Causal Paths in Temporal Networks of Face-to-Face Human Interactions

Agostino Funel
ENEA - Energy Technologies Department, ICT HPC Lab
Via E. Fermi, 1
80055 Portici (Naples), Italy

In a temporal network, causal paths are characterized by the fact that links from a source to a target must respect the chronological order. In this paper we study the causal paths structure in temporal networks of face-to-face human interactions in different social contexts. In a static network, paths are transitive; that is, the existence of a link from a to b and from b to c implies the existence of a path from a to c via b. In a temporal network, the chronological constraint introduces time correlations that affect transitivity. A probabilistic model based on higher-order Markov chains shows that correlations that can invalidate transitivity are present only when the time gap between consecutive events is larger than the average value and are negligible below such a value. The comparison between the densities of the temporal and static accessibility matrices shows that the static representation can be used with good approximation. Moreover, we quantify the extent of the causally connected region of the networks over time.

Keywords: temporal networks; human interactions; causal paths; Markov chains; probabilistic models

1. Introduction

Many real systems composed of elements connected to each other can be described as networks. Notable examples are the internet, transport networks and power grids. These technological and infrastructural networks can be considered static over timescales of many human daily activities because their structure does not change. On the contrary, human interactions can be thought of as events occurring over time. Product ratings, emails and social networks are examples of temporal networks [1] because the topology changes over time. The set of interactions between any two nodes of the network can be described as a set of links to each of which are associated the nodes that interacted and the time when the interaction occurred. Interactions involving many nodes can be treated by considering all links that traversed them during the temporal interval in which the interactions occurred.

https://doi.org/10.25088/ComplexSystems.30.1.33

The study of social dynamics is becoming increasingly important thanks to the availability of large datasets collected with technologies capable of monitoring human activities at different levels of space-temporal scales. Consider, for example, mobile calls, instant messaging and digital traces [2–4]. One might think that data analysis sheds light on human behavior [5, 6]. Many aspects of human interactions can be analyzed considering the static time-aggregated network in which a link between any two nodes exists if they interacted at least once. Additional statistical information can be obtained by associating to each link a weight that is the number of times the link has been active during the observation period. However, the (weighted) static representation may not be adequate to describe the underlying dynamics stemming from human interactions, which can often be described as a diffusive process over the social network or by means of models where individuals are agents who change their state over time according to specific rules [7–10]. Moreover, in these cases events obey the principle of causality. When studying human interactions, it is crucial to specify whether they are direct or mediated. With the advent of wearable devices, it is now possible to study human interactions at the level of face-to-face proximity. In this context, we can represent personal encounters as a collection of pathways that traverse all the people who interacted.

In this paper, we study the causal paths structure in temporal networks of face-to-face human interactions in different social contexts. When causality is involved, we must take into account the temporal order in which the links of a path occur. The ordering of the links introduces correlations that may invalidate results based on a static time-aggregated network [11–13]. For temporal networks, there are fewer well-established analysis techniques than for static representations. We investigate whether and to what extent the static representation is justified for the examined networks, using two different probabilistic models based on higher-order Markov chains and accessibility matrices. We also quantify the extent of the causally connected region of the networks during the period of observation.

The paper is organized as follows: in Section 2 we introduce the notation and the probabilistic models; we describe the datasets in Section 3 and report the results in Section 4. We make the final conclusions in Section 5.

2. Probabilistic Models

In this section, we introduce the notation and the probabilistic models used to analyze the datasets. A temporal network $G_T = (V, E_T)$ can be defined as a set of nodes (or vertices) $v_i \in V$ and edges (or links)

