Topological-temporal properties of evolving networks

ALBERTO CERIA*
Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands

SHLOMO HAVLIN
Department of Physics, Bar Ilan University, Ramat Gan, Israel; Department of Physics, Boston University, Boston, MA, USA and Tokyo Institute of Technology, Yokohama, Japan

ALAN HANJALIC
Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands

AND

HUIJUAN WANG
Faculty of Electrical Engineering, Mathematics, and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD, Delft, The Netherlands
*Corresponding author: A.Ceria@tudelft.nl

[Received on 3 November 2021]

Many real-world complex systems including human interactions can be represented by temporal (or evolving) networks, where links activate or deactivate over time. Characterizing temporal networks is crucial to compare such systems and to study the dynamical processes unfolding on them. A systematic method to characterize simultaneously the temporal and topological relations of active links (also called contacts or events), in order to compare different real-world networks and to detect their common patterns or differences is still missing. In this paper, we propose a method to characterize to what extent contacts that happen close in time occur also close in topology. Specifically, we study the interrelation between temporal and topological properties of contacts from three perspectives: (1) the autocorrelation of the time series recording the total number of contacts happened at each time step in a network; (2) the interplay between the topological distance and interevent time of two contacts; (3) the temporal correlation of contacts within local neighborhoods beyond a node pair. By applying our method on 13 real-world temporal networks, we found that temporal-topological correlation of contacts is more evident in virtual contact networks than in physical contact ones. This could be due to the lower cost and easier access of online communications than physical interactions, allowing and possibly facilitating social contagion, i.e., interactions of one individual may influence the activity of its neighbors. We also identify different patterns between virtual and physical networks and among physical contact networks at, e.g., school and workplace, in the formation of correlation in local neighborhoods. Detected patterns and differences may further inspire the development of more realistic temporal network models, that could reproduce jointly temporal and topological properties of contacts.

Keywords: Evolving networks, Structural analysis of networks, Social, socio-economic and political networks

1. Introduction

Complex systems can be represented as networks, where nodes and links represent the components of the system and their interactions respectively. In a temporal or evolving network [1, 2], the network topology changes over time, or equivalently, pairs of nodes interact at specific time stamps. Such time-stamped interactions between nodes are called contacts or events. Early work on evolving networks and their characterization methods have mostly focused on either temporal [3–7] or topological [8–13] dimension separately but rarely on combining both [14–19]. Regarding the topological aspect, the aggregated networks, where two nodes are connected if they have at least one contact or interaction, have been characterized using classical static network analysis methods. Scaling properties such as a scale-free degree distribution have been observed in many real networks [8–11]. From the perspective of time dimension, it has been found that individuals tend to execute actions like contacts in bursts within a short time duration and such high activity periods are separated by relatively long inactive ones. The approximate scale-free distribution of the inter-event times of contacts of a node or of a system, the so-called burstiness, seems to be common in real-world temporal networks [3–7, 20, 21]. The temporal correlation of the events of a network has been measured by e.g. auto-correlation [22] and the distribution of the number of contacts in a bursty period, the so-called event train. [23].

Recent studies have started to characterize both the topological and temporal properties together. Brot et al. [14], for instance, showed that events of addition and removal of links by users on a large online blogging platform do not occur sporadically at random nodes but rather occur in brief bursts in time and locally in topology. A similar conclusion has been drawn by Kikas et al. [15], in a study about dynamics of link addition and removal on a large Skype dataset. Temporal motifs are an ordered sequence of timestamp contacts among a small number of nodes conforming to a specific pattern as well as in a specific duration of time. The occurrence of diverse temporal motifs has been used to characterize and to classify evolving networks [16, 17]. Karsai et al. [18] characterized the sequence of contacts between each node and its neighbours using the distribution of the number of contacts in a bursty period, which is also called the event train size. However, it has been shown that bursty trains are usually formed by contacts between pair of nodes instead of in the aforementioned neighborhood of a node.

However, systematic methods to characterize simultaneously the temporal and topological relations of contacts/events for analyzing and better understanding real-world networks are still missing. In this work, we aim to develop methods to characterize to what extent contacts that happen close in time (topology) are also close in topology (time). Specifically, we characterize the relationship between temporal and topological properties of the contacts in real evolving networks from the following three perspectives: (a) The auto-correlation of the activity time series which records the total number of contacts in a network that happen at each time step; (b) The interplay between the topological distance and temporal delay of two contacts; (c) The temporal correlation of contacts within local neighborhoods beyond a node pair. These perspectives characterize simultaneously both the temporal and topological interrelations of contacts from a global level to a more granular level. In order to be able to characterize and compare real-world networks, normalization and three control network randomizations have been designed in our characterization methods. We apply our method on 13 real-world physical and virtual contact networks. We find that the temporal and topological correlation tends to be more evident in virtual contact networks compared to physical contact networks. This is likely because the online communications, which are of lower cost and easier to perform than physical contacts, allows and possibly facilitates social contagion, i.e. the interaction of one individual to influence the activity of its neighbors. Among physical contact networks, we also noticed that the number of node pairs that contribute to the temporal correlation at a local neighborhood seems to reflect the spatial constrains of individuals involved in the interactions. For example, the interactions of a student (employee) with more (fewer) peers in a primary school (workplace) contributes to the temporal correlation.

The detected patterns and differences could further guide the development of evolving network models, pushing the boundary of temporal network models towards reproducing jointly realistic temporal and topological properties. Moreover, temporal network properties influence the dynamic process which unfolds on the network [19, 24–34]. The temporal and topological correlation in an evolving network discovered using our methods could possibly better explain the dynamic process than topological property or temporal property alone.

2. Definitions

2.1 *Representation of a temporal network*

A network in which the interactions among nodes vary over time is called a temporal or evolving network. It can be represented by $\mathscr{G} = (\mathscr{N}, \mathscr{L})$, where $\mathscr{N}$ is the set of nodes (with size $|\mathscr{N}| = N$), $\mathscr{L} = \{\ell(i,j,t), t \in [0,T), i,j \in \mathscr{N}\}$ is the set of contacts, and each element $\ell(i,j,t)$ indicates that a contact or an interaction between node i and j occurs at time t. A temporal network can also be represented by a 3 dimensional adjacency matrix $\mathscr{A}_{N\times N\times T}$ whose elements $\mathscr{A}(i,j,t) = 1$ or $\mathscr{A}(i,j,t) = 0$ represent, respectively, the presence or the absence of a contact between node i and j at time t.

We consider undirected temporal networks, where $\ell(i,j,t) = \ell(j,i,t)$ and $\mathscr{A}(i,j,t) = \mathscr{A}(j,i,t)$. By aggregating the contacts between each node pair over the whole observation time $[0,T-1]$ one obtains the time aggregated network $G_W = (\mathscr{N}, \mathscr{L}_W)$. The aggregated network is static: two nodes i and j are regarded connected, i.e., $e(i,j) \in \mathscr{L}_W$, if there is at least one contact between i and j over the observation time $[0,T-1]$. The adjacency matrix of the unweighted aggregated network is denoted by $A_{N\times N}$ whose element $A(i,j) = 1$ or $A(i,j) = 0$ depending whether i and j are connected or not. Each link $e(i,j)$ in $\mathscr{L}_W$ can be further associated with a weight $W(i,j)$, which represents the total number of contacts between i and j over the time window $[0,T-1]$. The corresponding weighted adjacency matrix $W_{N\times N}$ has elements $W(i,j) = \sum_{t=0}^{t=T-1} \mathscr{A}(i,j,t)$.

2.2 *Temporal distance and topological distance between two contacts*

The contacts between two arbitrary nodes i and j can be regarded as the activation of the link $e(i,j) \in \mathscr{L}_W$ at the corresponding time stamps. The activity between i and j can be represented by a time series $X_{ij} = \{x_{ij}(t) = \mathscr{A}(i,j,t), t \in [0,T-1]\}$. The link $e(i,j)$ is active at time t if there is a contact between i and j at time t, i.e. $x_{ij}(t) = \mathscr{A}(i,j,t) = 1$. The total number of contacts in a network at each time stamp $Y = \{y(t) = \sum_{i,j\in\mathscr{N}, i<j} x_{ij}(t), t \in [0,T-1]\}$ reflects the global activity of the temporal network over time. The temporal distance between two contacts $\ell(i,j,t)$ and $\ell(k,l,s)$ is $\mathscr{T}(\ell(i,j,t), \ell(k,l,s)) = |t-s|$.

The topological distance, also called hopcount, between two nodes on a static network is the number of links contained in the shortest path between these two nodes. We define the topological distance $\eta(\ell(i,j,t), \ell(k,l,s))$ between two contacts $\ell(i,j,t)$ and $\ell(k,l,s)$ as the distance $\eta(e(i,j), e(k,l))$ between the corresponding two links $e(i,j)$ and $e(k,l)$ on the unweighted aggregated network, G_W. It can be derived as follows. The distance between the same link is zero, e.g. $\eta(e(i,j), e(i,j)) = 0$. The distance between two different links follows

$$\eta(e(i,j), e(k,l)) = \min_{u\in\{i,j\},\ v\in\{k,l\}} (h(u,v) + 1) \tag{2.1}$$

where $h(u,v)$ is the distance or hopcount between node u and v on the unweighted aggregated network G_W. The distance between two links is thus one plus the minimal distance between two end nodes of the two links. For example $\eta(e(i,j),e(i,k)) = 1$. Moreover, the line graph, e.g, G_W^L of a network G_W can be constructed by considering each link in G_W as a node, and two nodes are connected in G_W^L if the two corresponding links in G_W share a same end node. The distance (2.1) between two links in G_W equals the hopcount between their corresponding nodes in the line graph G_W^L.

2.3 *Network randomization -control methods*

In Section 4, we will explore diverse temporal-topological properties to understand the temporal and topological interrelations between contacts. However, real-world evolving networks may differ in, e.g., the number of nodes and the number contacts. In order to detect the non-trivial temporal-topological features and their interrelations in real-world networks, we compare each real-world network with its three controlled randomized networks which systematically preserve or remove specific topological and temporal correlation of contacts.

For a given temporal network $\mathscr{G}$, we introduce three randomized temporal networks $\mathscr{G}^1$, $\mathscr{G}^2$ and $\mathscr{G}^3$ respectively. Consider the set of contacts $\{\ell(i,j,t)\}$ in a temporal network $\mathscr{G}$, where each contact is described by its topological location, i.e., between pair of nodes (i,j) and its time stamp, t. Randomized network $\mathscr{G}^1$ is obtained by reshuffling the time stamps among the contacts, without changing the topological locations of the contacts. This randomization does not change the number of contacts between each node pair, only the timing is randomly changed, thus preserving the probability distribution of the topological distance of two randomly selected contacts. A temporal network can be also considered as an unweighted aggregated network and each link $e(i,j) \in \mathscr{L}_W$ is associated with its activity time series $\{\mathscr{A}(i,j,t), t \in [0,T-1]\}$. Randomized network $\mathscr{G}^2$ is obtained by iterating the step where two links are randomly selected from the aggregated network and their time series are swapped. This randomization does not change the distribution of the inter-event time of the activity of a random link, shown in Figures A.16 (virtual contacts) and A.17 (physical contacts). The third randomized network $\mathscr{G}^3$ is obtained by swapping the activity time series of two randomly selected links but with the same total number of contacts. This randomization preserves the number of contacts per node pair, the distribution of the inter-event time of contacts between a node pair and the distribution of the topological distance of two randomly selected contacts. The three randomized networks lead to the same unweighted aggregated network as the original network $\mathscr{G}$.

3. Datasets

All datasets of temporal networks are obtained from open access websites [1,2,3]. For each dataset, we consider nodes that belong to the largest connected component of the static aggregated network. The corresponding temporal network captures only the contacts between those nodes. Furthermore, we remove the long periods without any contact in the network, corresponding to e.g. night or weekend: we recognized these periods as outliers in the inter-event time [4] distribution of the global activity series

[1] http://www.sociopatterns.org/

[2] http://konect.uni-koblenz.de/

[3] https://snap.stanford.edu/data/index.html

[4] The inter-event time t_{ie} is the time interval between the occurrence of two consecutive events. A global activity time series Y with total number of events $k = \sum_{t=0}^{T} y(t)$ has $k-1$ inter-event times. If two events are contemporary, their corresponding inter-event time is 0.

Y that are far from the bulk. (see Figure 1). Finally, multiple contacts between the same pair of nodes at the same time step are accounted as a single contact. Details of the datasets are given in Table 1. In the original DNC Mail dataset [5], more than 96% of the total contacts forming the largest connected component occur in the last 33 days out of the 982 days. Hence, we include only the contacts of the last 33 days in our DNC Mail data.

Network	N	$\lvert\mathscr{L}_W\rvert$	$\lvert\mathscr{S}\rvert$	$\lvert\mathscr{C}\rvert$	T	dt	*contact type*
DNC Mail Part 2 (DNC_ 2 *) [35]	1598	4085	17300	30091	2861358	1	*virtual*
Manufacturing Email (ME*)[36]	167	3250	57791	82281	23430482	1	*virtual*
College Messages (CM*)[37]	1892	13833	58905	59789	16362751	1	*virtual*
Email EU (EEU*)[16, 38]	986	16025	206311	324933	44719809	1	*virtual*
Infectious (Infectious)[39]	410	2765	1392	17298	1421	20	*physical*
Primary School (PS)[40]	242	8317	3099	125771	3098	20	*physical*
High School 2012 (HS2012)[41]	180	2220	11267	45047	14114	20	*physical*
High School 2013 (HS2013)[42]	327	5818	7371	188504	7370	20	*physical*
Hypertext 2009 (HT2009)[39]	113	2196	5243	20818	7226	20	*physical*
SFHH Conference (SFHH)[43, 44]	403	9565	3508	70261	3799	20	*physical*
Workplace 2013 (WP)[45]	92	755	7095	9827	17844	20	*physical*
Workplace 2015 (WP2)[46]	217	4274	18479	78246	20946	20	*physical*
Hospital (Hospital)[47]	75	1139	9452	32424	16026	20	*physical*

Table 1: Basic features of the empirical networks after data processing. The number of nodes ($N = |\mathscr{N}|$), the number of links in $\mathscr{L}_W$ ($|\mathscr{L}_W|$), the number of snapshots ($|\mathscr{S}|$), the total number of contacts ($|\mathscr{C}|$), the length of the observation time window in time steps (T), the time resolution or duration of each time step (dt) in seconds and contact type are shown.

