HAL
open science

Estimating the outcome of spreading processes on networks with incomplete information: a mesoscale approach

Anna Sapienza, Alain Barrat, Ciro Cattuto, Laetitia Gauvin

To cite this version:

Anna Sapienza, Alain Barrat, Ciro Cattuto, Laetitia Gauvin. Estimating the outcome of spreading processes on networks with incomplete information: a mesoscale approach. Physical Review E , 2018, 98 (1), pp.012317. 10.1103/PhysRevE.98.012317 . hal-02065351

HAL Id: hal-02065351

https://hal.science/hal-02065351

Submitted on 12 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating the outcome of spreading processes on networks with incomplete information: a mesoscale approach

Anna Sapienza
USC Information Sciences Institute, CA, USA and
Data Science Laboratory, ISI Foundation, Turin, Italy

Alain Barrat
Aix Marseille Univ, Université de Toulon, CNRS, CPT, Marseille, France and
Data Science Laboratory, ISI Foundation, Turin, Italy

Ciro Cattuto and Laetitia Gauvin
Data Science Laboratory, ISI Foundation, Turin, Italy
(Dated: September 7, 2017)

arXiv:1709.01806v1 [physics.soc-ph] 6 Sep 2017

Recent advances in data collection have facilitated the access to time-resolved human proximity data that can conveniently be represented as temporal networks of contacts between individuals. While this type of data is fundamental to investigate how information or diseases propagate in a population, it often suffers from incompleteness, which possibly leads to biased conclusions. A major challenge is thus to estimate the outcome of spreading processes occurring on temporal networks built from partial information. To cope with this problem, we devise an approach based on Non-negative Tensor Factorization (NTF) – a dimensionality reduction technique from multilinear algebra. The key idea is to learn a low-dimensional representation of the temporal network built from partial information, to adapt it to take into account temporal and structural heterogeneity properties known to be crucial for spreading processes occurring on networks, and to construct in this way a surrogate network similar to the complete original network.

To test our method, we consider several human-proximity networks, on which we simulate a loss of data. Using our approach on the resulting partial networks, we build a surrogate version of the complete network for each. We then compare the outcome of a spreading process on the complete networks (non altered by a loss of data) and on the surrogate networks. We observe that the epidemic sizes obtained using the surrogate networks are in good agreement with those measured on the complete networks. Finally, we propose an extension of our framework when additional data sources are available to cope with the missing data problem.

Keywords: Temporal network, tensor decomposition, joint factorization, missing data, partial information

I. INTRODUCTION

The advances made in data collection technologies have led to a wealth of high-resolution time-resolved data. Mobile sensing devices, social networking applications and wearable sensors have indeed significantly contributed to monitor social interactions and physical proximity of individuals in time [1–6]. Such fine-grained data monitoring is crucial for a deeper study of human proximity dynamics described by complex temporal networks – in which links are drawn between nodes representing individuals when they are in close-range [7] – and their interplay with contagion processes. Physical proximity interactions play indeed a fundamental role in conveying information or in the spread of diseases [8–11]. They can thus inform our understanding of how messages or infectious diseases such as flu-like illnesses propagate among individuals.

However, despite the efforts made to increase the accuracy in the data collection, relational data often suffer from incompleteness resulting in missing links in empirical networks [12]. This lack of information can arise for several reasons: limited participation during surveys, incomplete records (diary-based, device-based) [13–16], and technical issues occurring during the data collection process. In the case of smartphone sensing such as in [3, 5], proximity might be undetected during some time windows. For instance, people might turn off their Bluetooth or Communication Data Records might not provide access to the vicinity of individuals to a cell tower at each time, possibly leading to undetected co-presence events. In the case of self-reporting of sexual relationships, individuals might choose not to disclose all of their partners, which leads to biases when focusing on the spread of sexually transmitted disease [17]. Data incompleteness can affect the measured properties and structure of temporal networks [18, 19] and can reflect on the simulated evolutions of contagion processes, leading to inaccurate conclusions [20, 21]. The investigation of information or disease propagation processes on networks built from such data must thus be undertaken carefully.

Several approaches have been put forward to cope with missing links in networks [22]. These methods include: distributional models, which estimate the likelihood of the presence of a link on the basis of the observed links and nodes attributes [23], hierarchical structure methods [24], stochastic block models [25], and expectation maximization methods [26], which try to extract the con-

nectivity patterns in the available part of the network to infer and complete the unknown part. However, the goal of these methods is the exact recovery of single links, a complicated task that becomes nearly impossible to achieve when a large amount of data is missing. Moreover, this might actually not be necessary if the goal is to estimate the global outcome of a contagion process at the population level and not the risk concerning a specific individual. Other approaches have been developed to specifically estimate important properties of epidemic spread and information cascade without trying to recover the original network [15, 20, 27–29]. These methods are however either process specific or rely on the existence of known mesocale structures (such as groups in the population) in the network, together with the knowledge of the structure to which each population member belongs.

Here, we propose a self-contained approach that does not rely on such a priori knowledge. As in [15, 20], the aim of this work is not to recover the exact links which are missing in the data but instead to build a surrogate version of the network of interactions. The main difference with these previous works, that is a building block of our approach, is that we uncover *in the incomplete data* mesoscale structures that involve nodes affected by missing activity. To study the network at the mesoscale level, we take advantage of tensor decomposition techniques, already applied in several fields [30], to extract both the topological and temporal properties of the network [31]. An important aspect of our method is that it does not rely on the availability of metadata or external information on the nodes, and that it allows us to recover fundamental properties of the studied temporal network, such as the temporal activity of nodes with partial information (i.e., the temporal evolution of their number of contacts). We use this approach to build a surrogate version of the network that yields a correct estimate of the outcome of a simulated spreading process occurring on the network. Furthermore, we extend our method to take advantage of other sources of information that might be available, such as information deriving from subsidiary data sources.

The paper is organized as follows: in Section II we present the notations used in the paper; in section III we describe the problem statement; in Section IV we explain the method that we developed to carry out the study; in Section V we report the results achieved by our approach in the study of several temporal human proximity networks; in Section VI we discuss the performance and limitations of the method and future research directions.

II. NOTATIONS

The following notations are used throughout the present paper. Lower-case letters denote scalar variables, e.g., t, capital letters denote defined constants, e.g., T, and boldface lower-case letters denote vectors, e.g., $\mathbf{t}$.

Matrices are denoted by boldface capital letters, e.g., $\mathbf{T}$, where the i-th column of a matrix $\mathbf{T}$ is $\mathbf{t}_i$ and the (i, j)-entry is t_{ij}. Third order tensors are denoted by bold calligraphic letters, e.g., $\boldsymbol{\mathcal{T}}$, whose (i, j, k)-entry is t_{ijk}.

The tensor product is denoted by $\circ$, the Hadamard (element-wise) product by $*$, the Kronecker product by $\otimes$, and the outer product by $\cdot$. $\| \cdot \|_F$ stands for the Frobenius norm.

III. PROBLEM STATEMENT

The purpose of the present work is the development of a methodology able to reproduce the outcome of contagion processes on temporal networks of human interactions, starting from incomplete information on these networks. We consider a scenario in which part of the activity of a fraction of the network nodes (i.e. part of their interactions over time) is missing in the data. To provide an estimate of the outcome of spreading processes on the network, we do not try to recover the exact missing links and interaction events of these nodes. Our method aims instead at building a surrogate version of the complete network by taking advantage of the only information that is available. As we will describe in details in Sec. IV A and Sec. IV B, this information can be related either only to the partial network of human contacts or to richer data than can be used as a proxy for human proximity. For instance, we could have access both to the partial temporal contact network and to the approximated location of individuals provided by smartphones, through the GPS, Bluetooth or WiFi signals: such additional information can conveniently be represented as a temporal bipartite network between individuals and locations, in which a link is drawn between an individual and a location when the individual is detected in that location. We will show how such information can be integrated in our framework.

The underlying assumption of our method is that we can leverage the mesoscale properties of the partial network, such as the presence of correlations in the node activities, to build a surrogate version of the complete network. To extract these mesoscale properties from the incomplete data we rely on the Non-negative Tensor Factorization (NTF) technique [30, 31]. In particular, our method is based on a NTF framework handling missing values [32, 33]. By applying the NTF on a tensor representing a temporal network we can indeed identify groups of nodes having similar connectivity patterns and whose links have similar activation times. Each of these groups can be seen as a sub-network. By studying the structure and the temporal activity of each sub-network, we can infer the properties of the nodes whose activity is partially missing.

However, standard tensor decomposition techniques such as NTF are based on the assumption that noise in the data follows a Gaussian distribution. Thus, even if the data are intrinsically heterogeneous, as is typically

the case in human contact data [34], the decomposition will provide a low-rank approximation in which these heterogeneous characteristics are not preserved [35]. As a result, the NTF tends to approximate the network in a way that makes the properties of the sub-networks extracted more homogeneous than in the original network. This trend towards homogeneity changes the temporal and structural properties of the network such as the distribution of the number of contacts per link (i.e., the number of activation times of each link) in the aggregated network, and thus has to be carefully taken into account. Indeed, heterogeneity, both in the aggregated number or duration of contacts per link and in temporal properties (e.g., broad inter-event time distributions) have been shown to play a key role in spreading processes occurring over static and temporal networks [36–42].

Since we are interested in recovering the outcome of such processes, we need to preserve the heterogeneity properties of the network. Thus, instead of directly using the approximated network provided by the NTF, we will use the information about the nodes belonging to each sub-network and their related activity patterns as a guide to build a surrogate network. In this surrogate network we will reintroduce the heterogeneity properties as described below.

Our method to build a surrogate network from partial information and recover the outcome of contagion processes is thus divided in two main steps:

A . The extraction of mesoscale structures from the partial temporal network through NTF adapted to handle incomplete information;

B . the construction of a surrogate network, in which we use both the information provided by step A and the heterogeneity properties of the network.