$E_T = V \times V \times [0, T]$ where $[0, T]$ is the temporal interval of observation. The interaction between two nodes v_1 and v_2 that starts at time t and has duration ω is an event that can be represented as a time-stamped link (v_1, v_2, t, ω). In the case of the examined datasets, we can consider the time as a discrete variable. Events whose duration is less than the temporal resolution ϵ of the data acquisition system cannot be observed. In all other cases, ω can be divided into n elementary time steps $\omega = n\sigma$ $(\sigma \geq \epsilon)$ and the interaction (v_1, v_2, t, ω) can be replaced by the sequence of "instantaneous" events (v_1, v_2, t_i) where $t_i = t + i\sigma \leq t + \omega$ and $i = 0, \ldots, n$. A causal path between a source v_0 and a target v_l is any sequence

$$(v_0, v_1, t_1), (v_1, v_2, t_2), \ldots, (v_{l-2}, v_{l-1}, t_{l-1}), (v_{l-1}, v_l, t_l)$$

where $t_1 < t_2 < \cdots < t_{l-1} < t_l$. The length l of the path is the number of links that traverse its nodes. A single node is a path of length zero. It is sometimes useful to impose the constraint $0 < t_{i+1} - t_i \leq \delta$ between consecutive events to select only causal paths that contribute to dynamical processes whose characteristic timescale is δ. For example, we may be interested in studying the propagation of information during conversations of average duration δ by analyzing data collected during a period $T \gg \delta$.

2.1 Higher-Order Markov Chains

We now briefly present the higher-order Markov chains probabilistic model used to analyze the datasets. The model has been proposed in [14] and the reader is encouraged to read the paper for more details. In the static time-aggregated network representation, the existence of the links (v_{i-1}, v_i) and (v_i, v_{i+1}) implies the existence of the path (v_{i-1}, v_i, v_{i+1}); then we would be led to think that v_{i-1} can influence v_{i+1} via v_i. However, this transitivity may be invalidated in the temporal network by time correlations because it exists only if (v_{i-1}, v_i) occurred before (v_i, v_{i+1}). For any given δ, let $S = \{p_{(1)}, \ldots, p_{(N)}\}$ be the collection of all causal paths that satisfy the condition $0 < t_{i+1} - t_i \leq \delta$. A causal path p_l of length l can be considered a sequence of transitions over random variables $(v_0 \to v_1 \to \cdots \to v_l)$. In this model, we define a discrete-time Markov chain of order k and assume that for a vertex v_i of the path, the probability to reach it depends only on the k previously traversed vertices

$$P(v_i \mid v_0 \to \cdots \to v_{i-1}) = P(v_i \mid v_{i-k} \to \cdots \to v_{i-1}).$$

For a given order k, a maximum likelihood estimation is performed to find the transition probabilities

$$P^{(k)} := P(v_i \mid v_{i-k} \to \cdots \to v_{i-1}).$$

https://doi.org/10.25088/ComplexSystems.30.1.33

It is found that the probabilities that maximize the likelihood function can be calculated from the relative frequencies of subpaths of length k in S. The order $k = 0$ only generates vertices of the network. For $k = 1$, only subpaths of length 1 are considered, and the model reproduces the topology of the static weighted network where the weight of each edge is its frequency. For $k > 1$, the model captures correlations of longer paths that cannot be evaluated on the basis of the topology of the underlying static network. To a k-order model is associated a k-order aggregated network $G^{(k)}$ [15]. Each node of $G^{(k)}$ is a k-tuple $v^{(k)} = (v_1, \dots, v_k)$, and a link $\left(v^{(k)}, w^{(k)}\right)$ exists between two nodes if $v_{i+1} = w_i$ for $i = 1, \dots, k-1$. A k-order Markov chain is treated as a first-order Markov chain over nodes of $G^{(k)}$. To find which is the higher-order network abstraction that best models the observed paths, a procedure that combines multiple layers of Markov chains is used. For a given path p_l, the transition probability of the multilayer model of order K is

$$\mathcal{P}^{(K)}(v_0 \to \cdots \to v_l) = \prod_{k=1}^{K-1} P^{(k)}(v_k \mid v_0 \to \cdots \to v_{k-1}) \prod_{i=K}^{l} P^{(K)}(v_i \mid v_{i-K} \to \cdots \to v_{i-1}).$$

The likelihood function of the multilayer order K model is

$$\mathcal{L}_K = \prod_{j=1}^{N} \mathcal{P}^{(K)}\left(p_{(j)}\right),$$

and the optimal order K_{opt} is found by progressively evaluating the likelihood ratio $\mathcal{L}_K / \mathcal{L}_{K+1}$ of models having consecutive orders. If $K_{\text{opt}} = 1$, correlations preserve the transitivity, while it is broken if $K_{\text{opt}} > 1$. Thus in the first case the static time-aggregated network representation is justified to analyze data. We note that the set of observed causal paths S depends on the inter-events time gap δ, therefore K_{opt} also depends on it.