[5]http://konect.uni-koblenz.de/

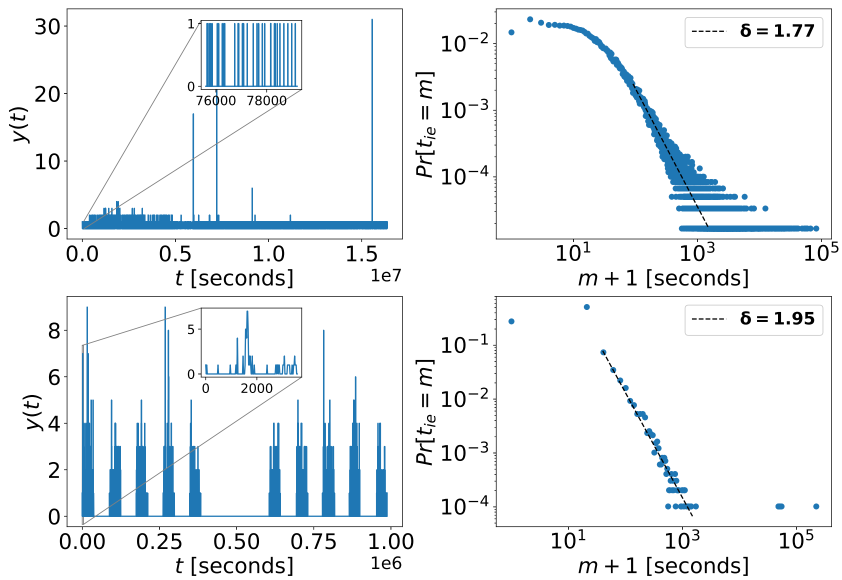

FIG. 1: Global activity (left) and its inter-event time distribution (right) in (a) physical contact network WP and (b) virtual contact network CM. The dashed line indicates the slope δ of the power-law fit and the scaling region, obtained via Clauset's method [48]. If the goodness of fit of the power law is significantly better than the exponential fit, the value of δ is reported in bold characters [6]. Time is expressed in seconds. Values of global activity are the total number of contacts occurred in each step of dt seconds. Insets in left figures show global activity for one hour. In WP, long time periods of null global activity correspond to night and weekend periods. These periods correspond to isolated outliers in the global inter-event time distribution with $m > 10^4 s$ and are removed in the data processing.

4. Characterizing topological-temporal properties of evolving networks

In this Section, we propose a systematic method to characterize topological-temporal properties of the contacts in an evolving network. In Subsection 4.1 we focus on the characterization of temporal properties, while in Section 4.2 and 4.3 we characterize the joint topological and temporal features of contacts.

4.1 *Temporal analysis of global activity*

The time series of global activity $Y = \{y(t), t \in [0, T-1]\}$ records the total number of contacts at each time step $t \in [0, T-1]$. In this Section, we aim to analyze the global activity time series in order to

[6] This evaluation is performed via the likelihood ratio test on power-law and exponential fit. If the test indicate a better performance of the power law with p-value $p < 0.05$, then the fitted exponent of power law δ is reported in bold characters.

discover whether the rate of contacts at different time steps are independent or correlated, irrespective of their location in the network.

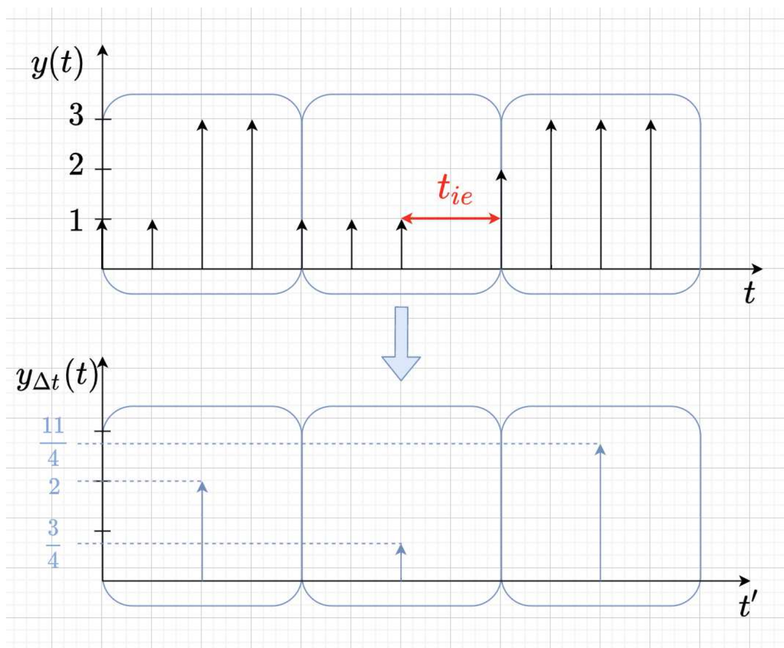

FIG. 2: Construction of the aggregated activity series $y_{\Delta t}(t')$ from the global activity time series $y(t)$, where $\Delta t = 4$ time steps. In the top sub-figure, we present the event sequence of $y(t)$, where each vertical line indicates the timing of one (or more -depending on thickness) event(s), while t_{ie} is the inter-event time and two events happening at the same time have an inter-event time zero.

We aggregate the global activity at each time window of duration Δt time steps as follows. The time steps $t \in [0, T-1]$ can be divided into a set of non-overlapping consecutive time windows/bins of duration Δt. The aggregated activity $\mathbf{y}_{\Delta t}(t')$ at a time period $[t'\Delta t, t'\Delta t + \Delta t)$ is the average activity of $\{y(t), t \in [t'\Delta t, t'\Delta t + \Delta t)\}$ where $0 \leqslant t' \leqslant \left\lfloor \frac{T-1}{\Delta t} \right\rfloor - 1$, see Figure 2.

To understand whether the number of contacts at each time step is independent on other time steps or not, we explore the variance of the aggregated activity time series $Var[Y_{\Delta t}]$ normalized by the variance of the un-aggregated $Var[Y]$ as a function of Δt (see Figure 3). First, we will derive how the normalized $\frac{Var[Y_{\Delta t}]}{Var[Y]}$ is related to the aggregation resolution Δt. The global activity Y can be considered as a realization of a set of T random variables $\{\hat{Y}(t)\}$, that are identically distributed as a variable $\hat{Y}$. Hence, $Var[\hat{Y}(t)] = Var[\hat{Y}]$, where $t \in [0, T-1]$. The aggregated activity $\hat{Y}_{\Delta t}(t')$ at a random period t' with $0 \leqslant t' \leqslant \left\lfloor \frac{T-1}{\Delta t} \right\rfloor - 1$ is $\hat{Y}_{\Delta t}(t') = \frac{1}{\Delta t} \sum_{t=t'\Delta t}^{t'\Delta t + \Delta t - 1} \hat{Y}(t)$ and its variance follows

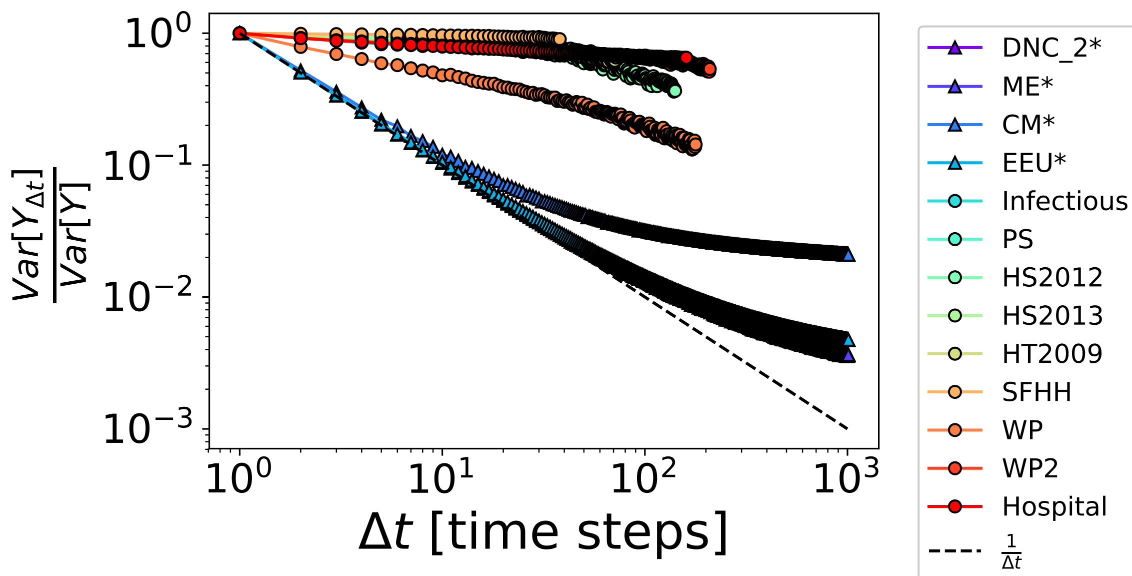

FIG. 3: The normalized variace $\frac{Var[Y_{\Delta t}]}{Var[Y]}$ as a function of the aggregation resolution Δt. Circles correspond to physical contact temporal networks, triangles correspond to virtual contact networks (online messages and mail), while the black dashed line ($\frac{\mathrm{Var}[Y_{\Delta t}]}{\mathrm{Var}[Y]} = \frac{1}{\Delta t}$) represents the uncorrelated curve. The resolution Δt is in units of time steps.

$$\begin{aligned}
\mathrm{Var}\left[\hat{Y}_{\Delta t}(t')\right] &= \mathrm{Var}\left[\frac{1}{\Delta t}\sum_{t=t'\Delta t}^{t'\Delta t+\Delta t-1}\hat{Y}(t)\right] \\
&= \frac{1}{(\Delta t)^2}\sum_{t=t'\Delta t}^{t'\Delta t+\Delta t-1}\left(\mathrm{Var}[\hat{Y}(t)] + 2\sum_{t'\Delta t\leqslant s<k\leqslant t'\Delta t+\Delta t-1}\mathrm{Cov}[\hat{Y}(s),\hat{Y}(k)]\right) \qquad (4.1) \\
&= \frac{\mathrm{Var}[\hat{Y}]}{\Delta t} + \frac{2}{(\Delta t)^2}\sum_{t=t'\Delta t}^{t'\Delta t+\Delta t-1}\ \sum_{t'\Delta t\leqslant s<k\leqslant t'\Delta t+\Delta t-1}\mathrm{Cov}[\hat{Y}(s),\hat{Y}(k)]
\end{aligned}$$

When the activity $\hat{Y}(t)$ at each time t is independently distributed, i.e. the set $\{\hat{Y}(t)\}$ are independent, the second term is zero and we have $\frac{\mathrm{Var}[\hat{Y}_{\Delta t}(t')]}{\mathrm{Var}[\hat{Y}]} = \frac{1}{\Delta t}$. This explains why $\frac{\mathrm{Var}[Y_{\Delta t}]}{\mathrm{Var}[Y]} = \frac{1}{\Delta t}$ in Figure 3 when we randomly re-shuffle the global activity $Y = \{y(t), t \in [0, T-1]\}$ in each of the thirteen temporal networks. Figure 3 shows that the normalized variance of all real-world temporal networks are well above the uncorrelated curve $\frac{\mathrm{Var}[Y_{\Delta t}]}{\mathrm{Var}[Y]} = \frac{1}{\Delta t}$, suggesting the existance of high correlation among the activity or number of contacts per time step at different times. Moreover, it is seen that the physical contact networks are further away from $\frac{\mathrm{Var}[Y_{\Delta t}]}{\mathrm{Var}[Y]} = \frac{1}{\Delta t}$ compared to the virtual contact networks, reflecting higher correlation in physical contacts than in virtual activities. The higher correlation in global activity in physical contacts networks than in virtual networks seems to be supported by the relatively higher probability for the global inter-event time to be relatively small in physical contact networks (see Figure A.12) than that in virtual contact ones (see Figure A.11).

4.2 *Topological and temporal distances between two contacts*

Next we wish to explore the relation between the topological distance and temporal distance of two contacts. Firstly, we will explore whether contacts that are close in topology (time) are also close in time (topology). Contacts of temporal networks are measured at discrete time steps. The duration of each time step is either 1 or 20 seconds in the datasets listed in Table 1. To compare between physical and virtual contact networks, we present the time distance between any two contacts in units of seconds.

The probability distribution $Pr[\mathscr{T}(\ell,\ell')=m]$ of the temporal distance $\mathscr{T}(\ell,\ell')$ between two randomly selected contacts is given in Appendix (Figures A.14 and A.15). The probability distribution $\Pr[\eta(\ell,\ell')=j]$ of the topological distance $\eta(\ell,\ell')$ of two randomly selected contacts is given for each real-world temporal network in Appendix (Figure A.13).

