topological information is removed. Unlike τ, the static-network structure creates a qualitative difference—but only for the *Prostitution* data. For this data set, the outbreak sizes are consistently underestimated for the fully connected networks, while for the static networks, $\lambda = \delta$ roughly separates two regions—for $\delta > \lambda$ the outbreak sizes are overestimated whereas for $\delta < \lambda$ they are underestimated. Below, we look for a structural explanation behind this phenomenon.

C. Structural explanations

In this section, we try to find which network structures affect the effects found above. First, the *Prostitution* and *Gallery* data sets differ from the others in that they lack a region of parameter space where the representations without temporal structure overestimate the time to extinction. Second, the *Prostitution* data set has a different response to removing the network structure than all the other data sets.

First, we investigate which network descriptors that separates $A = \{Prostitution, Gallery\ 1, Gallery\ 2\}$ from the rest [A refers to Eq. (2) and the discussion about it]. The top three quantities v with respect to x_v along with their values for the two groups of data sets are plotted in Fig. 9(a). (Values of the other quantities can be found in the Appendix.) These three quantities—the average life time of nodes, μ_{Nt}, and links, μ_{Lt}, and the fraction of nodes present at half of the contacts, f_{nC}—are all temporal in nature and all related to the turnover of individuals in the data, rather than higher frequency properties like the interevent time statistics. In more detail, we see that the data sets without regions of negative $\Delta \tau$ are characterized by a short average presence of the nodes and links in the data, and thus a high turnover of individuals. Representing such temporal networks as static networks destroys the long-time-scale effects, such as that a

node present early in the data cannot be infected by a node present only late in the data.

Our second investigation concerns how *Prostitution* differs from the other data sets [Fig. 9(b)]. We find that the quantities with the largest v values are the number of nodes, N, the average interevent time of nodes, μ_{Nt}, and the skewness of the degree distribution, γ_k. These three quantities are very different from the ones that explain the other effect [in Fig. 9(a)]. The number of nodes is probably not an explanation for this effect in itself, but it could help accentuate other effects. The long average interevent times of *Prostitution* come from a very skewed distribution of the number of contacts (a quantity we do not measure directly). The few-contact individuals can have long dormant periods and thus increase the average interevent time. (Individuals with only one contact, of which
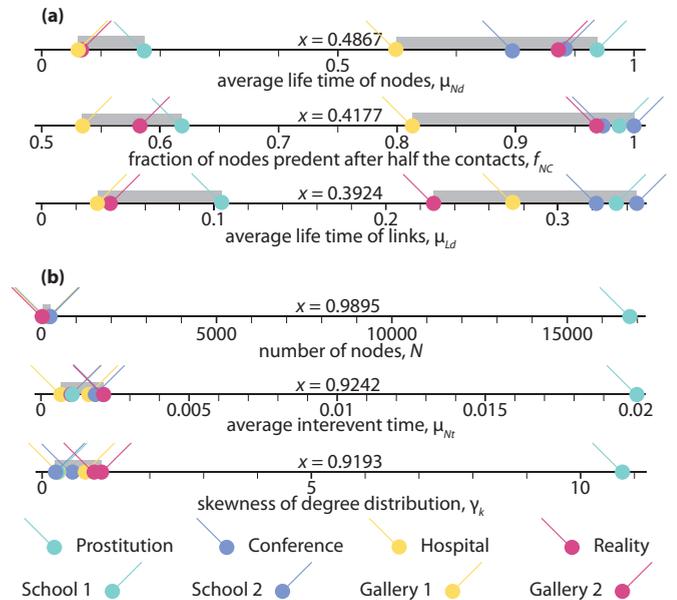

FIG. 9. (a) The top three temporal-network structures separating the *Prostitution*, *Gallery 1*, and *Gallery 2* data sets from the rest of the data sets and (b) separating *Prostitution* from the rest. x is the difference between the smallest value (of the network structural measure in question) of the set containing the largest value, and the largest value of the other set divided by the difference between the largest and smallest values in the union of the sets.

there are around 35%, do not contribute to μ_{Nt}.) The degree distribution is a very well studied quantity, responsible for many peculiar features in static-network epidemiology (such as the vanishing of epidemic thresholds or the emergence of superspreaders [6]). It is therefore reassuring to see its skewness as one of the top explanatory descriptors. It, furthermore, makes sense that the difference between the static networks and the fully connected networks is best explained by static-network quantities. However, except for the *Prostitution* data, the static and fully connected networks deviate from the temporal network in the same way, which means that the temporal structures are more influential with respect to disease spreading for these data sets, not only for τ but also for Ω.