2.2 Accessibility Matrices

This model has been proposed in [16] and the reader is advised to read the paper for the details. In a static network, the accessibility matrix provides a macroscopic view of the connections. Its elements are 1 if the nodes corresponding to the indices are connected by a path of any length, and zero otherwise. In the static case, the adjacency matrix $\mathbf{A}$ of the network is $\mathbf{A}_{ij} = 1$ if nodes (i, j) are connected by an edge; $\mathbf{A}_{ij} = 0$ otherwise. The elements of the matrix $\mathbf{A}^n$ are the number of paths of length n connecting the nodes. We define the

binary operator $\mathcal{U}$, which acts on matrices in the following way: $\mathcal{U}(\mathbf{M}_{ij}) = 1$ if $\mathbf{M}_{ij} \neq 0$ and $\mathcal{U}(\mathbf{M}_{ij}) = 0$ if $\mathbf{M}_{ij} = 0$. The accessibility matrix is

$$\mathbf{P}_N = \mathcal{U}\left(\sum_{i=1}^{N} \mathbf{A}^i\right),$$

where N is the number of nodes. Since the maximum distance is the diameter D of the network, the process of summing up saturates and one has $\mathbf{P}_N = \mathbf{P}_D$. The density of an $N \times N$ matrix $\mathbf{M}$ is

$$\rho(\mathbf{M}) = \frac{\text{nnz}(\mathbf{M})}{N^2},$$

where nnz($\mathbf{M}$) is the number of nonzero elements. The probability of finding a shortest path of length $l \leq n$ between two randomly chosen nodes is

$$P(l \leq n) = \rho(\mathbf{P}_n).$$

For a disconnected network, $\rho(\mathbf{P}_D) < 1$ because the number of nonzero entries of the accessibility matrices of the disjoint components is less than N^2. A temporal network can be represented by a sequence of chronologically ordered adjacency matrices

$$\mathcal{A} = \{\mathbf{A}_{t_1}, \mathbf{A}_{t_2}, \ldots, \mathbf{A}_{t_T}\}.$$

Causality is guaranteed by the condition $t_1 < t_2 < \cdots < t_T = T$. The accessibility matrix of the temporal network is

$$\mathcal{P}_T = \mathcal{U}\left(\prod_{i=1}^{T} (\mathbf{1} + \mathbf{A}_{t_i})\right).$$

The density of the accessibility matrix $\rho(\mathcal{P}_n)$ gives the fraction of the network causally connected for $t \leq t_n$. The model provides a criterion to quantify the goodness of the static time-aggregated approximation for a temporal network by evaluating the causal fidelity index $0 \leq \gamma = \rho(\mathcal{P}_T) / \rho(\mathbf{P}_N) \leq 1$. Low values of γ indicate bad approximation, meaning that the majority of paths in the static representation do not correspond to causal paths in the temporal network. Instead, high values of γ indicate good approximation.

3. Datasets

We use publicly available datasets provided by the SocioPatterns [17] collaboration. The data acquisition system is based on a sensing platform that uses wearable badges equipped with radio frequency

identification (RFID) devices. Contacts data is gathered at the level of face-to-face proximity (~1.5 meter) with a temporal resolution of 20 seconds [18]. We consider different social contexts:

- High school: contacts between students of five classes in a high school in Marseilles (France) collected during seven days in November 2012 [19].
- Hospital: contacts between patients and healthcare personnel in a hospital ward in Lyon (France) collected during five days in December 2010 [20].
- HT09: encounters between people who attended the ACM Hypertext Conference held in Turin (Italy) in 2009 [21].
- Infectious: encounters between people during the Infectious exhibition event held at the Science Gallery in Dublin (Ireland) from April 17 to July 17, 2009 [21].
- SFHH: interactions between participants at a scientific conference in Nice (France), June 4–5, 2009 [22].
- Workplace: contacts between individuals in an office building in France in 2015 [22].