We analyze the average topological distance $E[\eta(\ell,\ell')|\mathscr{T}(\ell,\ell')<\Delta t]$ of two contacts given that their temporal distance is less than Δt. Figures 4 and 5 show that $E[\eta(\ell,\ell')|\mathscr{T}(\ell,\ell')<\Delta t]$ increases with Δt in real-world temporal networks. That is, contacts that are close in time are typically also close in topology. Such an increasing trend or correlation between temporal and topological distances in each real-world temporal network is evidently higher than that in the corresponding three randomized networks. Network $\mathscr{G}^3$ (swapping the activity time series of the two randomly selected links but with the same total number of contacts) preserves more information of the original temporal network compared to $\mathscr{G}^1$ (swapping timestamps among contacts) and $\mathscr{G}^2$ (swapping the activity time series of two randomly selected links). Note that the swapping of time series of two node pairs with the same number of contacts in $\mathscr{G}^3$ already reduces significantly the tendency that contacts appear close in time and topology. Such temporal and topological distance correlation is reduced further in $\mathscr{G}^2$. The slight increase of $E[\eta(\ell,\ell')|\mathscr{T}(\ell,\ell')<\Delta t]$ with Δt in $\mathscr{G}^2$ is likely because $\mathscr{G}^2$ preserved the same inter-event times of consecutive contacts between a node pair as in $\mathscr{G}$ such that within a small time interval Δt many contacts may occur between the same couple of users, resulting in a smaller average distance. Temporal and topological distances correlation fully disappears in $\mathscr{G}^1$. The increase of $E[\eta(\ell,\ell')|\mathscr{T}(\ell,\ell')<\Delta t]$ with Δt in $\mathscr{G}$, in comparison with that in $\mathscr{G}^2$, is more significant in virtual contact networks and physical contact networks Infectious. In these networks, contacts that occur close in time tend to be close in topology. The strong correlation in network Infectious is related to the specific properties of this network. Network Infectious records the contacts among visitors of a museum. Only people that visit the museum at a similar time could have contact and be connected in the aggregated topology thus close in topology[39].

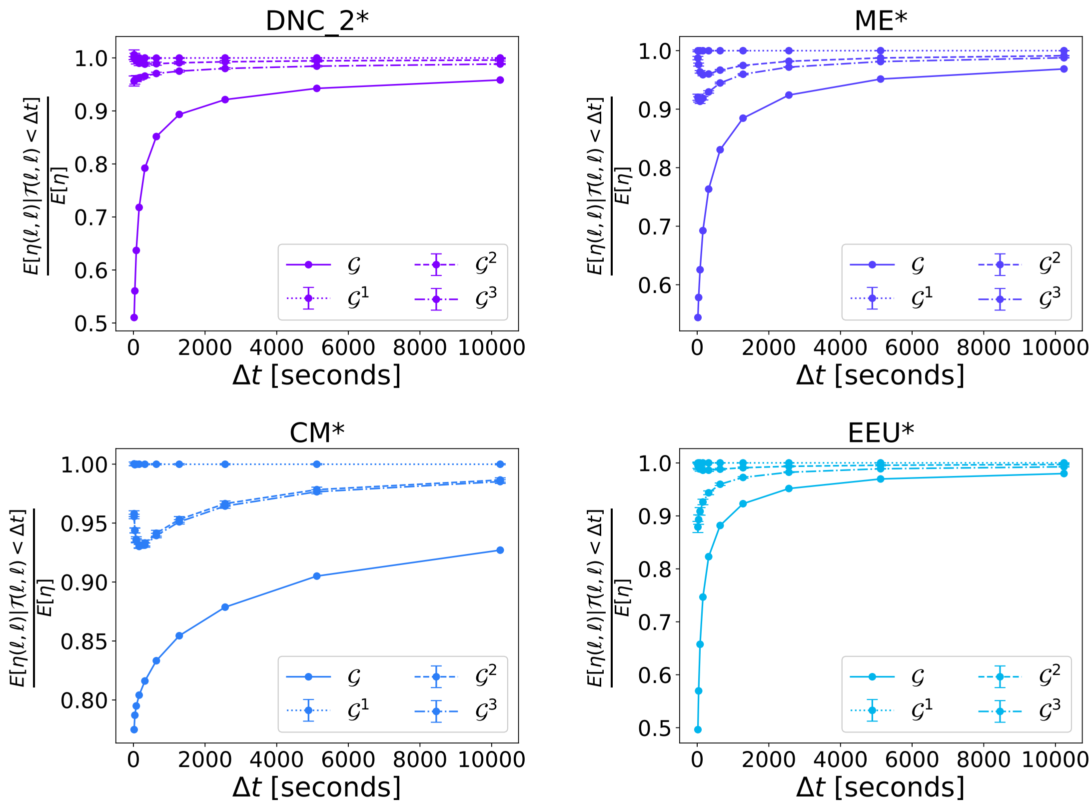

FIG. 4: $\frac{E[\eta(\ell,\ell')|\mathscr{T}(\ell,\ell')<\Delta t]}{E[\eta]}$ as a function of Δt in each virtual contact dataset. When $\frac{E[\eta(\ell,\ell')|\mathscr{T}(\ell,\ell')<\Delta t]}{E[\eta]} = 1$, topological and temporal distances are independent. Moreover, $\lim_{\Delta t\to\infty} E[\eta(\ell,\ell')|\mathscr{T}(\ell,\ell') < \Delta t] = E[\eta(\ell,\ell')]$. The results for each of the three randomized networks are obtained from 10 independent realizations of the randomized network.

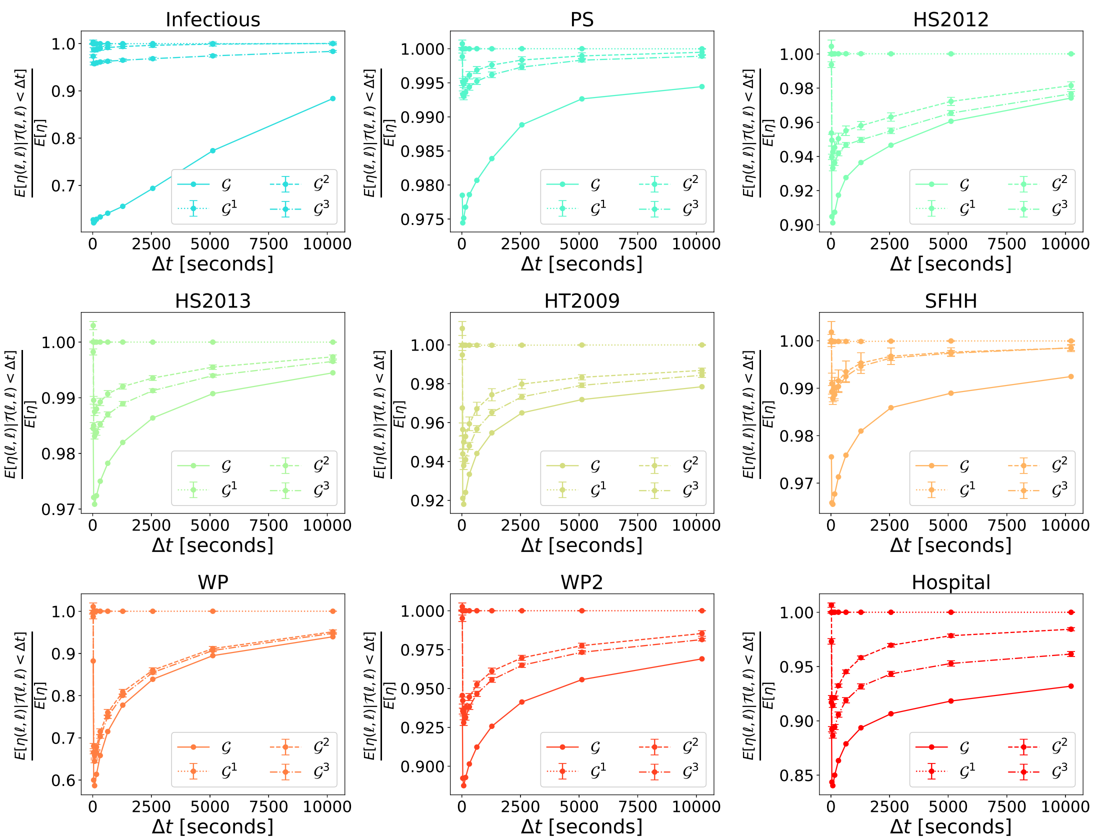

FIG. 5: $\frac{E[\eta(\ell,\ell')|\mathcal{T}(\ell,\ell')<\Delta t]}{E[\eta]}$ as a function of Δt in each physical contact dataset. For each of the three randomized models, the lines and corresponding error bars correspond to the average and standard deviation of the results obtained from 10 independent realizations.

We could also identify the temporal and topological correlation of contacts via $E[\mathcal{T}(\ell,\ell')|\eta(\ell,\ell') = j]$, the average temporal distance of two contacts given that their topological distance is j. However, this measure could be limited in distinguishing the difference among networks due to the small diameter, i.e., the maximal hopcount of real-world networks (see A.13).

4.3 *Temporal correlation of local events*

As explained in Section 2.2, a temporal network can be considered as an aggregated network, where each link $e(i,j) \in \mathcal{L}_W$ is associated with its activity time series $\{\mathcal{A}(i,j,t), t \in [0, T-1]\}$. In previous sections we show a non-trivial relation between the topological and temporal distance of contacts. In particular, contacts with a smaller topological distance tend to be closer in time, and vice versa. In this section, we explore the temporal correlation of the contacts in a local neighborhood such as a link and the ego network $ego(e(i,j))$ of a link $e(i,j)$ which consists of the link itself and all its neighboring links that share with $e(i,j)$ a common end node, either node i or j. The activity (sequence) in $ego(e(i,j))$ refers to the corresponding time series or time stamps of the contacts that associate with the links in

$ego(e(i,j))$. The topological distance of two contacts from the same ego network is maximally 2.

4.3.1 *Temporal correlation of contacts at an ego network* If a contact along a link influences or correlates with the contact(s) of its neighboring links, the temporal properties of the activity in $ego(e(i,j))$ in $\mathscr{G}$ are expected to differ from those of the activity in $ego(e(i,j)$ in the randomized networks, $\mathscr{G}^1$, $\mathscr{G}^2$ and $\mathscr{G}^3$. Specifically, we study the size distribution of event trains [18] of the activity (sequence) in an ego-network. A train of bursty events is a sequence of consecutive contacts/events whose inter-event times are shorter or equal than Δt and separated from the other contacts by an inter-event times larger than Δt. Given a Δt, trains can be identified for a given activity sequence. Given Δt and a temporal network, the train size distribution $Pr[\mathscr{E}_{\Delta t} = s]$ is derived from the activity sequences of the ego networks centered at every link. The train size distribution is compared between each real-world network and its three randomized networks.

The choice of Δt is not trivial. Consider firstly the simple case where the same Δt is chosen for a real-world network and its corresponding randomized networks. Karsai et al. [18] have observed a power law distribution $Pr[\mathscr{E}_{\Delta t} = s] \sim s^{-\beta}$ of the activity of a link in real-world networks, with $\beta = 0.39$ (0.42) for voice calls (SMS). This suggests temporal correlation of contacts between a node pair, since the train size distribution should follow an exponential function $Pr[\mathscr{E}_{\Delta t} = s] = Pr[t_{ie} \leqslant \Delta t]^{(s-1)}(1 - Pr[t_{ie} \leqslant \Delta t])$ when the inter-event times in the activity sequence are independent. Moreover, they found that the power-law exponent remains approximately the same when Δt varies within a broad range.

Our comparison of train size distribution with different Δt for virtual (Figure A.18) and physical contact datasets (Figure A.19) shows that when Δt is small ($\Delta t \leqslant 120s$), the distribution is approximately a power law. The exponent β of the power-laws seems more stable across different values of Δt in virtual contacts (FigureA.18) datasets than in physical contact ones (Figure A.19). The changes in the shape of the train size distribution of physical contact datasets are likely due to finite size effects which emerge because of limited duration of empirical temporal networks' observation window. When Δt is sufficiently large, for example, any ego network has a single train, whose value is the total number of contacts of the links belonging to the ego network. Figures 5 and 4 show that the positive correlation between topological and time separation of two successive contacts is more evident when the time separation is small. Moreover, the observation time windows of temporal networks, especially physical contact networks, are short in duration. All these perspectives motivate us to consider a small Δt, e.g. $\Delta t = 60s$. Moreover, our observations are similar when $\Delta t = 120s$ and when $\Delta t = 60s$ for all the analysis to be introduced. Hence, we focus our discussion on $\Delta t = 60s$ and all the results when $\Delta t = 120s$ are given in the appendix.

The distribution of the ego-networks in the real-world network has an evidently higher tail than that of the corresponding randomized networks in the four real-world virtual contact network (Figure 6) and the physical contact network Infectious (Figure 7). In these real-world networks, the activity sequences of ego-networks have a higher chance to form long trains, reflecting the stronger temporal correlation of local contacts than in their corresponding randomized networks. Randomized network $\mathscr{G}_2$ is obtained by shuffling the activity sequences among the links, thus preserving the activity sequences but removing their correlation with the network topology. The difference between the train size distribution in the ego-networks of $\mathscr{G}_2$ and an exponential distribution reflects solely the temporal correlation among the activity sequence of each link in real-world networks.

The different train size distributions in real-world networks $\mathscr{G}$ and their corresponding randomized networks $\mathscr{G}_2$ in Figures 6 and 7 indicate that temporal correlation of activities at each link is insufficient to explain the temporal correlation of contacts at ego-networks. Instead, the correlation between the activity sequences and topology also contributes. Such temporal correlation of local activities suggests that neighboring nodes tend to have contacts or activities within a short time. The evidently stronger correlation observed in virtual networks and the physical network Infectious is in line with the finding in Section 4.2.