IV. METHOD

A. Extracting mesoscale structures from a partial temporal network

Three-way tensors (i.e. three-dimensional arrays) are natural representations of temporal networks: given an undirected temporal network, composed by N nodes and $k = 1, \ldots, K$ time intervals, we can represent its snapshots $G_k = (V, E_k)$, which have $|V| = N$ nodes and set of links E_k, by K adjacency matrices $\mathbf{M}_k \in \mathbb{R}^{N\times N}$ of the form:

$$\mathbf{M}_k = \begin{cases} m_{ij} = 1 \text{ if } (i,j) \in E_k\,, \\ m_{ij} = 0 \text{ otherwise}\,. \end{cases}$$

These adjacency matrices form the slices of a tensor $\boldsymbol{\mathcal{T}} \in \mathbb{R}^{I\times J\times K}$, with here $I = J = N$. In the case of missing data, zero values in the tensor can either correspond to no activity or to undetected activity. We however need to factorize only the part of the tensor that corresponds either to measured activity or to actual inactivity (absence of contact). To this aim, the zero entries that correspond to possibly undetected activities have to be masked in the tensor. If a node with partial information has no activity at all measured during a given snapshot (in a given slice of the tensor), we consider that the related zero entries in the tensor $\boldsymbol{\mathcal{T}}$ might correspond to possibly undetected activities, i.e., possibly missing contacts. On the other hand, if a node, for which we know that only partial information is present, has at least one contact measured in a slice, we assume that no information was lost at all in that tensor slice because of that node. In other terms, we assume for each node and each time slice that either all or none of the activity recorded by that node in that time slice is present in the data. This would be for instance the case if a sensor measuring proximity is turned off or if the GPS coordinates of a mobile user are not collected during a time window: all the proximity information concerning this sensor in this time window is then lost. However, if the measuring device is not turned off, all the proximity relationships with other devices that are also on during that time window are present in the incomplete data.

We thus introduce a binary tensor $\boldsymbol{\mathcal{W}}$, of the same size of $\boldsymbol{\mathcal{T}}$, whose entries are defined as

$$w_{ijk} = \begin{cases} 0 \text{ if } i \text{ or } j \text{ has no activity in the time window } k, \\ 1 \text{ otherwise.} \end{cases}$$

This tensor is used to mask the part of the tensor $\boldsymbol{\mathcal{T}}$ that might be linked to possibly undetected activity. The approximation of the masked tensor $\boldsymbol{\mathcal{T}} * \boldsymbol{\mathcal{W}}$ consists in minimizing the following cost function with non-negative constraints [33]:

$$f_w\left(\boldsymbol{\lambda}, \mathbf{A}, \mathbf{B}, \mathbf{C}\right) = \left\| \boldsymbol{\mathcal{W}} * \left(\boldsymbol{\mathcal{T}} - \sum_{r=1}^{R} \lambda_r \mathbf{a}_r \cdot \mathbf{b}_r \cdot \mathbf{c}_r \right) \right\|_F^2 ,$$

where R is the rank of the approximation and is hereinafter called the number of components. The vectors $\mathbf{a}_r$, $\mathbf{b}_r$, and $\mathbf{c}_r$ with $r \in [1, R]$ form respectively the factor matrices $\mathbf{A} \in \mathbb{R}^{I\times R}$, $\mathbf{B} \in \mathbb{R}^{J\times R}$, $\mathbf{C} \in \mathbb{R}^{K\times R}$. Each tuple of vectors $(\mathbf{a}_r, \mathbf{b}_r, \mathbf{c}_r)$ describes a mesocale strucure. As we explain below, $\mathbf{a}_r$ and $\mathbf{b}_r$ indicate which nodes and links participate to the mesoscale structure r, while $\mathbf{c}_r$ describes the temporal activity of the structure r. For the sake of readability, the cost function is re-written in the following form:

$$f_w\left(\boldsymbol{\lambda}, \mathbf{A}, \mathbf{B}, \mathbf{C}\right) = \| \boldsymbol{\mathcal{W}} * \left(\boldsymbol{\mathcal{T}} - [\![\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!] \right) \|_F^2 . \quad (1)$$

Using the mask $\boldsymbol{\mathcal{W}}$ amounts to approximating the tensor $\boldsymbol{\mathcal{T}}$ based only on information of which we are certain: we are approximating only the part of the tensor composed of elements that are either 1 corresponding to measured events or 0 corresponding to real absence of contact, without taking into account all the 0 values that might correspond to undetected activity.

In other terms, minimizing the cost function f_w corresponds to finding the best approximation of the non-masked part of the tensor by a sum of components corresponding each to a mesoscale structure. The analysis of $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ yields then information on which link is involved in which mesocale structure, including in particular the links involving nodes with incomplete information. Several methods are available to estimate the factor matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ [43, 44], and details are given in Section VII.

B. Extracting mesoscale structures from coupled temporal networks

Here, we propose an extension of the mesoscale structure detection method to the case in which we have access to richer information – not expressible in one single temporal network. The integration of such information might help to better recover the missing entries in the tensor describing the temporal contact network: for instance, if we have access, in addition to the partial contact network, to the location of individuals involved in the contacts, the latter can contribute to recover the missing information on contacts. To integrate such additional data, we propose to use the so-called Joint Non-negative Tensor Factorization (JNTF) [45] that makes it possible to decompose multiple temporal networks at once in a coupled manner (in practice, we will consider the partial temporal contact network and a position network evolving with time).

Let us consider the general case of S different temporal networks, each represented by a tensor $\boldsymbol{\mathcal{T}}_s$, with possibly different dimensions as they might represent different types of information. We can approximate them in a coupled way by computing the following minimization problem with non-negative constraints:

$$\min \sum_{s=1}^{S} \|\boldsymbol{\mathcal{T}}_s - [\![\boldsymbol{\lambda}_s; \mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s]\!]\|_F^2 \qquad (2)$$
$$\text{s.t. } \boldsymbol{\lambda}_s, \mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s \geq 0\,.$$

Different couplings can be considered by imposing that some of the factor matrices $\mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s$ in the equation are equal for different values of s. The idea behind the introduction of this coupling is that different networks can provide partially redundant information, and that this redundancy is relevant for recovering missing entries. For instance, if we have access to the locations of nodes but only to a partial information concerning their contacts, we can couple the decomposition of the tensor representing the partial contact network to the decomposition of the tensor representing the time evolution of the location of nodes, and gain in this way information on the possible contacts over time. In practice, we would impose $\mathbf{A}_1 = \mathbf{A}_2$ and $\mathbf{C}_1 = \mathbf{C}_2$, which correspond respectively to imposing the same nodes' memberships and the same activity timeline for each mesoscale structure in the resulting approximations of the contact and location tensors.

The joint factorization of tensors – including one with missing information $\boldsymbol{\mathcal{T}}_{s'}$ –, can be adapted to handle missing values in the same way as the NTF:

$$\min\Big(\alpha_{s'}\|\boldsymbol{\mathcal{W}} * (\boldsymbol{\mathcal{T}}_{s'} - [\![\boldsymbol{\lambda}_{s'}; \mathbf{A}_{s'}, \mathbf{B}_{s'}, \mathbf{C}_{s'}]\!])\,\|_F^2$$
$$+ \quad \sum_{\substack{s=1\\ s\neq s'}}^{S} \alpha_s \|\boldsymbol{\mathcal{T}}_s - [\![\boldsymbol{\lambda}_s; \mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s]\!]\|_F^2\Big)$$
$$\text{s.t. } \boldsymbol{\lambda}_s, \mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s \geq 0 \;\forall s\,.$$

To solve this minimization problem we adapted the active-set-like method with Karush-Kuhn-Tucker optimality conditions [43, 46] (see Section VII for details). In the following, we set all the α_s parameters equal to 1. We note however that they could be used to tune the relevance of the information provided by each tensor. As an example, by setting $\alpha_{s'} > \alpha_s$, $\forall s \neq s'$ we would give more importance to the information provided by the network with partial information than to the one given by the subsidiary data.

C. Surrogate network

Both types of factorization, either from a partial network or from coupled temporal networks, provide approximations $\boldsymbol{\mathcal{T}}_{app}$ of the temporal networks as sums of components, each corresponding to a mesoscale structure. Each mesoscale structure is fully described by the respective columns of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ and can be seen as a sub-network composed of links with similar temporal properties, as we now describe.

To analyze the mesoscale structures revealed by the factorization step and interpret them as sub-networks, we estimate the memberships of the links using $\mathbf{A}$ and $\mathbf{B}$ and we determine the activation times by binarizing $\mathbf{C}$. First, the membership weight of each link (i, j) to the r-th component is given by the $(i, j)^{th}$ element of the following Kronecker product:

$$\mathbf{a}_r \otimes \mathbf{b}_r = \mathbf{a}_r \cdot \mathbf{b}_r^T\,.$$

As we consider undirected networks, the actual membership of each link is symmetrized in the following way:

$$\frac{\mathbf{a}_r \cdot \mathbf{b}_r^T + \mathbf{b}_r \cdot \mathbf{a}_r^T}{2}.$$

We consider that the links with the largest membership weights, composing 95% of the total sum of the squared memberships, belong to the component r (note that, depending on the membership values, a link could belong to more than one component). Then, to detect the times in which each mesoscale structure is active, we consider the matrix $\mathbf{C}$, which summarizes the temporal activity of each component: by using the Otsu method [47], a common way to perform binary thresholding, we transform

the temporal activity of each component in its binary version (1 if it is active, 0 otherwise).

These two steps allow us to determine for each mesoscale structure when it is active and which links it involves. For each structure, we then select the links involving at least one node with partial information and the times in which its activity was potentially lost (in the sense that no activity of that node was recorded at all for this particular time), and we add the corresponding elements of the structure to the partial network $\boldsymbol{\mathcal{T}}$ to create the surrogate tensor $\boldsymbol{\mathcal{T}}_{surrogate}$. The rationale is that the factorization has allowed us to determine, for the links for which only partial information is available, to which mesoscale structures they belong. Using the activity timelines of the mesoscale structures, we thus reconstruct the missing parts of the activity timelines of these links. Moreover, the binarization step enables to reintroduce the heterogeneity in the network. As mentioned previously, the properties of the sub-networks extracted tend to be more homogenous, in particular the distribution of weights of each sub-network is more homogenous than in the real data. Indeed, the factorization finds groups of links that have strongly correlated activity timelines and approximates them as fully correlated. As a result, the distribution of the link weights in the network described by $\boldsymbol{\mathcal{T}}_{app}$ – the number of times in which each link is active – becomes more homogeneous and the activity tends to be less bursty than in the empirical network. By binarizing the approximation in the way described above we are removing links which were associated to small weights and increasing weights of links with the highest weights in each component of the approximated case. This contributes to make the overall distribution of weights more heterogenous.