IV. DISCUSSION

We have compared SIR simulations (the entire parameter space) on three levels of representations of empirical contact data—temporal networks, static networks, and fully connected networks. We used two quantities characterizing the evolution of the outbreak—the time to extinction and the average outbreak size. We see that going from a temporal-network representation to static-network or fully connected network representations can lead to both a severe under- and overestimation of both the extinction time and the average outbreak size. In general, short disease durations and high transmission probabilities lead to an overestimation when the temporal information is discarded. Going from a static-network representation to a fully connected topology does not make much of a difference except for one data set (*Prostitution*) and one of the quantities (average outbreak size). Looking closer at the quantities determining the patterns of over- and underestimation of τ and Ω also gives that quantities describing the time evolution of the network are the most influential structures (in agreement with Ref. [28]). Static-network structure and shorter-time-scale temporal structure such as interevent times matters less. These observations are, of course, specific for the particular data sets we study (which is in line with other studies [13,27]). The results should be generalized with care. On the other hand, the contact data sets we use are as good as we can possibly obtain. There are no obvious structures in these data sets that disqualify them as representative of real data sets (except, perhaps, the limited sizes). At the very least, this should encourage more research into the role of time structures in disease spreading.

There are many possible extensions of this work. Even though we used a generous amount of 32 network descriptors,

TABLE II. Symbols and brief explanations of the network descriptors.

Descriptor	A_1	A_2
f_{NC}, fraction of nodes present when half of the contacts happened	0.4177	−0.8875
f_{NT}, fraction of nodes present at half the sampling time	0.0240	−0.0294
f_{LC}, fraction of links present when half of the contacts happened	0.1289	0.0020
f_{LT}, fraction of links present at half the sampling time	0.3113	−0.3005
F_{NC}, fraction of nodes present at both the first and last 5% of the contacts	0.0762	0.0702
F_{NT}, fraction of nodes present at both the first and last 5% of the sampling time	0.0263	−0.0989
F_{LC}, fraction of links present at both the first and last 5% of the contacts	0.0610	0.0332
F_{LT}, fraction of links present at both the first and last 5% of the sampling time	−0.2930	−0.0515
μ_{Lt}, mean link interevent time	−0.6539	0.3460
σ_{Lt}, standard deviation of interevent times of links	−0.0022	−0.6815
c_{Lt}, coefficient of variation of interevent times of links, i.e., the average link burstiness	0.1274	0.0839
γ_{Lt}, skewness of interevent times of links	−0.7223	0.1574
μ_{Ld}, mean duration (time between first and last contact) of links	0.3924	0.0205
σ_{Ld}, standard deviation of the duration of links	0.2138	0.3460
c_{Ld}, coefficient variation of the duration of links	−0.8187	−0.6815
γ_{Ld}, skewness of the duration distribution of links	−0.0097	0.0839
μ_{Nt}, like μ_{Lt} but for nodes	−0.0590	0.9242
σ_{Nt}, like σ_{Lt} but for nodes	−0.3793	0.6206
c_{Nt}, like c_{Lt} but for nodes, i.e., the node burstiness	0.1882	0.0382
γ_{Nt}, like γ_{Lt} but for nodes	−0.1489	0.2697
μ_{Nd}, like μ_{Ld} but for nodes	0.4867	−0.1257
σ_{Nd}, like σ_{Ld} but for nodes	−0.3577	−0.5342
c_{Nd}, like c_{Ld} but for nodes	0.1160	0.4076
γ_{Nd}, like γ_{Ld} but for nodes	0.1922	−0.9673
g, the longest gap between any two contacts in the data	0.1022	−0.1808
μ_k, average degree of the network of accumulated contacts	0.3378	0.0545
σ_k, standard deviation of the degree distribution of the network of accumulated contacts	0.0956	−0.3605
c_k, coefficient of variation of the degree distribution of the network of accumulated contacts	0.1009	0.8177
γ_k, skewness of the degree distribution of the network of accumulated contacts	−0.0169	0.9193
N, number of nodes	−0.0022	0.9895
C, clustering coefficient of the network of accumulated contacts	−0.1388	0.6517
r, degree assortativity of the network of accumulated contacts	−0.4554	−0.0989