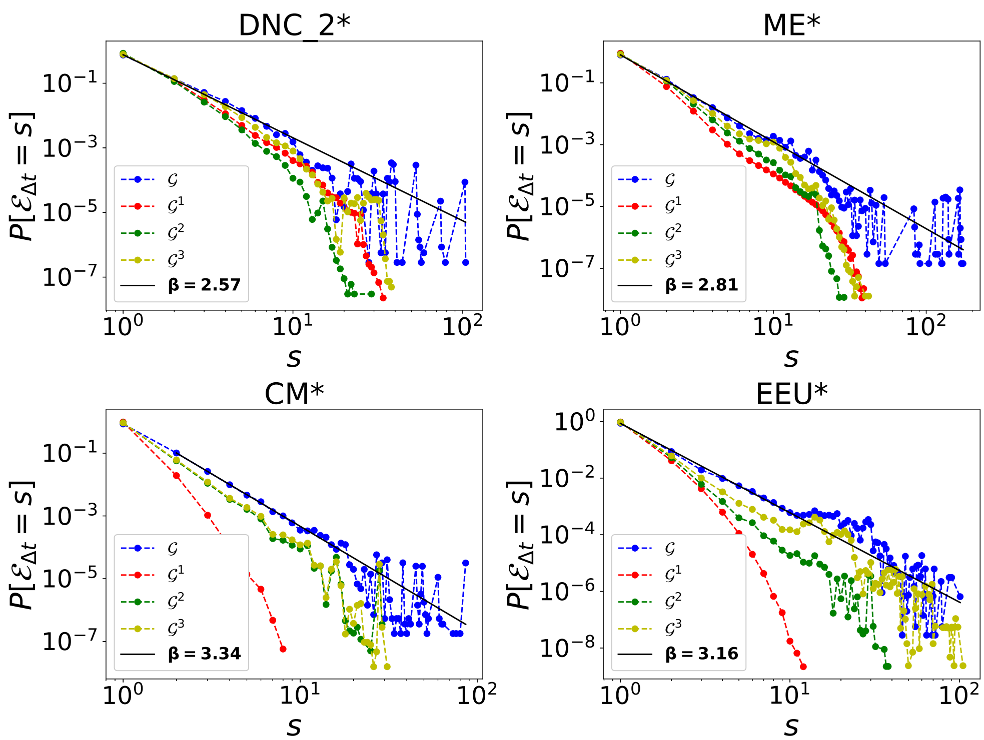

FIG. 6: Train size distribution ($\Delta t = 60s$) of ego network activity for $\mathscr{G}$ (blue), $\mathscr{G}_1$ (red), $\mathscr{G}_2$ (green), $\mathscr{G}_3$ (yellow) of virtual contact datasets. The black solid line represents the fit to the distribution $P[\mathscr{E}_{\Delta t} = s] \sim s^{-\beta}$, where β is the exponent of the power law fit of the train size distribution of $\mathscr{G}$. The power law fit and the scaling region were computed with Clauset's method [48]. If the goodness of fit of the power law is significantly better than the exponential fit (likelihood ratio test with p-value $p < 0.05$), the value of β is reported in bold characters. Note that the horizontal and vertical axes are presented in logarithmic scales.

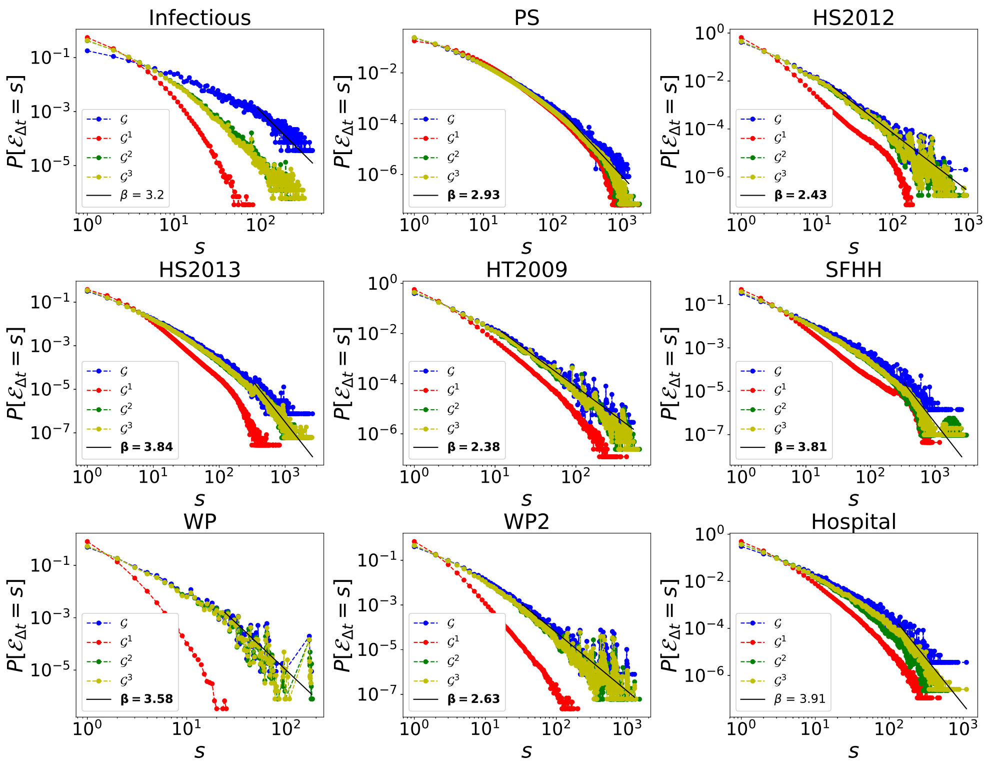

FIG. 7: Train size distribution ($\Delta t = 60s$) of ego network activity for $\mathscr{G}$ (blue), $\mathscr{G}_1$ (red), $\mathscr{G}_2$ (green), $\mathscr{G}_3$ (yellow) of physical contact datasets. The black solid line represents the fit to the distribution $P[\mathscr{E}_{\Delta t} = s] \sim s^{-\beta}$, where β is the exponent of the power law fit of the train size distribution of $\mathscr{G}$. The power law fit and the scaling region were computed with Clauset's method [48]. If the goodness of fit of the power law is significantly better than the exponential fit (likelihood ratio test with p-value $p < 0.05$), the value of β is reported in bold characters. Note that the horizontal and vertical axes are presented in logarithmic scales.

4.3.2 *Ego network activity versus link activity* We investigate further whether the temporal correlation of contacts at an ego network could be explained or introduced by the temporal correlation of the contacts within a link.

Firstly, we explore whether each activity train of an ego network contains the activities of a single link or of multiple links. We examine the number $\mathscr{M}$ of distinct active links that a train of an ego-network involves. Specifically, each identified train of an ego network is composed of a set of contacts, occurring at a subset of links within the ego-network, the so-called active links. For each real-world network and given $\Delta t = 60s$, trains are identified for every ego network centered at each link, and the number $\mathscr{M}$ of distinct active links of each train is counted. Figure 8 illustrates the average number of active links $\frac{E[\mathscr{M}|\mathscr{E}_{\Delta t}=s]}{s}$ for trains with size $\mathscr{E}_{\Delta t} = s$, normalized by the train size s, for virtual and physical contact networks, respectively. In all networks the fraction of active links $\frac{E[\mathscr{M}|\mathscr{E}_{\Delta t}=s]}{s}$ is above $\frac{E[\mathscr{M}|\mathscr{E}_{\Delta t}=s]}{s} = 1/s$ suggesting that a train usually involves far more than 1 active link.Interestingly, we observe in all 9 physical contact networks a seemingly power-law decay $\frac{E[\mathscr{M}|\mathscr{E}_{\Delta t}=s]}{s} \sim s^{-\alpha}$ (right plot of Figure 8). In contrast, $\alpha \approx 0$, or equivalently $E[\mathscr{M}|\mathscr{E}_{\Delta t} = s] \sim s$ in virtual contact networks, especially mail dataset, i.e. EEU, ME and DNC2 (left plot of Figure 8). This suggests that, in virtual contact networks, each train is mostly composed of the activities of many links in an ego network.

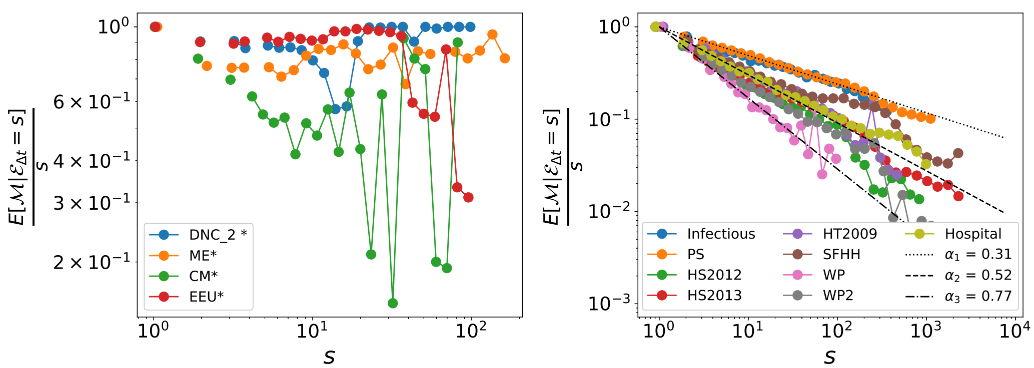

FIG. 8: The average number of active links $\frac{E[\mathscr{M}|\mathscr{E}_{\Delta t}=s]}{s}$ for trains with size $\mathscr{E}_{\Delta t} = s$, normalized by the train size s of the ego networks for virtual (left) and physical (right) contact datasets, when $\Delta t = 60s$. The three reference lines in right plot indicate $\frac{E[\mathscr{M}|\mathscr{E}_{\Delta t}=s]}{s} = s^{-\alpha}$ with slope $\alpha_1 = 0.31$ (dotted), $\alpha_2 = 0.52$ (dashed) and $\alpha_3 = 0.77$ (dash-dot). Note that the horizontal and vertical axes are presented in logarithmic scales. In total 30 logarithmic bins are split within the interval $[1, s_{max}]$, where s_{max} is the largest train size observed in the considered real temporal network.

We compare further the physical contact networks. Their power-law exponents are within $0.31 \leqslant \alpha \leqslant 0.77$. The slope of the power-law decay seems to be influenced by the type of human interaction and spacial constraints of the contact environment. Networks that lead to the slowest decay, i.e. $\alpha \approx 0.31$ are Infectious and PS datasets, which are contact networks in a museum and primary school respectively. The two contact networks of employees at a work place, WP and WP2 have the largest slope $\alpha \approx 0.77$. The other networks, i.e., contacts of high school students, conference participants have a power-law exponent in between $0.31 \leqslant \alpha \leqslant 0.77$. A similar trend has been observed when $\Delta t = 120s$ (see Figure A.24 in Appendix). These observations could be explained by the spatial constraints of contacts and the nature that younger students tend to interact with many others in an active period. The bursty events of a train tend to engage the largest number of links in an ego network in network Infectious and PS than the other physical contact networks. This could be due to the freedom for individuals to move in the museum and in the primary school (relative to the small museum/class room) and the tendency that primary school students interact with many others in an active period. The other way round, employees at a work place are confined in space (their offices) and tend to interact with limited number of colleagues during a train of activities. In this sense, virtual contacts are the least confined to space, leading thus to a larger number of active links than physical contacts.

Whether each activity train of an ego network contains the activities of a single link or of multiple links could also be reflected via the train size distribution in an ego network versus the train size distribution in a link. In Figures 9 and 10, we compare the train size distribution (with $\Delta t = 60s$) of the activity sequence of single links, of the most active single links (top 10% of links with the largest number of contacts) and of ego-networks. The trains of ego-networks tend to be longer than those of single links and the most active single links, in all networks except for WP. Therefore, the trains of the ego network are usually the results of the activity of more than one link. The same observations are obtained when $\Delta t = 120s$ (see Figures A.22 and A.23 in Appendix). The similar train size distribution in ego networks and in links in the WP dataset is consistent with the largest power-exponent observed in Figure 8. In WP, a train of bursty in an ego network events involve the activity of few links.