When the JNTF is used, an additional step turns out to be relevant. The JNTF indeed yields components influenced by both the activity in the partial proximity network and the subsidiary temporal network (such as a network linking individuals and locations). Thus, the JNTF will approximate the network in a way based also on co-presence events of individuals. As co-presence is a necessary but not sufficient condition for two individuals to be in contact, the links added when creating $\boldsymbol{\mathcal{T}}_{app}$ are less likely to correspond exactly to contacts that were missing than in the NTF case. Moreover, for the same reason the number of links in $\boldsymbol{\mathcal{T}}_{app}$ and their weights might be greater than in the original data and they might present less bursty activity patterns. As mentioned in IV B, the weight of the information provided by the second tensor can of course be tuned in the JNTF by adding coefficients so that the components extracted by the factorization are more representative of the network with partial information than of the other one. The investigation of this possibility is however outside the scope of this paper. The procedure of binarization of $\boldsymbol{\mathcal{T}}_{app}$ described above needs thus to be completed to make the properties of the network more heterogeneous. In order to do this, we have to discard some of the times in which the links involving nodes with partial information are active in $\boldsymbol{\mathcal{T}}_{app}$ to compensate for the overestimation and get closer to the empirical network. The key idea here is to zero out some elements in $\boldsymbol{\mathcal{T}}_{surrogate}$ to recover a weight distribution comparable to the empirical one, measured on the links involving only nodes with no missing information. We pick from this distribution a number of weights equal to the number of links having partial information, for which we have to adjust the weights, and we assign them at random to these links. Finally, we compare for each link the new value to its weight in $\boldsymbol{\mathcal{T}}_{surrogate}$: if the new weight is smaller than the old one, we erase at random parts of the link activity that are present in $\boldsymbol{\mathcal{T}}_{surrogate}$, until we reach the new weight; if instead the new weight is larger than the old one, we do not act on that link's activity. The reason why we can rely on the weight distribution measured on the partial network is due to its robustness to sampling, as shown in [4, 15, 20, 27] for various sampling procedures and in the Supplementary Information S.VII D for the sampling considered here. After the reassignment of weights, the resulting tensor $\boldsymbol{\mathcal{T}}_{surrogate}$ is used to approximate the whole temporal network of contacts and perform simulations of spreading processes.

V. RESULTS

We apply the method described in the previous sections to three temporal human proximity networks. The data were collected by the SocioPatterns collaboration (www.sociopatterns.org) in two conferences in Italy (HT09) and France (SFHH), and in a primary school in France (LSCH). In each case, the proximity of individuals in the network was measured with wearable sensors able to capture face-to-face contacts occurring in a $1-2$-meter range with a 20-second time resolution. As for the purpose of the present paper we do not need such a high-resolution, we aggregate the data to a 15-minute resolution. The HT09 dataset was collected during the ACM Hypertext 2009 conference [48]. The resulting temporal network has $N = 113$ nodes and $K = 237$ snapshots. The SFHH dataset was collected during the conference of the Société Française d'Hygiène Hospitalière, yielding a temporal network with $N = 417$ nodes and $K = 129$ time snapshots [4]. Finally, the LSCH dataset was collected in a primary school in Lyon [49]. The resulting network has $N = 241$ nodes and $K = 130$ snapshots. We consider all the snapshots as unweighted networks (meaning that each element of the tensor is either 0 or 1).

We simulate on each dataset a loss of data determined by a fraction of nodes $p_{nodes} \in [0.1, 0.2, 0.4]$ for which we zero out the activity occurring during the first half of the total timespan, i.e., the information concerning each of these nodes is lost over a fraction $p_{times} = 0.5$ of the temporal snapshots. In practice, for each dataset we start from the complete tensor $\hat{\boldsymbol{\mathcal{T}}}$ and we create a tensor with partial information $\boldsymbol{\mathcal{T}}$ by erasing, in the first half of the temporal slices of $\hat{\boldsymbol{\mathcal{T}}}$, the elements related to Np_{nodes}

chosen at random. To test the limits of our method, we also considered cases where either all the nodes lost some information or where a fraction of the nodes lost all their activity. More details on the cases dealt with are provided below.

A. Approximated network

For each dataset and for each data loss scenario, we performed the approximation of $\boldsymbol{\mathcal{T}}$, i.e., the minimization of Eq. (1), with a number of components selected using the so called Core Consistency Diagnostic [50] as a guide (see Sec. VII for details). We perform 20 decompositions in each case, varying the initial conditions, and we use the one with the highest core consistency value. Moreover, for each value of p_{nodes} we repeat the procedure on 10 different sets of nodes with missing information chosen at random, in order to evaluate the variations in the performance of our method.

From the factorization of the tensors with missing information $\boldsymbol{\mathcal{T}}_{HT09}$, $\boldsymbol{\mathcal{T}}_{SFHH}$, and $\boldsymbol{\mathcal{T}}_{LSCH}$, we recover a first approximated version of the complete network $\boldsymbol{\mathcal{T}}_{app}$ for each dataset. We evaluate the results provided by the decomposition by computing the Pearson's coefficient between the complete and approximated temporal activities of the nodes for which only partial information is available in $\boldsymbol{\mathcal{T}}$. The correlation is measured only on the part of the activity that is missing in the partial data. We report the correlation coefficients found in Table I and we show in the Supplementary Information S.VII D a representative example of the temporal activities of these nodes with partial information in the complete, partial, and approximated networks. The correlation coefficients shown are measured on 10 different sets of nodes with missing information for each dataset. The ranges of Pearson's coefficient indicate positive moderate to high correlation between the complete and approximated node activity, indicating a good recovery of the node temporal activities. Let us notice that by construction the approximation through the factorization relies on the existence of correlated activity patterns and consequently performs better in the presence of strong such patterns. This explains why the temporal activities are better recovered (larger Pearson correlation coefficients) in the LSCH case than in HT09 and SFHH. Indeed, in schools the schedule is quite constrained and all the students of a given class have highly correlated activity timelines, leading to stronger correlated activity patterns than in conferences, during which attendees are free to move and interact with different people at different times.

B. Surrogate network

As illustrated in the previous paragraphs, the approximation step achieves good results to recover single node temporal activities by using NTF handling missing information. However, as discussed above, heterogeneity properties of the link weights are crucial in determining the outcome of spreading processes. We show indeed in the Supplementary Information S.VII D that simulating a spreading process directly on an approximated network $\boldsymbol{\mathcal{T}}_{app}$ without the binarization step results in a strong underestimation of the original outcome of the spreading process. We thus apply the procedure described above and build a surrogate temporal network $\boldsymbol{\mathcal{T}}_{surrogate}$ relevant to estimate the outcome of spreading processes. For the JNTF, for each link involving nodes with partial information, we assign a weight extracted from the distribution of weights measured on the links with no missing information, and, if needed, we remove a part of the link's activity present in $\boldsymbol{\mathcal{T}}_{surrogate}$ but not in $\boldsymbol{\mathcal{T}}$ until we match this weight.

Table I. Range of values of the Pearson's coefficient computed by comparing the original and approximated node activity for each node in the different sets considered. The table also reports the median value obtained. The p-values are lower than 10^{-3}.

	p_{nodes}	p_{times}	Pearson's coeff.	median value
LSCH	0.1	0.5	[0.53, 0.96]	0.77
	0.2	0.5	[0.54, 0.96]	0.76
	0.4	0.5	[0.53, 0.96]	0.77
HT09	0.1	0.5	[0.38, 0.83]	0.54
	0.2	0.5	[0.37, 0.80]	0.52
	0.4	0.5	[0.37, 0.73]	0.50
SFHH	0.1	0.5	[0.40, 0.89]	0.57
	0.2	0.5	[0.48, 0.93]	0.57
	0.4	0.5	[0.40, 0.89]	0.57

C. Estimate of the outcome of spreading processes

To evaluate the performance of the method in estimating the outcome of spreading processes on the network, we simulate susceptible-infected-recovered (SIR) processes (see details in Sec. VII D) over three temporal networks for each dataset: the complete, the partial, and the surrogate network. In each case we run multiple simulations for each set of values (β, μ) of the infection and recovery probabilities per unit time. We focus on the couples of probabilities (β, μ) that satisfy the following criterion. The spreading has to be finished within the timespan of the dataset, and the epidemic size, defined as the final fraction of recovered individuals, has to be greater than 20% and lower than 80% (the selection is based on the median of the epidemic size). These conditions ensure to avoid the selection of parameters such that the simulations either never reach a significant epidemic size and/or are too slow with respect to the total timespan of the network. The limit for the final number of recovered individuals to 80% of the entire population prevents the selection of parameters leading to too fast spreading that are far from realistic conditions.

For each selected pair of parameter values β and μ,

we compute the distribution of the epidemic sizes in the three cases (here called "Complete", "Partial", "Surrogate"). As we simulate a loss of data by considering 10 different sets of randomly chosen nodes with partial information, we report the median distribution of the epidemic size on the simulations over these 10 cases as well as the 25-th and 75-th percentiles. We first consider the case in which the only available information is the partial temporal network $\boldsymbol{\mathcal{T}}$, and then a case in which an additional source of information is available, namely the location of individuals over time.

1. *Using only the partial temporal network*

In Fig. 1 we report the epidemic size distributions obtained for the LSCH dataset, for two couples of selected spreading parameters: a) $\beta = 0.3$, $\mu = 0.3$ and b) $\beta = 0.15$, $\mu = 0.25$ and for different fractions of nodes p_{nodes} with partial information. The figure shows that our method yields a good approximation of the distribution obtained in the original case, while the one obtained by simply simulating the SIR process on the network with partial information strongly underestimates the epidemic size. This underestimation becomes stronger as p_{nodes} increases while the estimation given by the surrogate network remains of good quality.