one can imagine many others—describing how static-network quantities change over the sampling time, how the activity level of nodes and their network position are correlated, etc. It would also be interesting to include a weighted network representation as an intermediate between the static and temporal representations. Ultimately, one would like to use results from this type of study to construct generative models for outbreak scenarios, retaining the important structures but not more. Indeed, some such models have already been proposed [29,30] but, to our knowledge, none that focuses on the longer-time-scale features that we find important.

APPENDIX

This Appendix contains the x values for all network descriptors and the two splits $A_1 = \{Prostitution, Gallery\ 1, Gallery\ 2\}$ and $A_2 = \{Prostitution\}$, respectively. See Table II.

[1] J. M. Drake, PLoS Med. **3**, e3 (2005).
[2] L. R. Petersen, D. J. Jamieson, A. M. Powers, and M. A. Honein, New Engl. J. Med. **374**, 1552 (2016).
[3] S. Janson, M. Luczak, and P. Windridge, Random Struct. Alg. **45**, 726 (2014).
[4] H. W. Hethcote, SIAM Rev. **42**, 599 (2000).
[5] M. J. Keeling and K. T. Eames, J. R. Soc. Interface **2**, 295 (2005).
[6] R. Pastor-Satorras, C. Castellano, P. van Mieghem, and A. Vespignani, Rev. Mod. Phys. **87**, 925 (2015).
[7] N. Masuda and P. Holme, F1000Prime Rep. **5**, 6 (2013).
[8] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, Nature (London) **411**, 907 (2001).
[9] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[10] A.-L. Barabási, Nature (London) **435**, 207 (2005).
[11] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, Phys. Rev. E **83**, 025102 (2011).
[12] B. Min, K.-I. Goh, and A. Vazquez, Phys. Rev. E **83**, 036102 (2011).
[13] P. Holme and F. Liljeros, Sci. Rep. **4**, 4999 (2014).
[14] P. Holme, PLoS Comput. Biol. **9**, e1003142 (2013).
[15] L. E. C. Rocha, F. Liljeros, and P. Holme, Proc. Natl. Acad. Sci. USA **107**, 5706 (2010).
[16] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J. F. Pinton, and W. van den Broeck, J. Theor. Biol. **271**, 166 (2011).

[17] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. van den Broeck, C. Régis, B. Lina *et al.*, PLoS ONE **6**, e23176 (2011).
[18] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B.-A. Kim, B. Comte, and N. Voirin, PLoS ONE **8**, e73970 (2013).
[19] W. van den Broeck, M. Quaggiotto, L. Isella, A. Barrat, and C. Cattuto, Leonardo **45**, 285 (2012).
[20] N. Eagle and A. Pentland, Pers. Ubiquit. Comput. **10**, 255 (2006).
[21] R. Pfitzner, I. Scholtes, A. Garas, C. J. Tessone, and F. Schweitzer, Phys. Rev. Lett. **110**, 198701 (2013).
[22] K.-I. Goh and A.-L. Barabási, Europhys. Lett. **81**, 48002 (2008).
[23] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
[24] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[25] T. Britton, M. Deijfen, A. N. Lagerås, and M. Lindholm, J. Appl. Probab. **45**, 743 (2008).
[26] P. Holme, PLoS ONE **8**, e84429 (2013).
[27] K. Hock and N. H. Fefferman, Ecol. Complexity **12**, 34 (2012).
[28] P. Holme, Sci. Rep. **5**, 14462 (2015).
[29] P. Holme and J. Saramäki, Phys. Rep. **519**, 97 (2012).
[30] P. Holme, Eur. Phys. J. B **88**, 1 (2015).