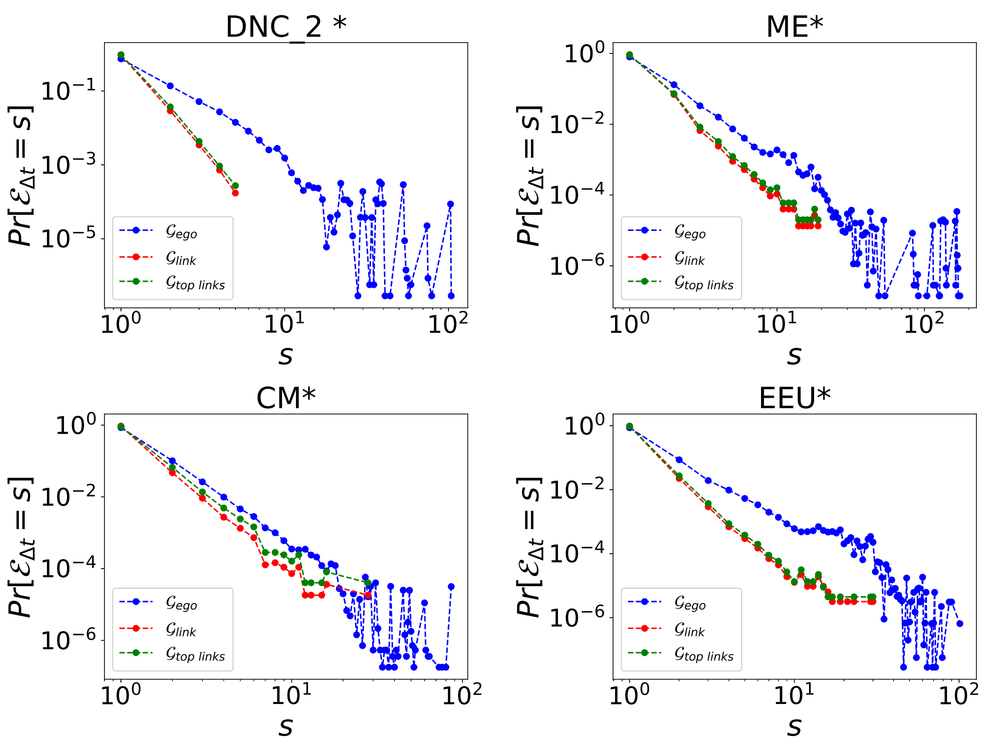

FIG. 9: Train size distribution ($\Delta t = 60s$) of ego network activity (blue), single link activity (red), most active link activity (green) of virtual contact datasets. Note that the horizontal and vertical axes are presented in logarithmic scales.

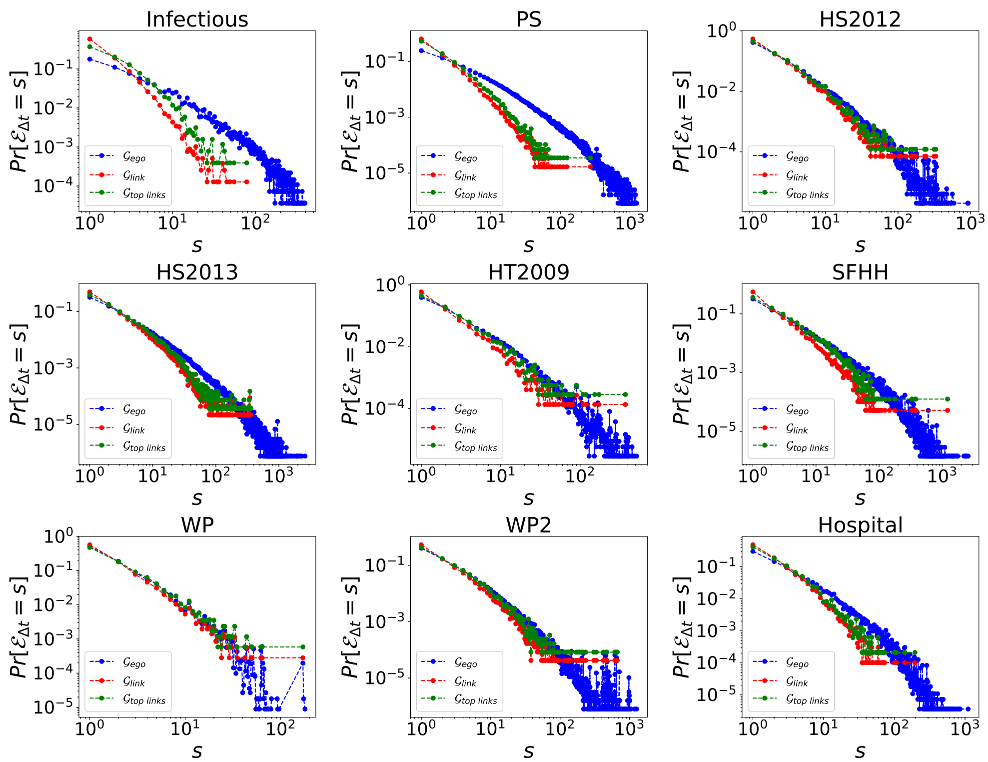

FIG. 10: Train size distribution ($\Delta t = 60s$) of ego-network activity (blue), single link activity (red), most active link activity (green) of physical contact datasets. Note that the horizontal and vertical axes are presented in logarithmic scales.

4.3.3 *Ego network activity versus node activity* We further address the question whether a train at an ego network $ego(e(i,j))$ involves the activity of both end nodes i and j, or only one of them. Event trains at ego networks engaging activities of both end nodes may suggest a possible social contagion in activity between nodes.

For each train of an ego network $ego(e(i,j))$, we consider the events that associate with only one end node but not both. Among these events, we count the fraction of events ϕ_i and ϕ_j that associate with end node i and j respectively and $\phi_i + \phi_j = 1$. The maximum of the two fractions $B = max(\phi_i, \phi_i)$ quantifies how unbalanced the activities of the two end nodes i and j contribute to a train and is called the activity balance of a train.

Table 2 shows the average activity balance $E[B]$ and the probability $Pr[B \leqslant 0.95]$ of the activity balance for all contact networks, accounting all trains (whose sizes are larger than 1) of all ego networks. We find that the average activity $E[B] < 1$, suggesting that an activity train in an ego network $ego(e(i,j))$ engages in the activity of both end nodes i and j. This is in line with the previous finding that the activity correlation in an ego network cannot be explained by the activity of a single link. Moreover, the activity is found to be larger, thus more unbalanced, in virtual contacts than in physical contact networks. This is reasonable because, in a virtual contact network like email contact network, an individual tends to contact many others at a similar time. A train of events at an ego network $ego(e(i,j))$ in a virtual network contains mainly the activities of a single end node i or j.

Virtual Contacts			Physical Contacts		
Dataset	$P[B \leqslant 0.95]$	$E[B]$	Dataset	$P[B \leqslant 0.95]$	$E[B]$
DNC 2*	0.05	0.98	Infectious	0.38	0.87
ME*	0.12	0.95	PS	0.44	0.86
CM*	0.03	0.99	HS2012	0.16	0.95
EEU*	0.12	0.94	HS2013	0.25	0.93
			HT2009	0.21	0.94
			SFHH	0.19	0.95
			WP	0.06	0.98
			WP2	0.1	0.97
			Hospital	0.12	0.97

Table 2: Probability $P[B \leqslant 0.95]$ and average $E[B]$ of the activity balance B in virtual contact (left) and physical contact (right) networks.

5. Conclusions

In this paper, we developed systematically methods to characterize jointly the topological and temporal properties of contacts in a time-evolving network, ranging from global network level to local neighborhoods. Via applying these methods to real-world networks, we identified substantial differences between virtual and physical contact networks.

We find that contacts that occur close in time tend to be close in topology and this trend is more evident in virtual contact networks compared to physical contact networks. This is in line with the observation that the contacts of ego-networks tend to have a higher chance to form long trains and thus happen closely in time in real-world networks. Such activity correlation is significantly more evident in virtual contact networks. Moreover, an event train of an ego network $ego(e(i,j))$ is mostly composed of the activities of multiple component links. Interestingly, more links tend to be engaged in e.g., virtual networks and physical contact network primary school where the contacts are less constrained in space, in contrast to e.g., the contact network at workplace. These may suggest that contacts with a low cost may better facilitate social contagion, i.e. influence of neighboring nodes in the activity. Finally, an event train of an ego network $ego(e(i,j))$ usually contains the activity of both ends, node i and j. Two connected nodes, thus, tend to have contacts with their neighbors close in time. The two end nodes' contributions are more unbalanced in virtual contacts than in physical contacts, likely driven by the nature that in a virtual (e.g. email) contact network, an individual tends to contact many others close in time.

Our methods are confined to undirected networks. A full-fledged directed temporal network characterization method is deemed as promising to develop. The application of these methods may enhance our understanding of diverse time-evolving systems and allow exploration of the influence of detected properties/patterns on a dynamic process upon the network. Finally, the detected patterns may further inspire the development of more realistic temporal network models that reproduce key realistic temporal and topological properties of contacts.

References

[1] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, 2012.

[2] Petter Holme. Modern temporal network theory: a colloquium. *The European Physical Journal B*, 88(9):234, 2015.

[3] K-I Goh and A-L Barabási. Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, 81(4):48002, 2008.

[4] Jean-Pierre Eckmann, Elisha Moses, and Danilo Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *Proceedings of the National Academy of Sciences*, 101(40):14333–14337, 2004.

[5] Joao Gama Oliveira and Albert-László Barabási. Darwin and einstein correspondence patterns. *Nature*, 437(7063):1251–1251, 2005.

[6] Julián Candia, Marta C González, Pu Wang, Timothy Schoenharl, Greg Madey, and Albert-László Barabási. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224015, 2008.

[7] Anders Johansen. Probing human response times. *Physica A: Statistical Mechanics and its Applications*, 338(1-2):286–291, 2004.

[8] Albert-László Barabási and Eric Bonabeau. Scale-free networks. *Scientific American*, 288(5): 60–69, 2003.

[9] Albert-László Barabási. Scale-free networks: a decade and beyond. *Science*, 325(5939):412–413, 2009.

[10] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[11] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, 2006.

[12] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11): 3747–3752, 2004.

[13] J-P Onnela, Jari Saramäki, Jorkki Hyvönen, György Szabó, David Lazer, Kimmo Kaski, János Kertész, and A-L Barabási. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336, 2007.

[14] Hilla Brot, Lev Muchnik, Jacob Goldenberg, and Yoram Louzoun. Evolution through bursts: Network structure develops through localized bursts in time and space. *Network Science*, 4(3): 293–313, 2016.

[15] Riivo Kikas, Marlon Dumas, and Márton Karsai. Bursty egocentric network evolution in skype. *Social Network Analysis and Mining*, 3(4):1393–1401, 2013.

[16] Ashwin Paranjape, Austin R Benson, and Jure Leskovec. Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 601–610, 2017.

[17] Lauri Kovanen, Márton Karsai, Kimmo Kaski, János Kertész, and Jari Saramäki. Temporal motifs in time-dependent networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(11): P11005, 2011.

[18] Márton Karsai, Kimmo Kaski, and János Kertész. Correlated dynamics in egocentric communication networks. *Plos One*, 7(7):e40612, 2012.

[19] Raj Kumar Pan and Jari Saramäki. Path lengths, correlations, and centrality in temporal networks. *Physical Review E*, 84(1):016105, 2011.

[20] Albert-Laszlo Barabasi. The origin of bursts and heavy tails in human dynamics. *Nature*, 435 (7039):207–211, 2005.

[21] Alexei Vazquez, Balazs Racz, Andras Lukacs, and Albert-Laszlo Barabasi. Impact of non-poissonian activity patterns on spreading processes. *Physical Review Letters*, 98(15):158702, 2007.

[22] Diego Rybski, Sergey V Buldyrev, Shlomo Havlin, Fredrik Liljeros, and Hernán A Makse. Scaling laws of human interaction activity. *Proceedings of the National Academy of Sciences*, 106(31): 12640–12645, 2009.

[23] Márton Karsai, Kimmo Kaski, Albert-László Barabási, and János Kertész. Universal features of correlated bursty behaviour. *Scientific Reports*, 2:397, 2012.

[24] Xiu-Xiu Zhan, Alan Hanjalic, and Huijuan Wang. Information diffusion backbones in temporal networks. *Scientific Reports*, 9(1):1–12, 2019.

[25] Xiu-Xiu Zhan, Alan Hanjalic, and Huijuan Wang. Suppressing information diffusion via link blocking in temporal networks. In *International Conference on Complex Networks and Their Applications*, pages 448–458. Springer, 2019.

[26] René Pfitzner, Ingo Scholtes, Antonios Garas, Claudio J Tessone, and Frank Schweitzer. Betweenness preference: Quantifying correlations in the topological dynamics of temporal networks. *Physical Review Letters*, 110(19):198701, 2013.

[27] Giovanna Miritello, Esteban Moro, and Rubén Lara. Dynamical strength of social ties in information spreading. *Physical Review E*, 83(4):045102, 2011.

[28] Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, Jari Saramäki, and Márton Karsai. Multiscale analysis of spreading in a large communication network. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(03):P03005, 2012.

[29] Ingo Scholtes, Nicolas Wider, René Pfitzner, Antonios Garas, Claudio J Tessone, and Frank Schweitzer. Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nature Communications*, 5(1):1–9, 2014.

[30] Oliver E Williams, Fabrizio Lillo, and Vito Latora. How auto-and cross-correlations in link dynamics influence diffusion in non-markovian temporal networks. *arXiv preprint arXiv:1909.08134*, 2019.

[31] Ville-Pekka Backlund, Jari Saramäki, and Raj Kumar Pan. Effects of temporal correlations on cascades: Threshold models on temporal networks. *Physical Review E*, 89(6):062815, 2014.

[32] Roni Parshani, Mark Dickison, Reuven Cohen, H Eugene Stanley, and Shlomo Havlin. Dynamic networks and directed percolation. *EPL (Europhysics Letters)*, 90(3):38004, 2010.

[33] Dávid X Horváth and János Kertész. Spreading dynamics on networks: the role of burstiness, topology and non-stationarity. *New Journal of Physics*, 16(7):073037, 2014.

[34] Jean-Charles Delvenne, Renaud Lambiotte, and Luis EC Rocha. Diffusion on networked systems is a question of time or structure. *Nature Communications*, 6(1):1–10, 2015.

[35] Jérôme Kunegis. Konect: the koblenz network collection. In *Proceedings of the 22nd international conference on World Wide Web*, pages 1343–1350, 2013.

[36] Radosław Michalski, Sebastian Palus, and Przemysław Kazienko. Matching organizational structure and social network extracted from email communication. In *Lecture Notes in Business Information Processing*, volume 87, pages 197–206. Springer Berlin Heidelberg, 2011.