We obtain similar results for the HT09 and SFHH datasets, for which we report the results of numerical simulations of the SIR process for various values of the spreading parameters (β,μ) and of p_{nodes} in Fig.s 2 and 3. In all cases, using only the partial temporal network in the numerical simulations of the SIR process leads to a clear underestimation of the epidemic sizes; this underestimation becomes worse as p_{nodes} increases. The simulations performed on the surrogate network are systematically much closer to the ones based on the complete information, leading thus to a strong improvement in the prediction of the epidemic risk. For the SFHH case however, we observe that the agreement becomes worse when p_{nodes} increases. This is due to the fact that during large conferences attendees tend to follow the schedule less rigorously than in small ones and move around and engage in contacts more freely and in a more random way, thus leading to less correlated activity patterns. In such a case, auxiliary data could be useful, as in the case we describe in the next section.

2. *Using both the partial network and an additional proxy*

By construction, the method based on the NTF cannot handle extreme cases such as missing information for all nodes at the same time or activity fully missing for some nodes (as considered in [20]). To address this limitation, we propose an extension of our method that makes it possible to take advantage of additional information that can be used as a proxy for human proximity. To this aim, we consider the JNTF method described in Sec. IV B. We test this extended method on the LSCH dataset, for which we have access to the approximate position of individuals in time. There are indeed 15 locations in the school: 10 classes, the cafeteria, the playground, two staircases and a control room. The resulting bipartite temporal network relating individuals and locations is composed by $N = 241$ nodes representing individuals, 15 nodes representing locations, and 130 temporal snapshots: the tensor representing this additional information has dimensions $I = 241$, $J = 15$ and $K = 130$. Let us notice that due to the temporal resolution selected, a node might appear in several locations in the same snapshot.

To compare the methods of construction of surrogate data based on the JNTF and on the NTF decompositions, we simulated a loss of data on the contact network for a fraction of nodes $p_{nodes} = 0.2$ and for several fractions of the timespan $p_{times} = [0.6, 0.8, 1]$ selected consecutively on the temporal activity of the nodes. Here, the set of nodes with partial activity is the same for all values of p_{times}, so that we can compare the outcomes of the SIR process and the impact of incrementally removing larger fractions of the temporal activity of the same nodes. We build surrogate data using the methods based on the NTF and on the JNTF decompositions, performed respectively on the partial contact networks and on the joint partial contact and location networks. For the JNTF, we impose a coupling on the first and third dimensions (i.e., we impose the equality of the matrices obtained in the decompositions of the two tensors, for $\mathbf{A}$ on the one hand and for $\mathbf{C}$ on the other hand), as the two networks have the same nodes related to individuals and the same snapshots in time. The choice of this coupling relies on the reasonable assumption that co-located individuals are more likely to be in contact.

In Fig. 4, we display the results obtained for $p_{nodes} = 0.2$ and $p_{times} = 0.6, 0.8, 1$. The different subfigures in a) and b) correspond to SIR processes with the following infection and recovery probabilities: $\beta = 0.3$, $\mu = 0.3$, and $\beta = 0.15$ and $\mu = 0.25$. As in the previous cases, the distributions of epidemic sizes computed on the network with partial information show a clear underestimation of the number of individuals infected. For $p_{times} = 0.6$ and $p_{times} = 0.8$, the results achieved with both procedures based on the NTF and the JNTF are in agreement and give a better estimation of the distributions of epidemic sizes obtained in the original network than using the incomplete data. Finally, we report in the rightmost panels of each subfigure the extreme case for which no information about 20% of nodes is present in the partial contact network ($p_{times} = 1$). In this case, as no information is present at all for the selected nodes, no correlated activity pattern concerning them can be inferred by the NTF decomposition, which thus cannot help recover any information on these nodes. Using JNTF proves then helpful and the surrogate network built using this method, which combines contact and location information, yields

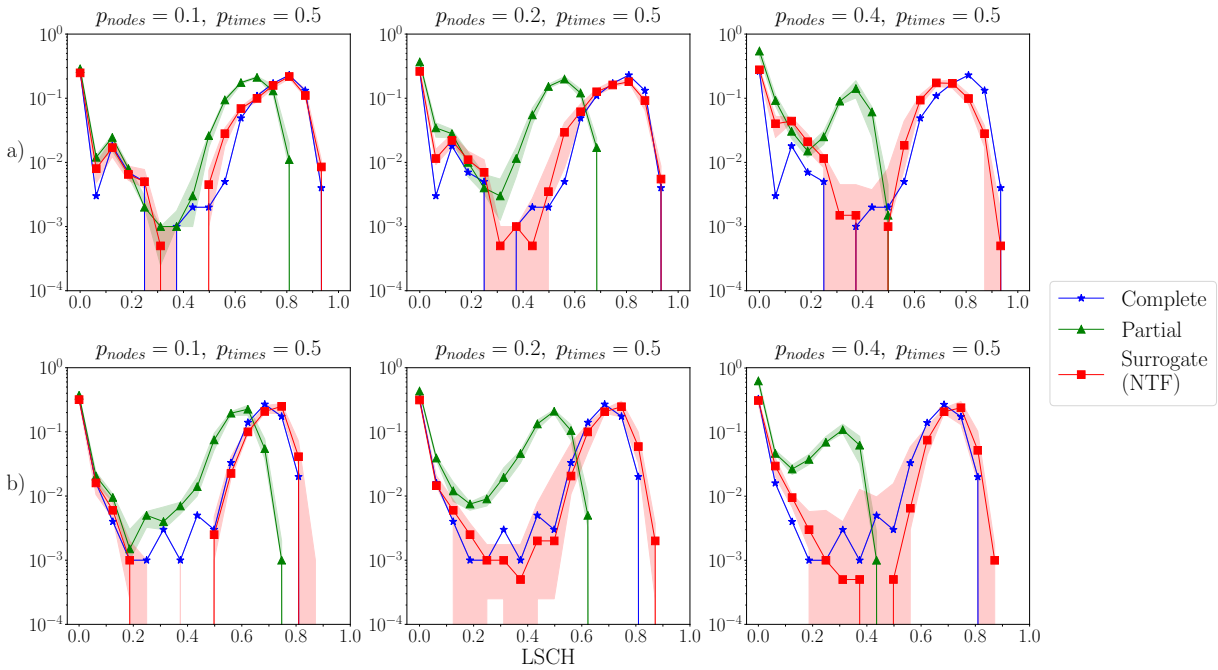

Figure 1. Distributions of epidemic sizes computed in the complete, partial, and surrogate cases for the LSCH dataset. Each panel corresponds to one value of the fraction p_{nodes} of nodes with partial information and one couple of spreading parameters: a) $\beta = 0.3$ and $\mu = 0.3$; b) $\beta = 0.15$ and $\mu = 0.25$. For the partial and surrogate cases, the symbols and lines show the median distribution of the epidemic size computed from the results relative to the 10 different sets of nodes with partial information, while the shaded area is delimited by the 25-th and 75-th percentiles.

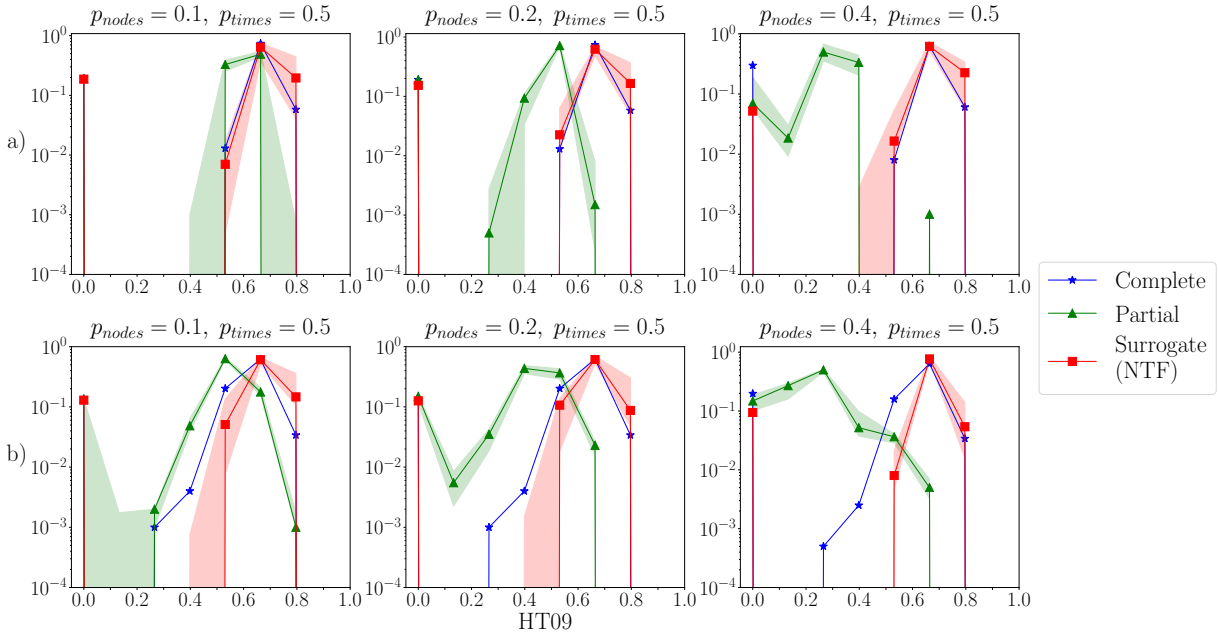

Figure 2. Distributions of epidemic sizes computed in the complete, partial, and surrogate cases for the HT09 dataset. Each panel corresponds to one value of the fraction p_{nodes} of nodes with partial information and one couple of spreading parameters: a) $\beta = 0.60$ and $\mu = 0.10$; b) $\beta = 0.25$ and $\mu = 0.05$. The symbols and lines show the median distribution of the epidemic size computed from the results relative to the 10 different sets of nodes with partial information, while the shaded area is delimited by the 25-th and 75-th percentiles.