Since all measurements of interest to us are invariant under time translation, to better compare the temporal networks for all of them, we make sure that events start at $t = 0$. Table 1 shows the size and temporal information of the datasets.

The software libraries we use in this paper are: tacoma at github.com/benmaier/tacoma to study the edge activity, pathpy at github.com/IngoScholtes/pathpy to study the causal paths statistics and perform the multilayer Markov chains analysis, and TemporalNetworkAccessibility classes at github.com/hartmutlentz/TemporalNetworkAccessibility to compute the accessibility matrices.

Network	Nodes	Time-Stamped Links	Observation Duration (sec)	δ_{avg}(sec)	δ_{max}(sec)
High school	180	45047	729500	64.72	220280
Hospital	75	32424	347500	36.76	26980
HT09	113	20818	212340	40.48	28900
Infectious	10972	415912	6946340	90.28	152980
SFHH	403	70261	114300	32.58	38320
Workplace	217	78249	993540	53.74	218600

Table 1. Human face to face interaction temporal networks analyzed in this paper. The minimum inter-events time δ_{min} is equal to 20 sec, which is the time resolution of the data acquisition system. The average (maximum) time gap between consecutive events is δ_{avg} (δ_{max}).

4. Results

The overall edge activity is plotted in Figure 1 for all networks. Inactivity periods during which data was not collected are clearly shown. The total number of causal paths depends on the inter-events time window δ. Looking at Table 1, we observe that for all networks, the minimum and average time gap between consecutive events $\delta_{\min}$ and δ_{avg} is of the order of tens of seconds, while the maximum $\delta_{\max}$ is two or three orders of magnitude larger. To sample the interval $[\delta_{\min}, \delta_{\max}]$, we divide it into subintervals of equal width. We choose the sampling width $\mu = 1$ minute and collect the path statistics at

$$\delta_i = \delta_{\min} + i\mu \leq \delta_{\max} \quad i = 0, 1, \ldots, m.$$

Figure 2 shows the probability distributions $P(l, \delta)$ of causal paths of length l as a function of δ. It is interesting to note that the distributions $P(l, \delta)$ have the same trend for all social contexts. The probability $P(1, \delta)$ decreases with δ, while $P(l > 1, \delta)$ increases in almost the whole range $[\delta_{\min}, \delta_{\max}]$. In the figure, only the probability distributions for paths of length $l \leq 5$ are plotted for better readability. The distributions $P(l > 5, \delta)$ all have the same shape of $P(l > 1, \delta)$ and $P(l + 1, \delta) < P(l, \delta)$. In the interval $[\delta_{\min}, \delta_{\text{avg}}]$, the percentage of paths of length $l = 1$ varies in the range ~[70–90]%; the percentages of paths of length $l = 2$ in the range ~[7–23]%, and paths of length $l \geq 3$ are rare. In the limit $\delta \to \delta_{\max}$, the contribution of paths of length $3 \leq l \leq 10$ to the statistics is significantly higher. However, even in this limit, paths of length $l > 10$ are rare. Figure 3 shows for $\delta = \delta_{\min}$, δ_{avg} and $\delta_{\max}$ the probabilities of causal paths of length $l \leq 10$. Time correlations in causal paths come into play only when $l \geq 2$, and their importance is greater the longer the path length. Therefore, we expect that transitivity is not lost around δ_{avg} but only for $\delta > \delta_{\text{avg}}$. This is confirmed by the higher-order, multilayer Markov chains model. It is difficult to predict the exact value of δ for which a first-order network abstraction is inadequate. However, we argue that since for the majority of the links of a causal path the time gap between consecutive events is around δ_{avg}, the first-order approximation could be acceptable and the static time-aggregated network model could be used with a good degree of confidence. We thus compute the causal fidelity index γ to quantify the goodness of the static representation. The results of both models are reported in Table 2. Apart from the Infectious network for which $\gamma = 0.57$, we observe a value of $\gamma \approx 1$ in all other cases. This indicates that regardless of the social context, the static time-aggregated network representation is a good approximation of the causal paths structure and may be trusted

to describe dynamical processes whose timescale does not differ too much from the average inter-events time gap. The density of the accessibility matrix $\rho(\mathcal{P}_n)$ is plotted in Figure 4 for all networks. It gives information on the extent of the causally connected region of the network over time.