[37] Pietro Panzarasa, Tore Opsahl, and Kathleen M Carley. Patterns and dynamics of users' behavior and interaction: Network analysis of an online community. *Journal of the American Society for Information Science and Technology*, 60(5):911–932, 2009.

[38] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution: Densification and shrinking diameters. *ACM transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2–es, 2007.

[39] Lorenzo Isella, Juliette Stehlé, Alain Barrat, Ciro Cattuto, Jean-François Pinton, and Wouter Van den Broeck. What's in a crowd? analysis of face-to-face behavioral networks. *Journal of theoretical biology*, 271(1):166–180, 2011.

[40] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Lorenzo Isella, Jean-François Pinton, Marco Quaggiotto, Wouter Van den Broeck, Corinne Régis, Bruno Lina, et al. High-resolution measurements of face-to-face contact patterns in a primary school. *PloS one*, 6(8):e23176, 2011.

[41] Julie Fournet and Alain Barrat. Contact patterns among high school students. *PloS one*, 9(9):e107878, 2014.

[42] Rossana Mastrandrea, Julie Fournet, and Alain Barrat. Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PloS one*, 10(9):e0136497, 2015.

[43] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one*, 5(7):e11596, 2010.

[44] Juliette Stehlé, Nicolas Voirin, Alain Barrat, Ciro Cattuto, Vittoria Colizza, Lorenzo Isella, Corinne Régis, Jean-François Pinton, Nagham Khanafer, Wouter Van den Broeck, et al. Simulation of an seir infectious disease model on the dynamic contact network of conference attendees. *BMC medicine*, 9(1):1–15, 2011.

[45] Mathieu Génois, Christian L Vestergaard, Julie Fournet, André Panisson, Isabelle Bonmarin, and Alain Barrat. Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. *Network Science*, 3(3):326–347, 2015.

[46] Mathieu Génois and Alain Barrat. Can co-location be used as a proxy for face-to-face contacts? *EPJ Data Science*, 7(1):1–18, 2018.

[47] Philippe Vanhems, Alain Barrat, Ciro Cattuto, Jean-François Pinton, Nagham Khanafer, Corinne Régis, Byeul-a Kim, Brigitte Comte, and Nicolas Voirin. Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PloS one*, 8(9):e73970, 2013.

[48] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

A. Appendix

A.1 *Datasets Description (* indicates virtual contacts)*

- **Manifacturing Email (ME) ***: Emails exchanged between 167 employees of a mid-size company in Poland, observation time: 270 days, time resolution 1 s
- **European Union Mail(EEU) ***: Emails exchanged between 986 accounts of a large European research institution during a period from October 2003 to May 2005 (18 months), time resolution 1 s
- **Democratic National Committee Mail ***: Emails of 1900 members (1598 after preprocessing) of the Democratic National Committee, in our case only final 33 days were considered, because they are more than 95% of the entire corpus of email, time resolution 1 s
- **College Messages (CM)** *: messages from an online community of 1899 (1892 after preprocessing) students at the University of California, Irvine. Time span of approximately 6 months, time resolution 1 s
- **Hypertext 2009 (HT09)** face-to-face interactions (Rfid sensors, range of 1.5-2 m, time resolution of 20s) of the 113 participants to Hypertext conference, during 3 days.
- **Infectious (Science Gallery, Dublin)** face-to-face interactions (Rfid sensors, range of 1.5-2 m, time resolution of 20s) of 14000 visitors (410 after preprocessing) at the Science Gallery of Dublin, during 3 months of observation (after preprocessing, i.e. selecting the largest connected component, 1 day). Community structure linked to time of visit (only visitors present at the same time can interact)
- **Workplace (WP)** face-to-face interactions (Rfid sensors, range of 1.5-2 m, time resolution of 20s) of 92 employees in one of the two office buildings of the InVS, located in Saint Maurice near Paris, France, during two weeks. Each participant belongs to a department (5 in total), so the network has community structure.

- **Workplace (WP2)** Second deployment of WP, same details as WP, but larger number of participants (217) and more departments included (12). Each participant belongs to a department, so the network has community structure.

- **SFHH Conference (SFHH)** face-to-face interactions (Rfid sensors, range of 1.5-2 m, time resolution of 20s) of 403 participants to the 2009 SFHH conference in Nice, France (June 4-5, 2009).

- **Primary School (PS)** face-to-face interactions (Rfid sensors, range of 1.5-2 m, time resolution of 20s) of 242 individuals (232 children and 10 teachers) in a primary school in Lyon, France during two days in October 2009. Each kid or teacher belongs to a class, so the network has community structure.

- **High school 2012 (HS2012)** face-to-face interactions (Rfid sensors, range of 1.5-2 m, time resolution of 20s) of 180 students of five classes of a high school in Marseilles, France, during 7 days (from a Monday to the Tuesday of the following week) in Nov. 2012. Each student belongs to a class, so the network has community structure.

- **High school 2013 (HS2013)** face-to-face interactions (Rfid sensors, range of 1.5-2 m, time resolution of 20s) of 327 students of nine classes of a high school in Marseilles, France, during 5 days in Dec. 2013. Each student belongs to a class, so the network has community structure.

- **Hospital** face-to-face interactions (Rfid sensors, range of 1.5-2 m, time resolution of 20s) between patients, patients and health-care workers (HCWs) and among HCWs in a hospital ward in Lyon, France, from Monday, December 6, 2010 at 1:00 pm to Friday, December 10, 2010 at 2:00 pm. The study included 46 HCWs and 29 patients.

A.2 *Global probability distribution of inter-event times*

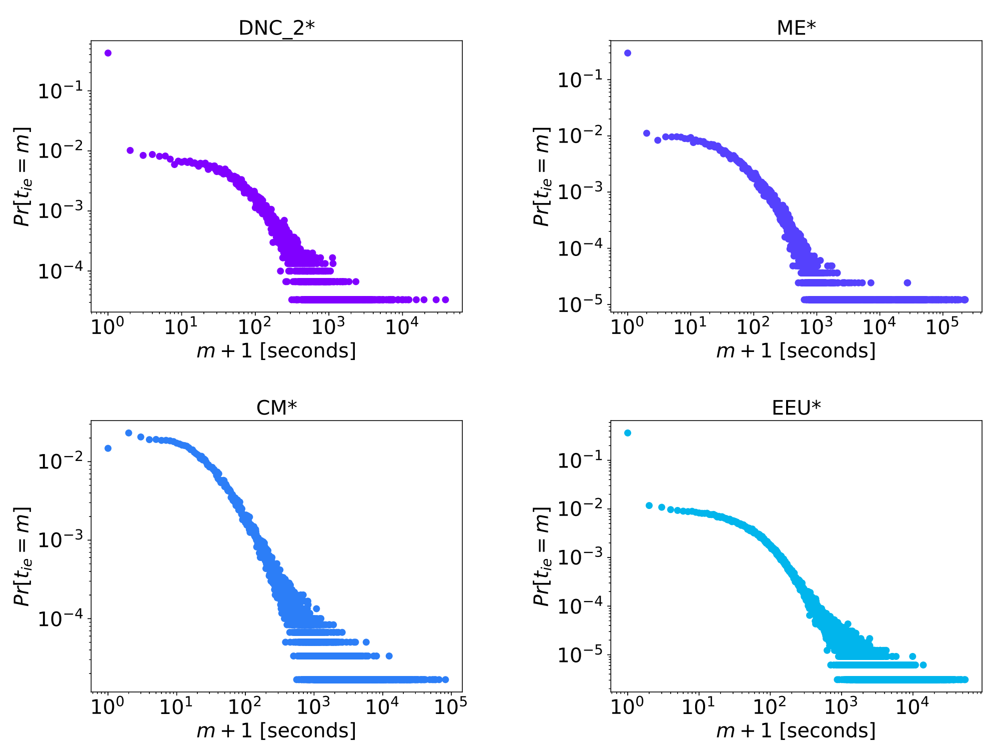

FIG. A.11: Probability distribution $Pr[t_{ia} = m]$ of the inter-event time of the global activity of virtual contact temporal networks. inter-event times are reported in seconds.

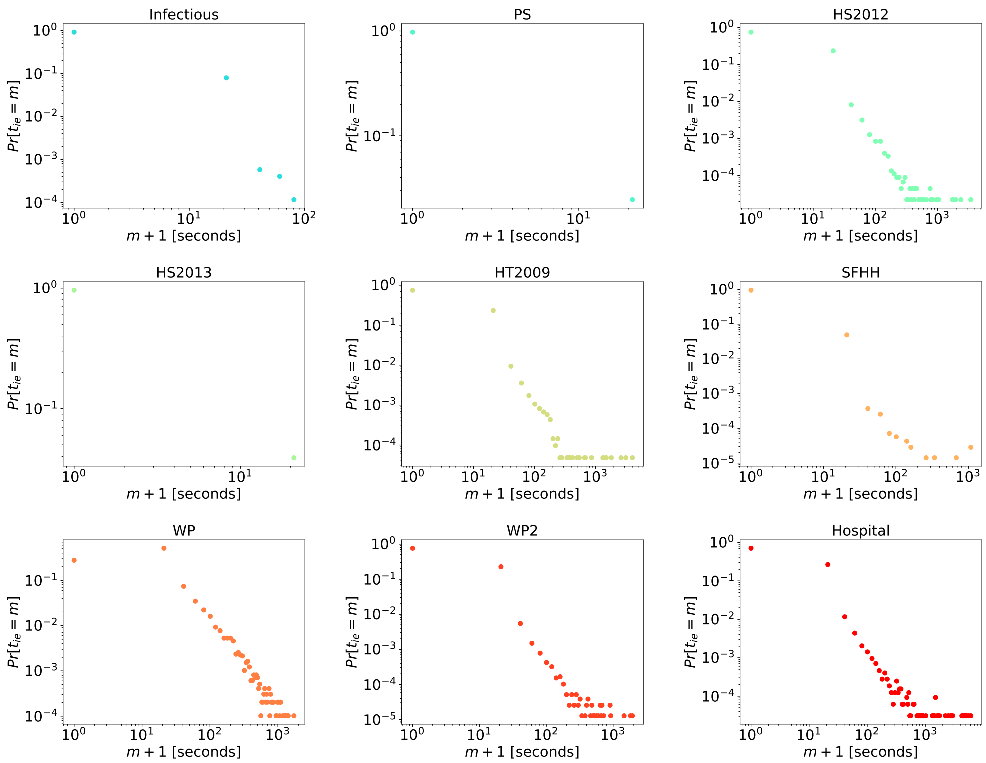

FIG. A.12: Probability distribution $Pr[t_{ia} = m]$ of the inter-event time of the global activity of physical contact temporal networks. inter-event times are reported in seconds.

A.3 *Global probability distribution of topological and temporal distances*

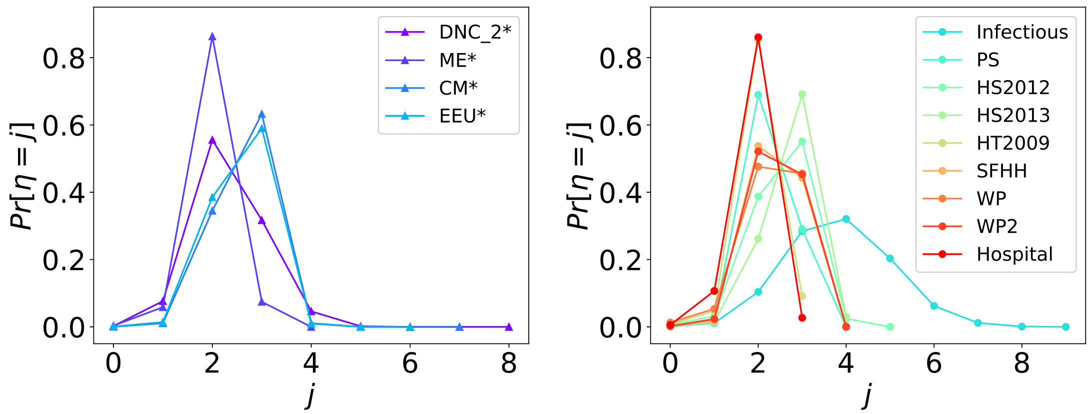

FIG. A.13: Probability distribution $\Pr[\eta = j]$ of the spatial/topological distance, j, between two randomly selected contacts in virtual (a) and physical (b) contact temporal network.