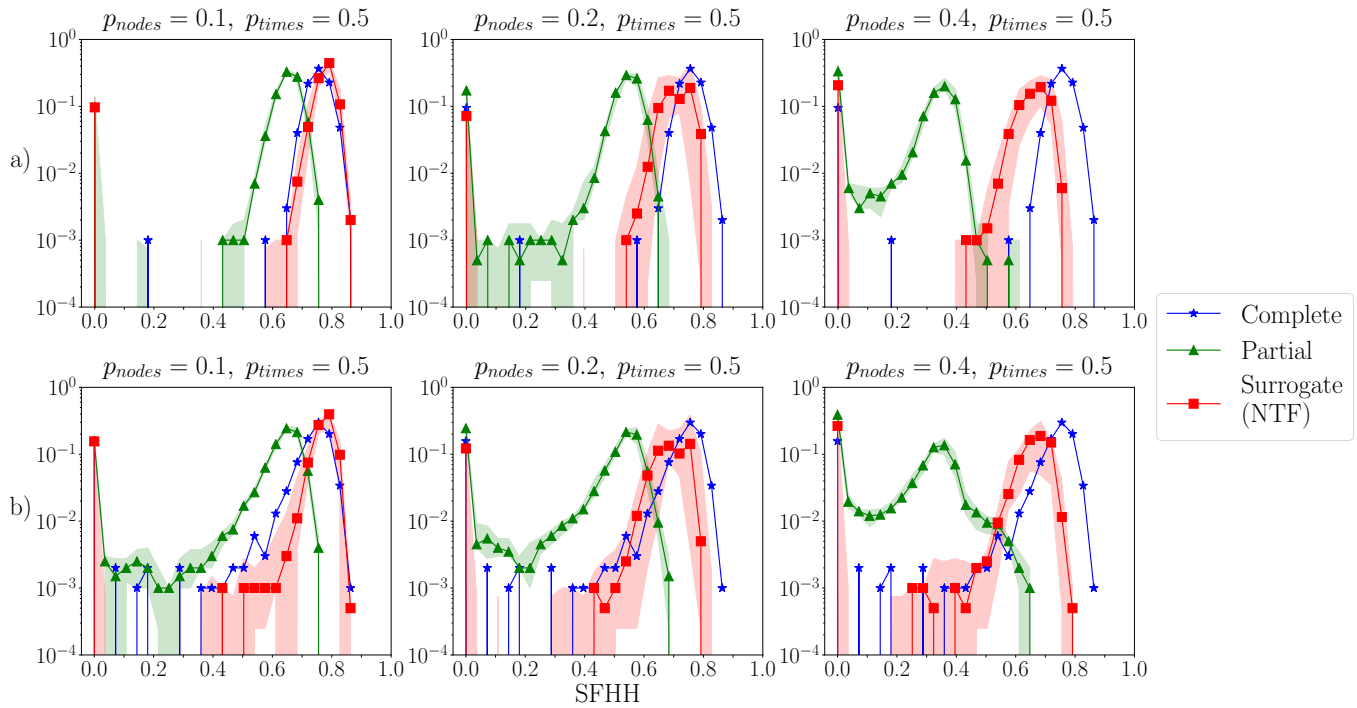

Figure 3. Distributions of epidemic sizes computed in the complete, partial, and surrogate cases for the SFHH dataset. Each panel corresponds to one value of the fraction p_{nodes} of nodes with partial information and one couple of spreading parameters: a) $\beta = 0.3$ and $\mu = 0.10$; b) $\beta = 0.25$ and $\mu = 0.08$. For the partial and surrogate cases, the symbols and lines show the median distribution of the epidemic size computed from the results relative to the 10 different sets of nodes with partial information, while the shaded area is delimited by the 25-th and 75-th percentiles.

a distribution of epidemic size closer to the one obtained on the complete network. This shows that the external information provided by the approximated location of individuals in the school helps to infer possible correlations in the contact activity, even for nodes for which no contact activity was initially known. We note however that the epidemics sizes remain underestimated with respect to the complete network.

Finally, we consider a case in which all nodes had some activity missing. We simulate a scenario in which all the nodes have lost half of their activity: for each node, we zero out its activity for half of the times taken at random among the times it was active. In that way each node has a different set of times during which its activity is erased. We apply the method based on the JNTF in such scenarios with 10 random generations of the missing times for each node. The epidemic size distributions obtained are represented on Fig. 5 together with the one measured with the complete network. While the distributions measured on the partial networks strongly differ from those obtained with the complete network, the approach based on the JNTF yields a much better estimation of the real epidemic size distribution.

VI. DISCUSSION

In this work, we have proposed a new approach to face the problem of estimating the outcome of spreading processes on temporal human proximity networks built from incomplete information. Our method leverages the existent correlations in the observed activity of the nodes to recover the contact properties of nodes whose activity is partially missing. To this aim, we rely on tensor decomposition techniques able to extract the mesoscale properties of temporal networks. In practice, the methodology we put forward follows two main steps: (i) the extraction of structures from the partial network through tensor decomposition and (ii) the construction of a surrogate network that is used in numerical simulations to estimate the outcome of spreading processes.

In the first step we use NTF handling missing values to decompose the partial networks into a sum of components, each describing a mesoscale structure. In this step, we take into account the fact that the tensor describing the temporal contact network is based on incomplete information, and the decomposition is thus performed using only the part of the tensor composed of known contacts or known non-contacts. This leads to a good approximation of the temporal activity of the nodes with partial information. In the second step, we determine which links belong to each mesoscale structure, and use

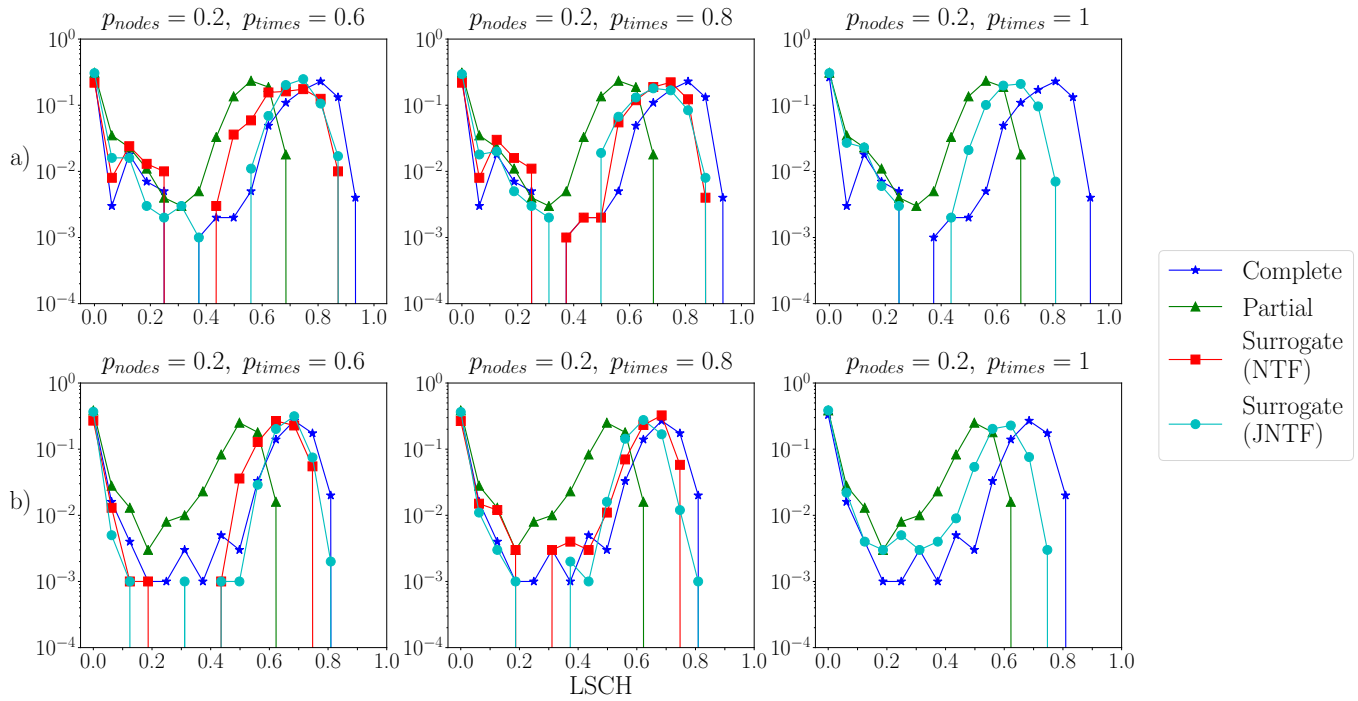

Figure 4. Results achieved by the surrogate construction methods using either the NTF or the JNTF. Each subfigure corresponds to a fixed fraction of nodes $p_{nodes} = 0.2$ with increasing loss of information for these nodes: $p_{times} = 0.6, 0.8, 1$. The JNTF is computed on the two tensors representing the temporal social network of the LSCH dataset and the related temporal location network. Here, we report the epidemic size distributions of the complete, partial and surrogate networks for SIR processes with two different couples of infection/recovery probabilities: a) $\beta = 0.3$ and $\mu = 0.3$, b) $\beta = 0.15$ and $\mu = 0.25$. In the last panels of both a) and b), as $p_{times} = 1$ no information can be gained on the nodes with missing data using simply NTF, so the results of the NTF-based method coincide with the ones obtained by simulating the SIR over the network with partial information.

the binarized activity timeline of each structure to fill in the unknown part of the timeline of each link with missing information. This is enough to obtain a surrogate network comparable to the empirical one when we use the NTF procedure. However, the JNTF might overestimate the activity of the links with missing activity and leads to a distribution of weights (number of contacts per pair of individuals) that is more homogeneous than in the original case. As the heterogeneity properties of contact networks are well known to have a crucial role in determining the outcome of spreading processes, we adjust the weight distribution of the links for which only partial information was available: this step is made possible by the robustness under sampling of the contact network weight distribution. We can thus rely on the weight distribution measured on those links for which no information is missing in the incomplete network to extract at random values and assign them to the links with missing information.

We tested our method on three different datasets describing face-to-face contacts between individuals in different contexts (two conferences and a primary school), represented as temporal human proximity networks, on which we simulated a loss of data, namely a loss of information concerning a fraction of the individuals, each for a fraction of the total timeline. We have shown that the method is able to well estimate the outcome of a spreading process, even when half of the activity is missing for 40% of the nodes, while the epidemic sizes are strongly underestimated when we simulate the process on the networks with incomplete information.