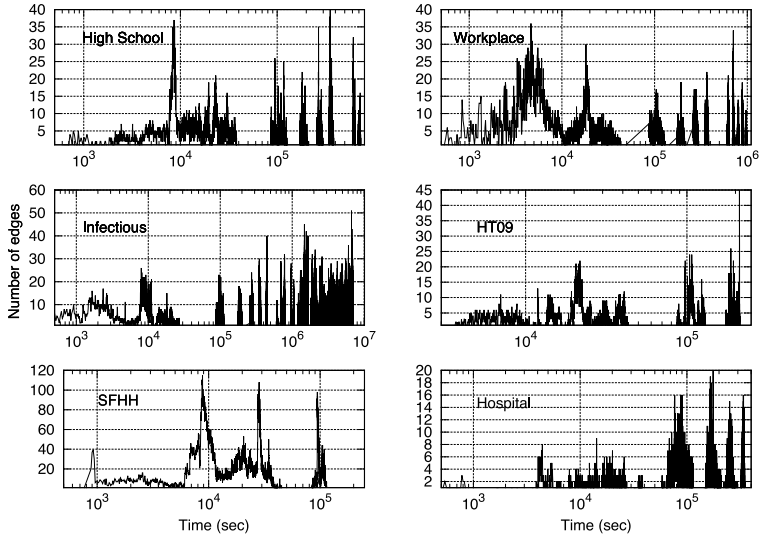

Figure 1. Number of active edges during the observation period.

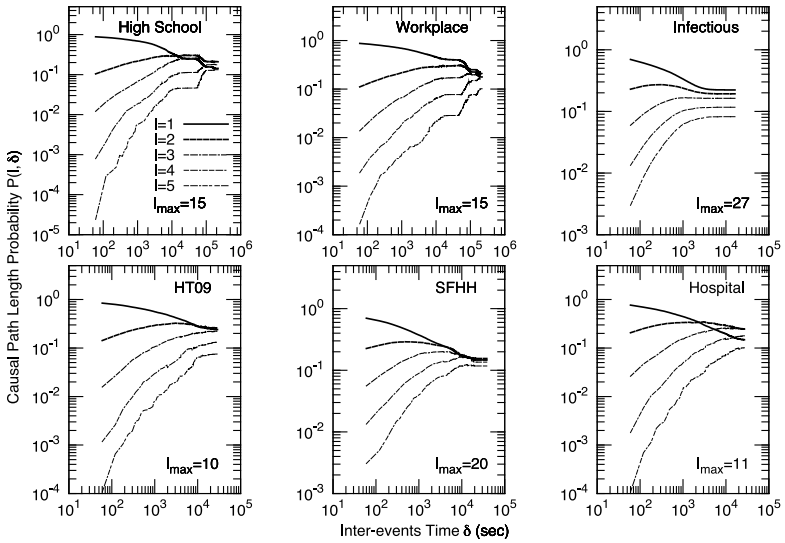

Figure 2. Probability distributions $P(l, \delta)$ of causal paths of length l as a function of the inter-events time δ. The maximum observed path length is $l_{\max}$.