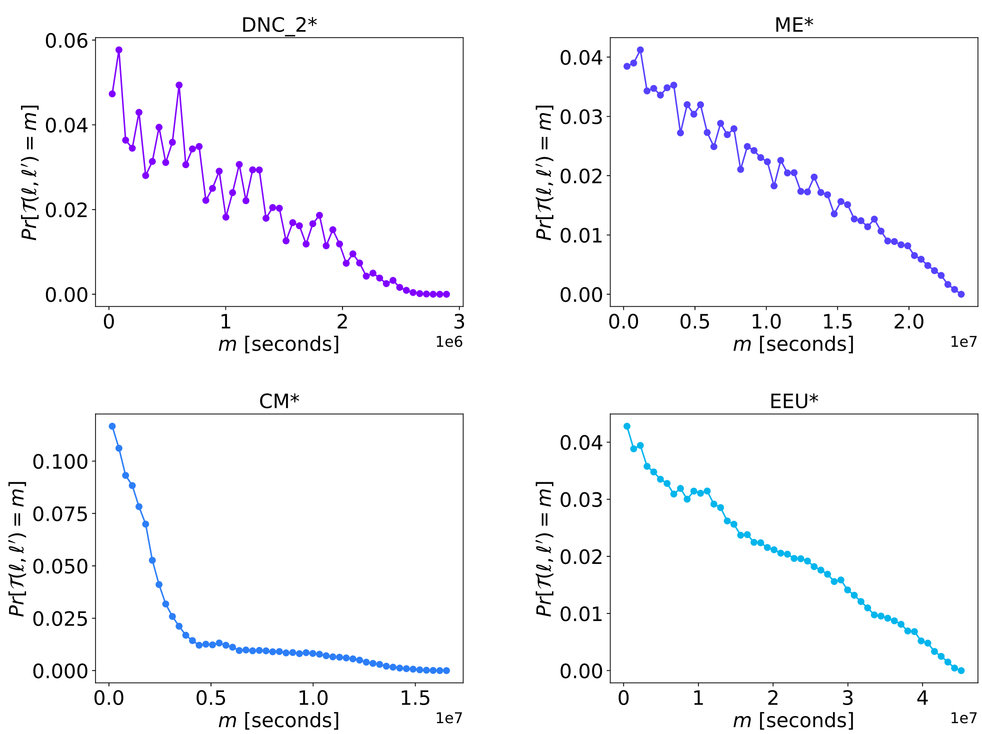

FIG. A.14: Probability distribution $Pr[\mathscr{T}(\ell, \ell') = m]$ of the temporal distance between two randomly selected contacts in each of virtual contact temporal networks. Temporal distance is reported in seconds.

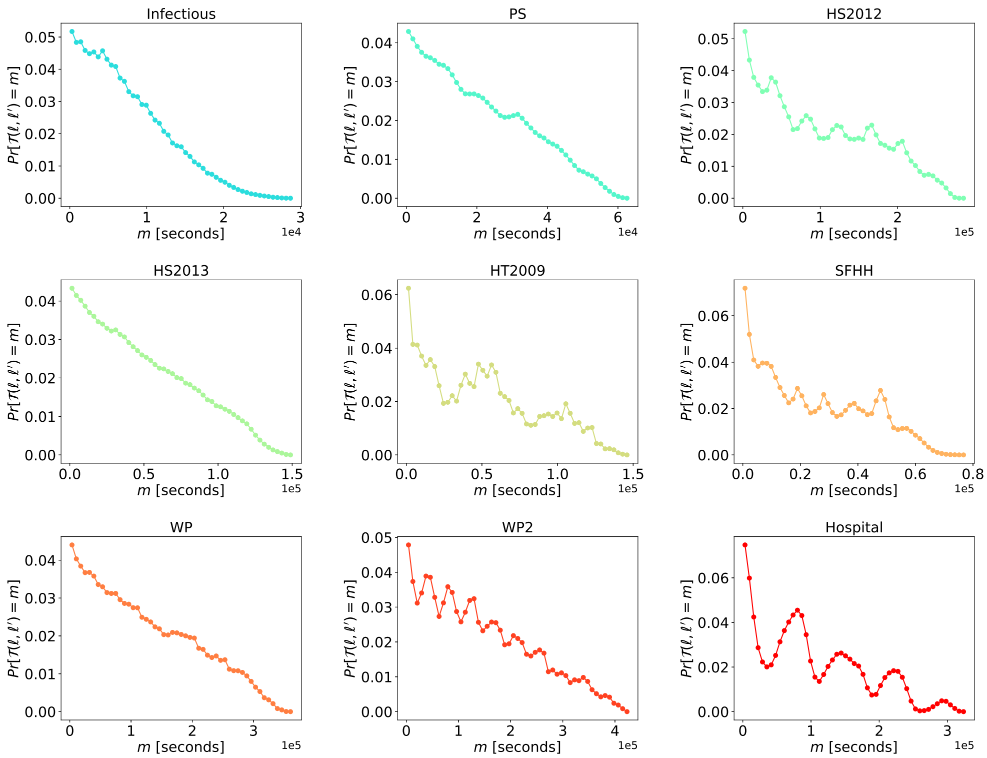

FIG. A.15: Probability distribution $Pr[\mathscr{T}(\ell, \ell') = m]$ of the temporal distance between two randomly selected contacts in each of physical contact temporal networks. Temporal distance is reported in seconds.

A.4 *Inter event time distribution of links*

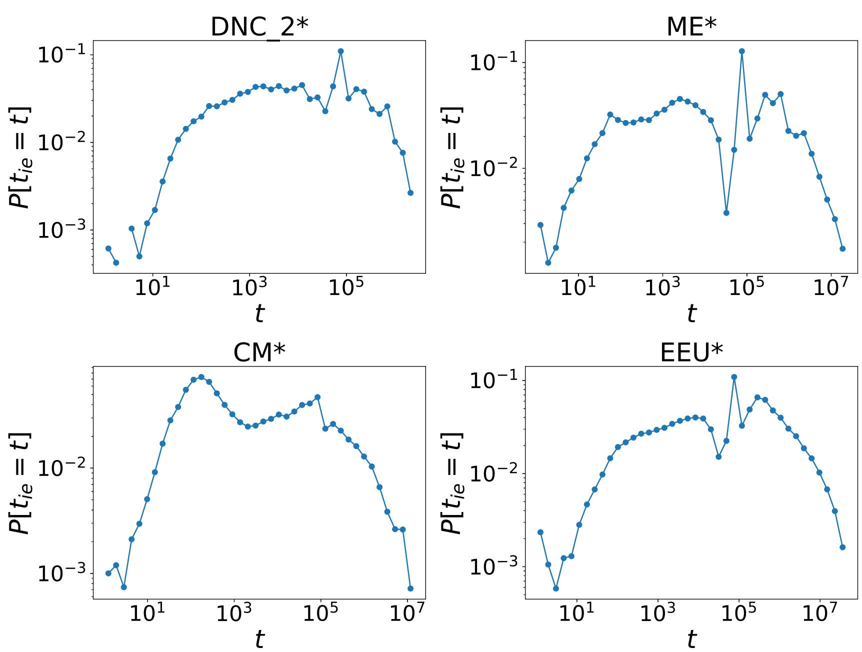

FIG. A.16: Inter-event time distribution of single link activity of virtual contact datasets. Note that the horizontal and vertical axes are presented in logarithmic scales. Inter-event times are measured in seconds. In total 40 logarithmic bins are split within the interval $[t_{min}, t_{max}]$ where t_{min} and t_{max} are, respectively, the minumum and maximum inter-event time observed in the considered dataset.

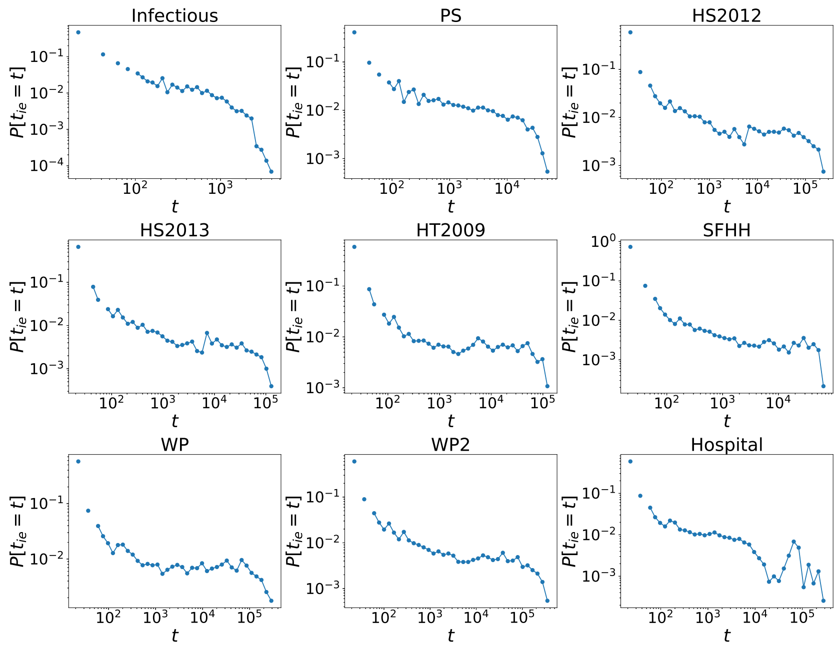

FIG. A.17: Inter-event time distribution of single link activity of physical contact datasets. Note that the horizontal and vertical axes are presented in logarithmic scales. Inter-event times are measured in seconds. In total 40 logarithmic bins are split within the interval $[t_{min}, t_{max}]$ where t_{min} and t_{max} are, respectively, the minumum and maximum inter-event time observed in the considered dataset.

A.5 *Temporal correlation of local events, additional figures*

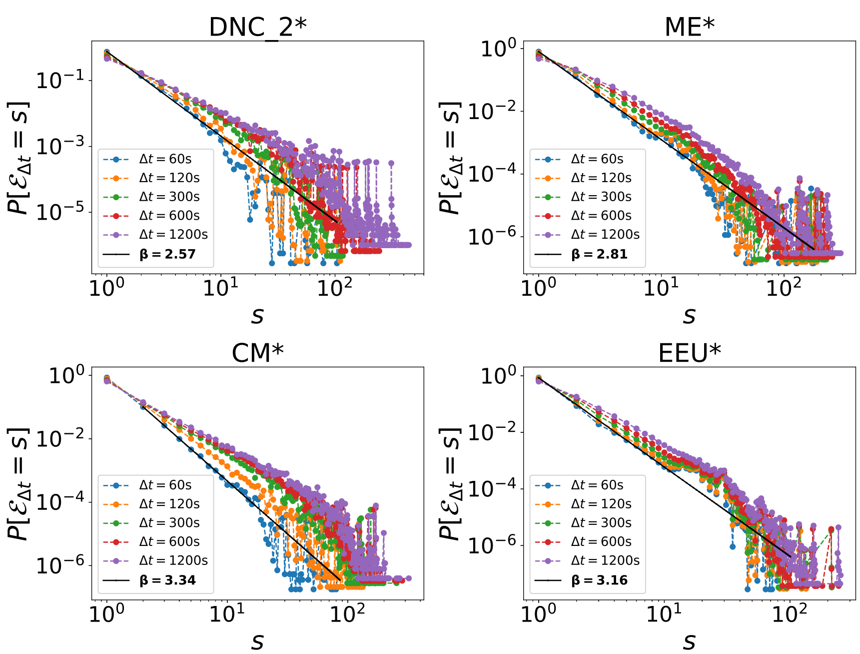

FIG. A.18: Train size distributions of ego network activity of $\mathscr{G}$ of virtual contact datasets with $\Delta t =$ 60 (blue), 120 (red), 300 (green), 600 (yellow), 1200 (purple) seconds. The black solid line represents the fit to the distribution $P[\mathscr{E}_{\Delta t} = s] \sim s^{-\beta}$, where β is the exponent of the power law fit of the train size distribution of $\mathscr{G}$ with $\Delta t = 60s$. The power law fit and the scaling region of the power law were computed with Clauset's method [48]. If the goodness of fit of the power law is significantly better than the exponential fit (likelihood ratio test with p-value $p < 0.05$), the value of β is reported in bold characters. Note that the horizontal and vertical axes are presented in logarithmic scales.

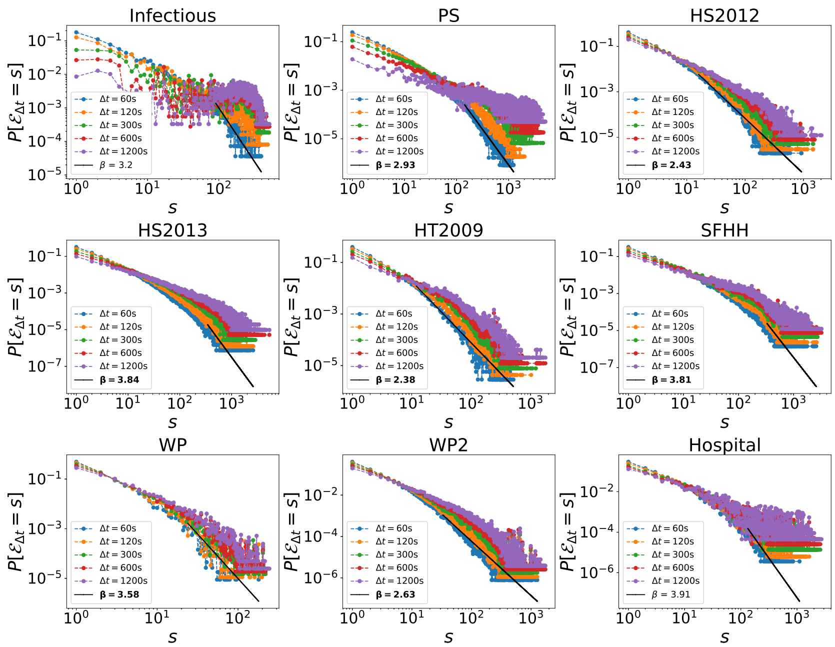

FIG. A.19: Train size distributions of ego network activity of $\mathscr{G}$ of physical contact datasets with $\Delta t =$ 60 (blue), 120 (red), 300 (green), 600 (yellow), 1200 (purple) seconds. The black solid line represents the fit to the distribution $P[\mathscr{E}_{\Delta t} = s] \sim s^{-\beta}$, where β is the exponent of the power law fit of the train size distribution of $\mathscr{G}$ with $\Delta t = 60s$. The power law fit and the scaling region were computed with Clauset's method [48]. If the goodness of fit of the power law is significantly better than the exponential fit (likelihood ratio test with p-value $p < 0.05$), the value of β is reported in bold characters. Note that the horizontal and vertical axes are presented in logarithmic scales.