The performance of the method based on the NTF however decreases when the amount of missing information drastically increases. In particular, it is not applicable if no information at all is available for some of the nodes. To deal with this issue, we have proposed an adaptation of the method based on the joint factorization of multiple tensors (JNTF). The JNTF is indeed a natural extension that allows to integrate information encoded in multiple networks. It is particularly adequate if the information available concerns on the one hand the contacts of individuals and on the other hand their (approximate) location, encoded in a bipartite temporal network in which a link is drawn between a node representing an individual and a node representing a location when the individual is detected in that location. The joint factorization of the tensors representing the partial temporal network of contacts and the temporal network encoding positions, constrained to extract mesoscale structures with the same nodes and the same activity timelines in both tensors, can then help to recover the missing information about the contacts. We have tested this alternative method in a

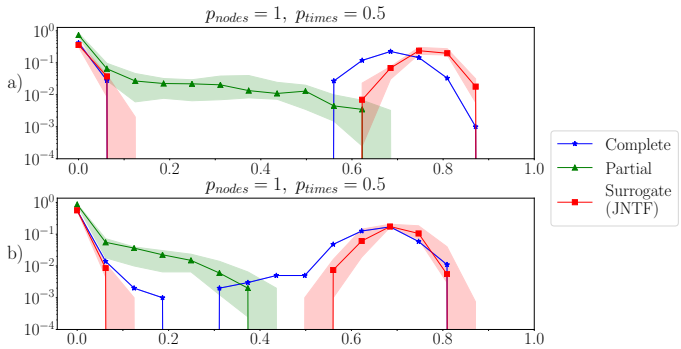

Figure 5. Results achieved by the surrogate construction method using the JNTF. Each subfigure corresponds to the case in which $p_{nodes} = 1$ and $p_{times} = 0.5$. The JNTF is computed on the two tensors representing the temporal human proximity network of the LSCH dataset and the related temporal location network. Here, we report the results for two different couples of infection/recovery probabilities: a) $\beta = 0.3$ and $\mu = 0.3$, b) $\beta = 0.15$ and $\mu = 0.25$. The symbols and lines show the median distributions of epidemic sizes computed over the 10 random generations of the missing times, while the shaded area is delimited by the 25-th and 75-th percentiles.

dataset for which both contacts and approximated positions of nodes in time are available. We have shown that it yields results similar to the NTF when both methods can be applied. When information about some nodes is completely lost, i.e., when the method based on NTF alone cannot be used, using JNTF allows us to recover part of the missing information and yields a distribution of epidemic sizes closer to the original than when using the partial network in the simulations. It is also the case when as much as 50% of the contact activity of each node has been lost in the data.

The methodology we have presented manages to cope with missing information in various contexts. Importantly, and contrarily to other methods, it does not rely on the availability of metadata or any knowledge on the structure of the population into groups (such as classes in a school) as in [20]: the NTF or JNTF decompositions are indeed able to extract the effective structure (both in groups and temporal) of the temporal network of interactions and to assign each individual to one or multiple mesoscale structures, each with its activity timeline.

Some limitations of our method stem from the tensor decomposition itself. For instance, if a whole group of correlated links or nodes is missing in the data, it will not be uncovered by the NTF. In such an unfavorable case, the JNTF method can compensate the lack of correlated activity in the incomplete contact network by relying on auxiliary data when available. Moreover, when the temporal network of contacts lacks structure or when the nodes with partial information behave in a random way, i.e., do not exhibit any activity correlation with other nodes, neither decomposition (NTF nor JNTF) might be able to clearly assign them to any mesoscale structure and thus to determine their activity timeline. A way to cope with this issue might be to attribute some random activity (based for instance on the average behaviour) to those nodes with missing information that are not found in any mesoscale structure.

Another limitation, which can be easily overcome though, is that for the JNTF we rely on the distribution of weights measured in the partial network. Indeed, if the remaining information is not representative enough (too much information missing), the heterogeneity properties of the surrogate network might be affected as the weights assigned to the links with missing information are taken among those measured. We could however take advantage of the known robustness of the weight distributions in different contexts [15, 34] to use publicly available weight distributions collected in other contexts.

Finally, a natural direction for future research is the extension of our technique to infer mesoscale properties of human proximity networks even when no direct information on contacts is available, but only proxies such as approximate locations of individuals have been collected.

VII. METHODOLOGY

Here we describe some technical details regarding the different steps of our procedure.

A. NTF Rank Selection

The selection of the number of components R for the decomposition is guided by the Core Consistency Diagnostic [50], which estimates to which extent the PARAFAC model $[\![\boldsymbol{\lambda}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$ with a given rank r (i.e., with a sum of r components) is appropriate to represent the data. The core consistency is a measure that has 100 as an upper bound and values above 50 are usually considered acceptable. Here, we computed the core consistency values between the tensor with partial information and its approximation for $r \in [2, \ldots, R_{max}]$, for 5 realizations of the optimization procedure starting from different initial conditions for $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ for each value of r. We select a rank $R = R_{cc} - 1$, where R_{cc} is the smallest rank for which the core consistency value for each of the 5 realizations is lower than 85. This threshold is selected to ensure an approximation as faithful as possible to the original tensor. For the JNTF case, we use as a hint for the number of components to be selected, the one obtained with the NTF on the temporal contact network with partial information. It is worth noting that, when the amount of available information in the data decreases, the value of R determined by the core consistency can vary. This is due to the fact that by erasing an increasing percentage of node activities the correlated activity patterns are increasingly perturbed and might be destroyed: a smaller number of components

(sub-networks composed by links having a correlated activity in time) is then detected.

B. JNTF Computation

To integrate data from multiple sources we use the Joint Non-negative Tensor Factorization (JNTF), described in Sec. VII B. To this aim, we adapt the Alternating large-scale Non-negativity-constrained Least Squares (ANLS) framework and in particular the way to compute the Karush-Kuhn-Tucker (KKT) optimality conditions in a Block Principal Pivoting (BPP) framework [46]. Here, we illustrate the adaptation details to solve Eq. (2) in the case of two data sources that are coupled in two dimensions (case study in Sec. V), i.e. $S = 2$, $\mathbf{A} = \mathbf{A}_1 = \mathbf{A}_2$, and $\mathbf{C} = \mathbf{C}_1 = \mathbf{C}_2$, such that Eq. (2) becomes:

$$\begin{aligned} \min \Bigg(& \frac{1}{2} \left\| \boldsymbol{\mathcal{T}}_1 - [\![\boldsymbol{\lambda}_1; \mathbf{A}, \mathbf{B}_1, \mathbf{C}]\!] \right\|_F^2 + \frac{\alpha_1}{2} \left\| \boldsymbol{\lambda}_1 \right\|_2^2 \\ & + \frac{1}{2} \left\| \boldsymbol{\mathcal{T}}_2 - [\![\boldsymbol{\lambda}_2; \mathbf{A}, \mathbf{B}_2, \mathbf{C}]\!] \right\|_F^2 + \frac{\alpha_2}{2} \left\| \boldsymbol{\lambda}_2 \right\|_2^2 \Bigg) \\ & \text{s.t.} \boldsymbol{\lambda}_{1,2}, \mathbf{A}, \mathbf{B}_{1,2}, \mathbf{C} \geq 0 \; . \end{aligned} \tag{3}$$

Note that we added here the regularization terms $\frac{\alpha_{1,2}}{2} \left\| \boldsymbol{\lambda}_{1,2} \right\|_F^2$, which are sparsity penalties.

To solve the problem through the BPP algorithm we rewrite the minimization problem as a multiple right hand side problem of the form:

$$\min \left\| \mathbf{V}\mathbf{X} - \mathbf{W} \right\|_F^2 \; , \tag{4}$$

and solve it for each of the factor matrices (here we include $\boldsymbol{\lambda}_{1,2}$ in the factor matrices). To this aim, we start by solving the problem for the factor matrices $\mathbf{B}_1$ and $\mathbf{B}_2$ which are respectively present in the first and third term of Eq. (3). With respect to these factor matrices we can rewrite the equation by using the 2-mode matricization [30] of $\boldsymbol{\mathcal{T}}_1$ and $\boldsymbol{\mathcal{T}}_2$, which leads to the following approximations:

$$\begin{aligned} \mathbf{T}_{1,(2)} &\approx \mathbf{B}_1 \boldsymbol{\Lambda}_1 \left(\mathbf{C} \odot \mathbf{A} \right)^T \\ \mathbf{T}_{2,(2)} &\approx \mathbf{B}_2 \boldsymbol{\Lambda}_2 \left(\mathbf{C} \odot \mathbf{A} \right)^T \quad , \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{\Lambda}_1 &= \operatorname{diag} \left(\lambda_{1,1}, \ldots, \lambda_{1,R} \right) \\ \boldsymbol{\Lambda}_2 &= \operatorname{diag} \left(\lambda_{2,1}, \ldots, \lambda_{2,R} \right) \; . \end{aligned}$$

We can rewrite the approximations as

$$\begin{aligned} \mathbf{T}_{1,(2)}^T &\approx \left(\mathbf{C} \odot \mathbf{A} \right) \boldsymbol{\Lambda}_1 \mathbf{B}_1^T \\ \mathbf{T}_{2,(2)}^T &\approx \left(\mathbf{C} \odot \mathbf{A} \right) \boldsymbol{\Lambda}_2 \mathbf{B}_2^T \quad , \end{aligned}$$

where $\boldsymbol{\Lambda}_{1,2}^T = \boldsymbol{\Lambda}_{1,2}$, and thus

$$\begin{aligned} \mathbf{B}_1^T &\approx \boldsymbol{\Lambda}_1^{-1} \left(\mathbf{C} \odot \mathbf{A} \right)^\dagger \mathbf{T}_{1,(2)}^T \\ \mathbf{B}_2^T &\approx \boldsymbol{\Lambda}_2^{-1} \left(\mathbf{C} \odot \mathbf{A} \right)^\dagger \mathbf{T}_{2,(2)}^T \quad . \end{aligned}$$

By using the following property of the Khatri-Rao product, for which

$$\left(\mathbf{C} \odot \mathbf{A} \right)^\dagger = \left(\mathbf{C}^T \mathbf{C} * \mathbf{A}^T \mathbf{A} \right)^\dagger \left(\mathbf{C} \odot \mathbf{A} \right)^T$$

we can write the approximation as

$$\begin{aligned} \boldsymbol{\Lambda}_1 \left(\mathbf{C}^T \mathbf{C} * \mathbf{A}^T \mathbf{A} \right) \mathbf{B}_1^T &\approx \left(\mathbf{C} \odot \mathbf{A} \right)^T \mathbf{T}_{1,(2)}^T \; , \\ \boldsymbol{\Lambda}_2 \left(\mathbf{C}^T \mathbf{C} * \mathbf{A}^T \mathbf{A} \right) \mathbf{B}_2^T &\approx \left(\mathbf{C} \odot \mathbf{A} \right)^T \mathbf{T}_{2,(2)}^T \quad . \end{aligned}$$