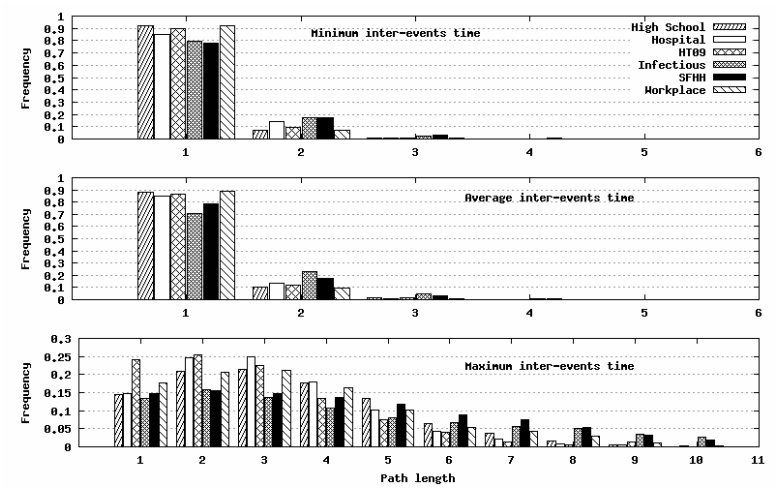

Figure 3. Probability distributions of causal paths as a function of the length for the inter-events times $\delta = \delta_{\min}$ (top), $\delta = \delta_{\text{avg}}$ (middle) and $\delta = \delta_{\max}$ (bottom).

	K_{opt}			
Network	$\delta_{\min}$	δ_{avg}	$\delta_{\max}$	γ
High school	1	1	2	0.97
Hospital	1	1	2	0.93
HT09	1	1	2	0.99
SFHH	1	1	2	0.99
Workplace	1	1	2	0.99
Infectious	1	1	2	0.57

Table 2. The table shows, for all examined temporal networks, the optimal maximum order K_{opt} provided by the multilayer Markov chains model evaluated for minimum, average and maximum inter-events times, and the causal fidelity index γ.

Table 3 shows the extent of the causally connected region of the networks during the observation period T. The Infectious network is ~1% causally connected at the end of the observation period. Its static time-aggregated network is not connected. This means that there are components that never were connected during the observation period. These components contain isolated individuals or disjoint groups of individuals. Note that the analyzed data of the Infectious network was collected from April 17 to July 17, 2009; therefore it is possible that people who participated in the exhibition event on different days never met. We observe that the High school and SFHH networks are 90% causally connected after about one-quarter of the

observation period, meaning a high level of mobility and interaction between individuals. We may think that students of a class had frequent relationships not only between themselves but also with students of other classes, for example during recess time after courses, and people who attended the scientific conference frequently met to exchange opinions. The Hospital network becomes 90% causally connected only at the end of the observation period and the extent of the network region connected by causal paths increases quite slowly. The HT09 network becomes 90% causally connected after ~50% of the observation period, which is a sign that the majority of people had already had a face-to-face contact during the first half of the conference duration time. The Workplace network is 25% causally connected just after ~1% of the observation period and becomes 90% causally connected after ~37%. This is an indication that the employees who worked in the office building had frequent relationships, probably because their activities required strong team collaboration.

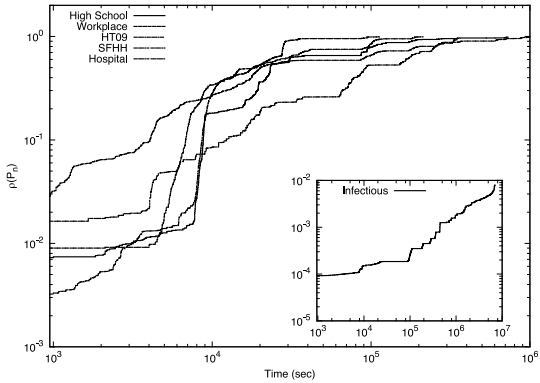

Figure 4. Densities $\rho(\mathcal{P}_n)$ of the accessibility matrices of the temporal networks as a function of the time.

	Time to Reach			
Network	25%	50%	75%	90%
High school	2.4	3.2	13.0	24.8
Hospital	10.6	26.7	70.0	95.3
HT09	4.0	8.9	20.0	48.9
SFHH	8.4	15.8	24.2	25.6
Workplace	0.88	2.0	18.0	37.3
Infectious	~1% causally connected			

Table 3. The table shows the percentage of the observation period T taken to have one-quarter, half, three-quarters and 90% of the network causally connected. The Infectious network is only ~1% causally connected at the end of T.