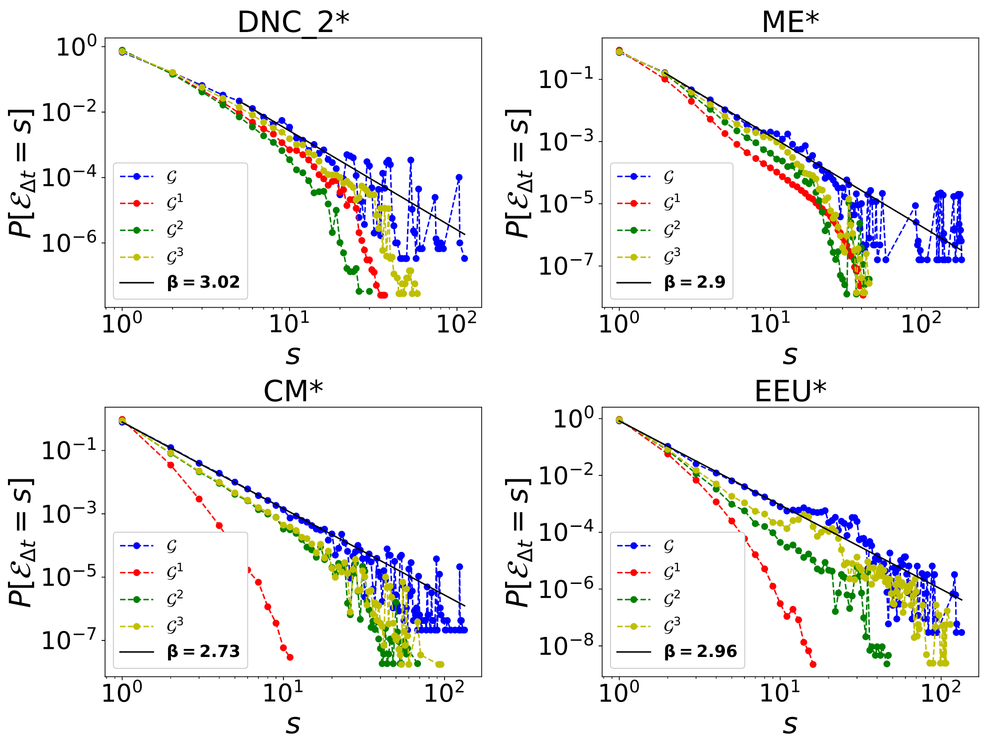

FIG. A.20: Train size distribution ($\Delta t = 120s$) of ego network activity for $\mathscr{G}$ (blue), $\mathscr{G}_1$ (red), $\mathscr{G}_2$ (green), $\mathscr{G}_3$ (yellow) of virtual contact datasets. The black solid line represents the fit to the distribution $P[\mathscr{E}_{\Delta t} = s] \sim s^{-\beta}$, where β is the exponent of the power law fit of the train size distribution of $\mathscr{G}$ with $\Delta t = 60s$. The power law fit and the scaling region were computed with Clauset's method [48]. If the goodness of fit of the power law is significantly better than the exponential fit (likelihood ratio test with p-value $p < 0.05$), the value of β is reported in bold characters. Note that the horizontal and vertical axes are presented in logarithmic scales.

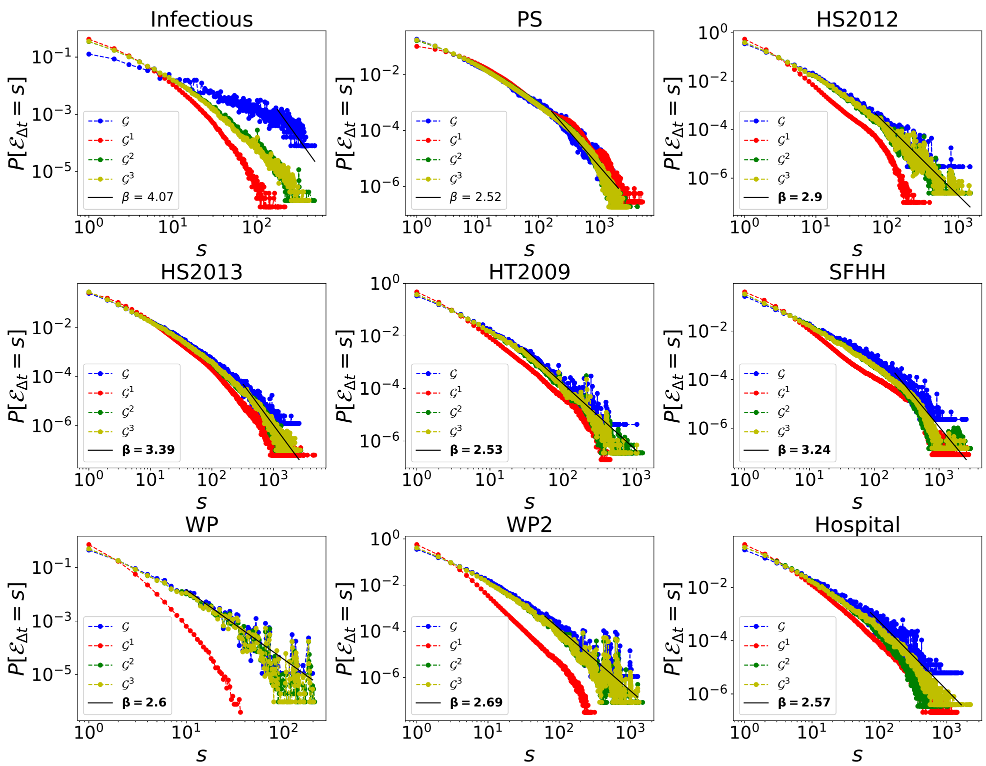

FIG. A.21: Train size distribution ($\Delta t = 120s$) of ego network activity for $\mathscr{G}$ (blue), $\mathscr{G}_1$ (red), $\mathscr{G}_2$ (green), $\mathscr{G}_3$ (yellow) of physical contact datasets. The black solid line represents the fit to the distribution $P[\mathscr{E}_{\Delta t} = s] \sim s^{-\beta}$, where β is the exponent of the power law fit of the train size distribution of $\mathscr{G}$. The power law fit and the scaling region were computed with Clauset's method [48]. If the goodness of fit of the power law is significantly better than the exponential fit (likelihood ratio test with p-value $p < 0.05$), the value of β is reported in bold characters. Note that the horizontal and vertical axes are presented in logarithmic scales.

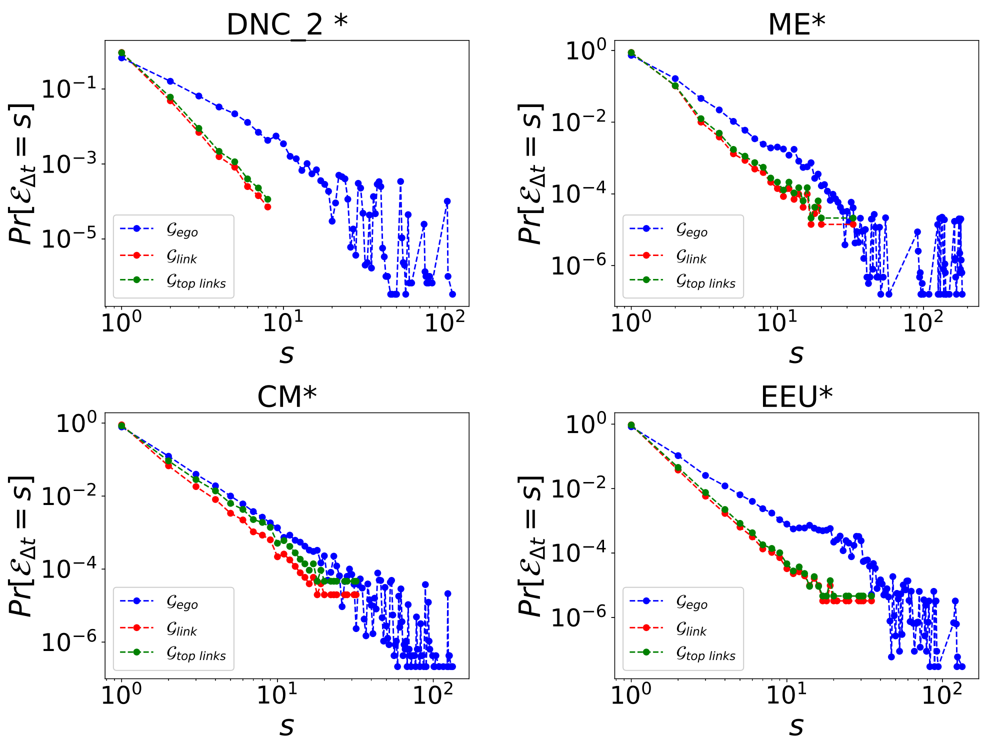

FIG. A.22: Train size distribution ($\Delta t = 120s$) of ego network activity (blue), single link activity (red), most active link activity (green) of virtual contact datasets. Note that the horizontal and vertical axes are presented in logarithmic scales.

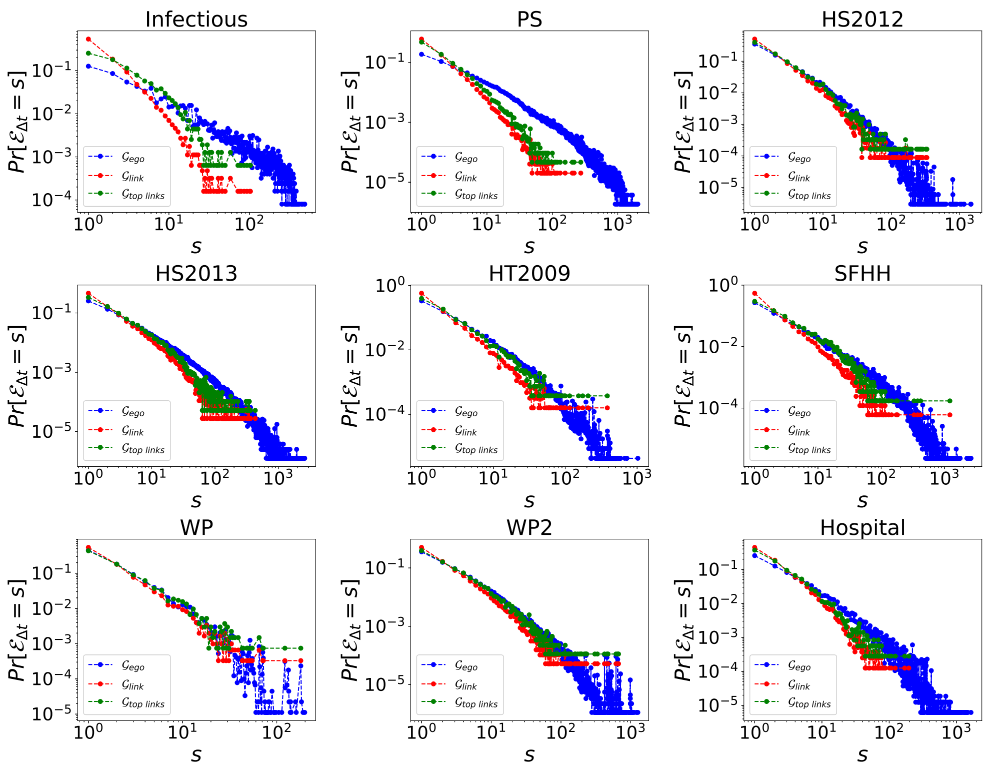

FIG. A.23: Train size distribution ($\Delta t = 120s$) of ego network activity (blue), single link activity (red), most active link activity (green) of physical contact datasets. Note that the horizontal and vertical axes are presented in logarithmic scales.

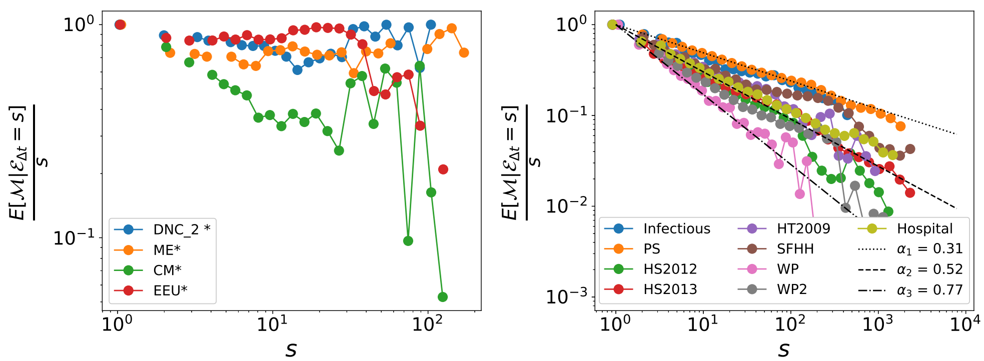

FIG. A.24: The average number of active links $\frac{E[\mathcal{M}|\mathcal{E}_{\Delta t}=s]}{s}$ for trains with size $\mathcal{E}_{\Delta t} = s$ ($\Delta t = 120s$), normalized by the train size s of the ego networks for virtual (left) and physical (right) contact datasets. The three reference lines in right plot indicate $\frac{E[\mathcal{M}|\mathcal{E}_{\Delta t}=s]}{s} = s^{-\alpha}$ with slope $\alpha_1 = 0.31$ (dotted), $\alpha_2 = 0.52$ (dashed) and $\alpha_3 = 0.77$ (dash-dot). Note that the horizontal and vertical axes are presented in logarithmic scales. In total 30 logarithmic bins are split within the interval $[1, s_{max}]$, where s_{max} is the largest train size observed in the considered real temporal network.