The related subproblems of Eq. (3) for $\mathbf{B}_1$ and $\mathbf{B}_2$ are then reduced to:

$$\begin{aligned} \min_{\mathbf{B}_1} \frac{1}{2} \left\| \boldsymbol{\Lambda}_1 \left(\mathbf{C}^T \mathbf{C} * \mathbf{A}^T \mathbf{A} \right) \mathbf{B}_1^T - \left(\mathbf{C} \odot \mathbf{A} \right)^T \mathbf{T}_{1,(2)}^T \right\|_F^2 \; , \\ \min_{\mathbf{B}_2} \frac{1}{2} \left\| \boldsymbol{\Lambda}_2 \left(\mathbf{C}^T \mathbf{C} * \mathbf{A}^T \mathbf{A} \right) \mathbf{B}_2^T - \left(\mathbf{C} \odot \mathbf{A} \right)^T \mathbf{T}_{2,(2)}^T \right\|_F^2 \; . \end{aligned}$$

Finally, the subproblems can be respectively written in the form of Eq. (4) by assigning

$$\begin{aligned} & \mathbf{V} = \boldsymbol{\Lambda}_1 \left(\mathbf{C}^T \mathbf{C} * \mathbf{A}^T \mathbf{A} \right) \; , \; \mathbf{X} = \mathbf{B}_1^T \\ & \text{and} \; \mathbf{W} = \boldsymbol{\Lambda}_2 \left(\mathbf{C} \odot \mathbf{A} \right)^T \mathbf{T}_{1,(2)}^T \; , \\ & \mathbf{V} = \left(\mathbf{C}^T \mathbf{C} * \mathbf{A}^T \mathbf{A} \right) \; , \; \mathbf{X} = \mathbf{B}_2^T \\ & \text{and} \; \mathbf{W} = \left(\mathbf{C} \odot \mathbf{A} \right)^T \mathbf{X}_{2,(2)}^T \; . \end{aligned}$$

The solution to the subproblems is now straightforward as illustrated in [46].

We now need to rewrite the subproblems for the factors $\mathbf{A}$ and $\mathbf{C}$ that are present in both first and third terms of Eq. (3). We show the procedure for $\mathbf{A}$ (which is analogous for $\mathbf{C}$). First, we write the minimization problem by using the 1-mode matricization of $\boldsymbol{\mathcal{T}}_1$ and $\boldsymbol{\mathcal{T}}_2$ (3-mode for $\mathbf{C}$):

$$\begin{aligned} \min_{\mathbf{A}} \Bigg(& \frac{1}{2} \left\| \left(\mathbf{C} \odot \mathbf{B}_1 \right) \boldsymbol{\Lambda}_1 \mathbf{A}^T - \mathbf{T}_{1,(1)}^T \right\|_F^2 \\ & + \frac{1}{2} \left\| \left(\mathbf{C} \odot \mathbf{B}_2 \right) \boldsymbol{\Lambda}_2 \mathbf{A}^T - \mathbf{T}_{2,(1)}^T \right\|_F^2 \Bigg) . \end{aligned}$$

By following the procedure shown above, we can write the problem in the following way:

$$\begin{aligned} \min_{\mathbf{A}} \Bigg(& \frac{1}{2} \left\| \boldsymbol{\Lambda}_1 \left(\mathbf{C}^T \mathbf{C} * \mathbf{B}_1^T \mathbf{B}_1 \right) \mathbf{A}^T - \left(\mathbf{C} \odot \mathbf{B}_1 \right)^T \mathbf{T}_{1,(1)}^T \right\|_F^2 + \\ & + \frac{1}{2} \left\| \boldsymbol{\Lambda}_2 \left(\mathbf{C}^T \mathbf{C} * \mathbf{B}_2^T \mathbf{B}_2 \right) \mathbf{A}^T - \left(\mathbf{C} \odot \mathbf{B}_2 \right)^T \mathbf{T}_{2,(1)}^T \right\|_F^2 \Bigg) , \end{aligned}$$

in which we can assign

$$\begin{aligned} & \mathbf{V}_1 = \boldsymbol{\Lambda}_1 \mathbf{C}^T \mathbf{C} + \mathbf{B}_1^T \mathbf{B}_1 \; , \mathbf{W}_1 = \left(\mathbf{C} \odot \mathbf{B}_1 \right)^T \mathbf{T}_{1,(1)}^T \; , \\ & \mathbf{V}_2 = \boldsymbol{\Lambda}_2 \mathbf{C}^T \mathbf{C} + \mathbf{B}_2^T \mathbf{B}_2 \; , \mathbf{W}_2 = \left(\mathbf{C} \odot \mathbf{B}_2 \right)^T \mathbf{T}_{2,(1)}^T \; , \\ & \text{and} \; \mathbf{X} = \mathbf{A}^T \; , \end{aligned}$$

leading to

$$f \left(\mathbf{X} \right) = \min_{\mathbf{X}} \left(\frac{1}{2} \left\| \mathbf{V}_1 \mathbf{X} - \mathbf{W}_1 \right\|_F^2 + \frac{1}{2} \left\| \mathbf{V}_2 \mathbf{X} - \mathbf{W}_2 \right\|_F^2 \right) . \tag{5}$$

To solve the problem in Eq. (5) we need to adapt the KKT conditions, which result to be

$$\nabla f(\mathbf{X}) = \underbrace{\left(\mathbf{V}_1^T\mathbf{V}_1 + \mathbf{V}_2^T\mathbf{V}_2\right)}_{\mathbf{V}_{1,2}}\mathbf{X} - \underbrace{\left(\mathbf{V}_1^T\mathbf{W}_1 + \mathbf{V}_2^T\mathbf{W}_2\right)}_{\mathbf{W}_{1,2}}$$

$$\nabla f(\mathbf{X}) \geq 0\ ,\ \nabla f(\mathbf{X})^T\mathbf{X} = 0\ ,\ \mathbf{X} \geq 0\ ,$$

whose solution is given by solving

$$\mathbf{X}^T\mathbf{V}_{1,2}^T - \mathbf{W}_{1,2}^T = 0\ .$$

Since Eq. (3) includes regularization terms, we have to adapt the KKT conditions for the minimization problem with respect to $\boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_2$. We show the procedure for $\boldsymbol{\lambda}_1$, which is analogous for $\boldsymbol{\lambda}_2$. We consider the cost function built from the terms in which $\boldsymbol{\lambda}_1$ is involved (the first and second in Eq.(3)):

$$f_{\boldsymbol{\lambda}_1} = \frac{1}{2}\left\|\boldsymbol{\mathcal{X}}_1 - [\![\boldsymbol{\lambda}_1; \mathbf{A}, \mathbf{B}_1, \mathbf{C}]\!]\right\|_F^2 + \frac{\alpha_1}{2}\left\|\boldsymbol{\lambda}_1\right\|_2^2\ ,$$

and we rewrite it through the vectorization:

$$f_{\boldsymbol{\lambda}_1} = \frac{1}{2}\left\|\mathrm{vec}\left(\mathbf{C}\odot\mathbf{B}\odot\mathbf{A}\right)\boldsymbol{\lambda}_1 - \mathrm{vec}\left(\boldsymbol{\mathcal{X}}\right)\right\|_F^2 + \frac{\alpha_1}{2}\left\|\boldsymbol{\lambda}_1\right\|_2^2\ .$$

Minimizing the cost function $f_{\boldsymbol{\lambda}_1}$ is equivalent to minimize $f'_{\boldsymbol{\lambda}_1}$, obtained by incorporating the regularization term as follows:

$$f'_{\boldsymbol{\lambda}_1} = \frac{1}{2}\left\|\underbrace{\begin{pmatrix}\mathrm{vec}\left(\mathbf{C}\odot\mathbf{B}_1\odot\mathbf{A}\right)\\ \sqrt{\alpha_1}\end{pmatrix}}_{\mathbf{V}}\underbrace{\boldsymbol{\lambda}_1}_{\mathbf{x}} - \underbrace{\begin{pmatrix}\mathrm{vec}\left(\boldsymbol{\mathcal{X}}\right)\\ 0\end{pmatrix}}_{\mathbf{w}}\right\|_F^2\ ;$$

The solution to the minimization of $f'_{\boldsymbol{\lambda}_1}$ follows from [46].

C. Otsu Method

The Otsu method [51] is commonly used in image processing to recover the different levels of grey in pixels. The same idea can be used on the temporal activities of the components by thinking of them as images composed by K pixels of different values (level of activation). In particular, the method for 2-dimensional functions assumes that the function given as an input contains values that follow a bimodal distribution. The method computes then the optimal threshold, defined by minimizing the intra-class variance and by maximizing the inter-class variance. The resulting optimal threshold can be used to divide the values of the function into two groups.

Here, we use the threshold given by the Otsu method to define whenever a component is active or not. This is done by applying the Otsu method on the temporal activity of each component. The values above the threshold correspond to the temporal activation of the component, while the values below the threshold correspond to the times in which the component is inactive.

D. SIR processes

To simulate how an infectious disease propagates in a population, and thus how the related dynamical process spreads over the network, we run an SIR process over the network. The SIR model assumes that each individual in the population can be in one of three states: Susceptible, Infectious, Recovered. The propagation of the disease starts from a single individual, who is in the Infectious state (the seed of the process) while all others are initially susceptible. At each timestep, each susceptible individual in contact with an infectious one has a probability β to become infectious. Each infectious individual recovers with probability μ per timestep. For simulation purposes, the first node to be infected is chosen among the nodes that are not in the set of those with partial information. For each network and each set of values of the parameters β and μ, we run 1000 simulations of the spreading process. In particular, the parameter space we consider a priori is given by all the couples of probabilities (β, μ), with β, μ in the range $[0.001, 1]$. We select the suitable couples of parameters according to the conditions described in Sec. V C and we show in the main text the results corresponding to two representative examples of parameter couples for each dataset. Results related to other couples of parameters are similar to the ones shown in the figures.

[1] N. D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. T. Campbell, IEEE Communications magazine **48** (2010).