5. Conclusion

We studied the causal paths structure in temporal networks of human face-to-face proximity interactions in different social contexts. The importance of causal paths lies in the fact that they are the underlying background on top of which diffusive processes evolve. Temporal networks are often analyzed by applying well-established algorithms and techniques to the static time-aggregated representation. Although this approach provides information on global properties, it implicitly assumes link transitivity and does not capture time correlations. Therefore, it may lead to misleading results in the analysis of causal paths. To quantify time correlations and the goodness of the static approximation, we applied two independent models to the datasets, one based on multilayer, higher-order Markov chains, and the other on the unfolding of the accessibility matrices. We found that for all examined networks, the probability distributions of causal paths as a function of the waiting time δ between consecutive events have the same shape. The number of causal paths of length one decreases with δ and that of longer paths increases. The number of causal paths of length one considerably exceeds that of longer paths for δ less than or equal to the average value δ_{avg}. Time correlations are present for paths of length $l \geq 2$. Their number increases with δ but exceeds that of a path of length one only for $\delta \gg \delta_{\text{avg}}$. For all examined networks, the multilayer, higher-order Markov chains model shows that transitivity is not lost when $\delta = \delta_{\text{avg}}$, indicating that for temporal scales of this order, the static network representation may be a good approximation. We evaluated the ratio of the densities of the temporal and static accessibility matrices, which provides an index $0 \leq \gamma \leq 1$. High values of γ indicate that the static representation is a good approximation and low values indicate a bad approximation. Except for one of the examined networks (exhibition event lasting about three months) for which $\gamma \sim 0.6$, for all other networks we found that $\gamma \sim 1$. We also quantified the extent of the causally connected region of the networks during the observation period, obtaining an overall view of the people mobility in the different social contexts.

Acknowledgments

The computing resources and the related technical support used for this work were provided by CRESCO/ENEAGRID High Performance Computing infrastructure and its staff [23]. CRESCO/ENEAGRID High Performance Computing infrastructure is funded by ENEA, the Italian National Agency for New Technologies, Energy and

Sustainable Economic Development and by Italian and European research programs; see www.eneagrid.enea.it for information.

References

[1] P. Holme and Jari Saramäki, "Temporal Networks," *Physics Reports*, **519**(3), 2012 pp. 97–125. doi:10.1016/j.physrep.2012.03.001.

[2] D. Monsivais, A. Ghosh, K. Bhattacharya, R. I. M. Dunbar and K. Kaski, "Tracking Urban Human Activity from Mobile Phone Calling Patterns," *PLOS Computational Biology*, **13**(11), 2017 e1005824. doi:10.1371/journal.pcbi.1005824.

[3] J. Leskovec and E. Horvitz, "Planetary-Scale Views on a Large Instant-Messaging Network," in *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*, Beijing, China, 2008, New York: Association for Computing Machinery, 2008 pp. 915–924. doi:10.1145/1367497.1367620.

[4] A. Badawy, E. Ferrara and K. Lerman, "Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Barcelona, 2018 (U. Brandes, C. Reddy and A. Tagarelli, eds.), Piscataway, NJ: 2018 pp. 258–265. doi:10.1109/ASONAM.2018.8508646.

[5] N. Tremblay, A. Barrat, C. Forest, M. Nornberg, J.-F. Pinton and P. Borgnat, "Bootstrapping under Constraint for the Assessment of Group Behavior in Human Contact Networks," *Physical Review E*, **88**(5), 2013 052812. doi:10.1103/PhysRevE.88.052812.

[6] X. Lu and C. Brelsford, "Network Structure and Community Evolution on Twitter: Human Behavior Change in Response to the 2011 Japanese Earthquake and Tsunami," *Scientific Reports*, **4**(1), 2015 6773. doi:10.1038%2 Fsrep06773.

[7] Z.-K. Zhang, C. Liu, X.-X. Zhan, X. Lu, C.-X. Zhang and Y.-C. Zhang, "Dynamics of Information Diffusion and Its Applications on Complex Networks," *Physics Reports*, **651**, 2016 pp. 1–34. doi:10.1016/j.physrep.2016.07.002.

[8] X.-X. Zhan, A. Hanjalic and H. Wang, "Information Diffusion Backbones in Temporal Networks," *Scientific Reports*, **9**(1), 2019 6798. doi:10.1038/s41598-019-43029-5.

[9] A. Sirbu, V. Loreto, V. D. P. Servedio and F. Tria, "Opinion Dynamics: Models, Extensions and External Effects," *Participatory Sensing, Opinions and Collective Awareness* (V. Loreto, M. Haklay, A. Hotho, V. D. P. Servedio, G. Stumme, J. Theunis and F. Tria, eds.), Cham: Springer, 2017, p. 363–401. doi:10.1007/978-3-319-25658-0_17.

[10] G. Deffuant, D. Neau, F. Amblard and G. Weisbuch, "Mixing Beliefs among Interacting Agents," *Advances in Complex Systems*, **3**(1), 2000 pp. 87–98. doi:10.1142/S0219525900000078.

[11] R. Pfitzner, I. Scholtes, A. Garas, C. J. Tessone and F. Schweitzer, "Betweenness Preference: Quantifying Correlations in the Topological Dynamics of Temporal Networks," *Physical Review Letters*, **110**(19), 2013 198701. doi:10.1103/PhysRevLett.110.198701.

[12] I. Scholtes, N. Wider, R. Pfitzner, A. Garas, C. J. Tessone and F. Schweitzer, "Causality-Driven Slow-Down and Speed-Up of Diffusion in Non-Markovian Temporal Networks," *Nature Communications*, **5**(1), 2014 5024. doi:10.1038/ncomms6024.

[13] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West and R. Lambiotte, "Memory in Network Flows and Its Effects on Spreading Dynamics and Community Detection," *Nature Communications*, **5**(1), 2014 4630. doi:10.1038/ncomms5630.

[14] I. Scholtes, "When Is a Network a Network? Multi-order Graphical Model Selection in Pathways and Temporal Networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'17)*, Halifax, NS, Canada, 2017, New York: Association for Computing Machinery, 2017 pp. 1037–1046. doi:10.1145/3097983.3098145.

[15] I. Scholtes, N. Wider and A. Garas, "Higher-Order Aggregate Networks in the Analysis of Temporal Networks: Path Structures and Centralities," *The European Physical Journal B*, **89**(3), 2016 61. doi:10.1140/epjb/e2016-60663-0.

[16] H. H. K. Lentz, T. Selhorst and I. M. Sokolov, "Unfolding Accessibility Provides a Macroscopic Approach to Temporal Networks," *Physical Review Letters*, **110**(11), 2013 118701. doi:10.1103/physrevlett.110.118701.

[17] SocioPatterns. (Dec 22, 2020) www.sociopatterns.org.

[18] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton and A. Vespignani, "Dynamics of Person-to-Person Interactions from Distributed RFID Sensor Networks," *PLoS ONE*, **5**(7), 2010 e11596. doi:10.1371/journal.pone.0011596.

[19] J. Fournet and A. Barrat, "Contact Patterns among High School Students," *PLoS ONE*, **9**(9), 2014 e107878. doi:10.1371/journal.pone.0107878.

[20] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B.-a Kim, B. Comte and N. Voirin, "Estimating Potential Infection Transmission Routes in Hospital Wards Using Wearable Proximity Sensors," *PLoS ONE*, **8**(9), 2013 e73970. doi:10.1371/journal.pone.0073970.

[21] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton and W. Van den Broeck, "What's in a Crowd? Analysis of Face-to-Face Behavioral Networks," *Journal of Theoretical Biology*, **271**(1), 2011 pp. 166–180. doi:10.1016/j.jtbi.2010.11.033.

[22] M. Génois and A. Barrat, "Can Co-location be Used as a Proxy for Face-to-Face Contacts?," *EPJ Data Science*, **7**(1), 2018 11. doi:10.1140/epjds/s13688-018-0140-1.

[23] F. Iannone, F. Ambrosino, G. Bracco, M. De Rosa, A. Funel, G. Guarnieri, S. Migliori, et al., "CRESCO ENEA HPC Clusters: A Working Example of a Multifabric GPFS Spectrum Scale Layout," in *2019 International Conference on High Performance Computing Simulation (HPCS)*, Dublin, Ireland 2019, Piscataway, NJ: IEEE, 2019 pp. 1051–1052. doi:10.1109/HPCS48598.2019.9188135.