[2] E. Miluzzo, N. D. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eisenman, X. Zheng, and A. T. Campbell, in *Proceedings of the 6th ACM conference on Embedded network sensor systems* (ACM, 2008) pp. 337–350.

[3] N. Eagle and A. S. Pentland, Personal and ubiquitous computing **10**, 255 (2006).

[4] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, PloS one **5**, e11596 (2010).

[5] A. Stopczynski, V. Sekara, P. Sapiezynski, A. Cuttone, M. M. Madsen, J. E. Larsen, and S. Lehmann, PloS one **9**, e95978 (2014).

[6] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, Proceedings of the National Academy of Sciences **107**, 22020 (2010).

[7] P. Holme and J. Saramäki, Physics reports **519**, 97 (2012).

[8] J. M. Read, K. T. Eames, and W. J. Edmunds, Journal of The Royal Society Interface **5**, 1001 (2008).

[9] T. Obadia, R. Silhol, L. Opatowski, L. Temime, J. Legrand, A. C. Thiébaut, J.-L. Herrmann, E. Fleury, D. Guillemot, P.-Y. Boëlle, *et al.*, PLoS Comput Biol **11**, e1004170 (2015).

[10] N. Voirin, C. Payet, A. Barrat, C. Cattuto, N. Khanafer, C. Régis, B.-a. Kim, B. Comte, J.-S. Casalegno, B. Lina, and P. Vanhems, Infection Control & Hospital Epidemiology **36**, 254 (2015).

[11] W. He, Y. Huang, K. Nahrstedt, and B. Wu, in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on* (IEEE, 2010) pp. 141–146.

[12] J. L. Schafer and J. W. Graham, Psychological methods **7**, 147 (2002).

[13] T. Smieszek, E. Burri, R. Scherzinger, and R. W. Scholz, Epidemiol. Infect. **21**, 1 (2011).

[14] V. Ouzienko and Z. Obradovic, Machine learning **95**, 329 (2014).

[15] R. Mastrandrea and A. Barrat, PLoS Comput Biol **12**, e1005002 (2016).

[16] T. Smieszek, S. Castell, A. Barrat, C. Cattuto, P. J. White, and G. Krause, BMC infectious diseases **16**, 341 (2016).

[17] A. Ghani, C. Donnelly, and G. Garnett, Statistics in medicine **17**, 2079 (1998).

[18] G. Kossinets, Social networks **28**, 247 (2006).

[19] L. Rocha, N. Masuda, and P. Holme, arXiv preprint , arXiv:1306.4626 (2017).

[20] M. Génois, C. L. Vestergaard, C. Cattuto, and A. Barrat, Nature communications **6** (2015).

[21] C. L. Vestergaard, E. Valdano, M. Génois, C. Poletto, V. Colizza, and A. Barrat, European Journal of Applied Mathematics **27**, 941 (2016).

[22] M. Huisman, Journal of Social Structure **10**, 1 (2009).

[23] R. Guimerà and M. Sales-Pardo, Proceedings of the National Academy of Sciences **106**, 22073 (2009).

[24] A. Clauset, C. Moore, and M. E. Newman, Nature **453**, 98 (2008).

[25] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, Journal of Machine Learning Research **9**, 1981 (2008).

[26] M. Kim and J. Leskovec, in *SDM*, Vol. 11 (SIAM, 2011) pp. 47–58.

[27] E. Sadikov, M. Medina, J. Leskovec, and H. Garcia-Molina, in *Proceedings of the fourth ACM international conference on Web search and data mining* (ACM, 2011) pp. 55–64.

[28] J. Fournet and A. Barrat, Sci. Rep. (2017).

[29] S. Ghonge and D. C. Vural, Phys. Rev. E **96**, 012319 (2017).

[30] T. G. Kolda and B. W. Bader, SIAM review **51**, 455 (2009).

[31] L. Gauvin, A. Panisson, and C. Cattuto, PLoS ONE **9** (2014); R. Bro and S. De Jong, Journal of chemometrics **11**, 393 (1997).

[32] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, in *SDM* (SIAM, 2010) pp. 701–712.

[33] J.-P. Royer, N. Thirion-Moreau, and P. Comon, in *20th European Signal Processing Conference (EUSIPCO-2012)* (Elsevier, 2012) pp. 1–5.

[34] A. Barrat and C. Cattuto, "Face-to-face interactions," in *Social Phenomena*, edited by B. Gonçalves and N. Perra (Springer International Publishing Switzerland, 2015) Chap. 3, pp. 37–57.

[35] K. Hayashi, T. Takenouchi, T. Shibata, Y. Kamiya, D. Kato, K. Kunieda, K. Yamada, and K. Ikeda, in *Data Mining (ICDM), 2010 IEEE 10th International Conference on* (IEEE, 2010) pp. 216–225.

[36] A. Vazquez, B. Racz, A. Lukacs, and A.-L. Barabasi, Physical review letters **98**, 158702 (2007).

[37] J. L. Iribarren and E. Moro, Physical review letters **103**, 038702 (2009).

[38] T. Smieszek, Theoretical Biology and Medical Modelling **6**, 25 (2009).

[39] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, and P. Vanhems, BMC medicine **9**, 87 (2011).

[40] A. Machens, F. Gesualdo, C. Rizzo, A. E. Tozzi, A. Barrat, and C. Cattuto, BMC infectious diseases **13**, 1 (2013).

[41] L. Gauvin, A. Panisson, C. Cattuto, and A. Barrat, Sci. Rep. **3**, 3099 (2013).

[42] C. L. Vestergaard, M. Génois, and A. Barrat, Physical Review E **90**, 042805 (2014).

[43] H. Kim, H. Park, and L. Eldén, *Bioinformatics and Bioengineering, 2007. BIBE 2007. Proceedings of the 7th IEEE International Conference on*, , 1147 (2007).

[44] K. Balasubramanian, J. Kim, A. Puretskiy, M. W. Berry, and H. Park, Text Mining.

[45] E. Acar, T. G. Kolda, and D. M. Dunlavy, arXiv preprint arXiv:1105.3422 (2011).

[46] J. Kim and H. Park, "Fast nonnegative tensor factorization with an active-set-like method," in *High-Performance Scientific Computing: Algorithms and Applications*, edited by M. W. Berry, K. A. Gallivan, E. Gallopoulos, A. Grama, B. Philippe, Y. Saad, and F. Saied (Springer London, London, 2012) pp. 311–326.

[47] M. Fang, G. Yue, and Q. Yu, in *International Symposium on Information Processing (ISIP)* (Citeseer, 2009) pp. 109–112.

[48] M. Szomszor, C. Cattuto, W. Van den Broeck, A. Barrat, and H. Alani, in *Extended Semantic Web Conference* (Springer, 2010) pp. 196–210.

[49] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina, *et al.*, PloS one **6**, e23176 (2011).

[50] R. Bro and H. A. Kiers, Journal of chemometrics **17**, 274 (2003), definition of the core consistency diagnostic.

[51] M. Fang, G. Yue, and Q. Yu, in *International Symposium on Information Processing (ISIP)* (Citeseer, 2009) pp. 109–112.

SUPPLEMENTARY INFORMATION

Approximated network

By decomposing a network affected by missing information through the Non-negative Tensor Factorization (NTF) we are able to reconstruct some of the structural and temporal properties observed in the complete network. One of these characteristics is the overall contact activity of the nodes. In particular, given a node whose information is partially missing, we are able to reconstruct its overall activity in time, i.e., the number of contacts related to that node at each time. We have indeed

shown that the activities in the complete network and in the approximated one are significantly correlated and have high Pearson correlation coefficients. In Fig. 6, we report a representative example for the overall temporal activity of a single node in the complete, partial, and approximated network for each dataset: LSCH, HT09, and SFHH. As we consider a scenario in which the missing activity concerns, in each case, the first half of the timeline, the activity of the nodes with missing information is 0 in the partial data, while it shows non-trivial patterns in the complete network, which are partially recovered in the approximated network.

SIR on the approximated network

We have shown that by approximating a network with partial information via NTF we are able to recover some of its properties, such as the overall temporal activity of the nodes. However, as discussed in the main text, this information is not enough to obtain outcomes of a spreading process close to the ones obtained on the complete network, because the approximation tends to make the distributions of weights more homogeneous than the empirical one. We illustrate this fact by showing in Fig. 7 the distributions of epidemic sizes of an SIR process simulated on the complete, partial and approximated networks for the LSCH dataset, for different spreading parameters ($\beta = 0.30$ and $\mu = 0.30$ and $\beta = 0.15$ and $\mu = 0.25$), different fractions of nodes missing $(0.1, 0.2, 0.4)$ and fraction of times with missing data $p_{times} = 0.5$.

Here, we reported the results only for the LSCH dataset, as it is the one characterized by highly correlated activity patterns and thus the one for which the approximations obtained through the NTF are closer to the complete case. In the primary school indeed, students' activity is determined by the school schedule and they are divided in classes. This makes their activity highly correlated as their contacts are more homogeneous during the daily class activities. However, as we can see from Fig. 7, even if the approximated network is close to the complete one in terms of the overall activity patterns displayed, this is not the case for the outcome of the spreading process. The epidemic sizes in the case of the partial and approximated networks are far from the complete network case. This strong underestimation is due by the fact that when we are approximating the network via NTF, the method tends to approximate the activity patterns of the nodes in the same component as fully correlated, thus connecting the nodes in the same component and rendering the network more homogeneous.

Weight Distribution

For the JNTF case, in order to obtain outcomes of SIR processes close to the case of the complete network, we have to reintroduce the heterogeneity properties into the approximated network, and use the resulting surrogate network to simulate the SIR process. We reintroduce the heterogeneity properties by reassigning the weights, i.e., the total number of contacts, to the links involving nodes with partial information. In particular, we measure the weight distribution on the available part of the partial network. Then, we use this distribution to pick at random a weight that will be reassigned to the link whose activity was approximated via the JNTF. We rely on such a process as the weight distribution of a network is robust to various sampling procedures, as shown in Fig. 8 for the sampling considered in this paper: for each dataset, we compare the weight distribution of the complete network with the corresponding weight distributions computed on the links of partial networks not involving any nodes with missing information, with $p_{nodes} \in [0.1, 0.2, 0.4]$. By construction, these distributions do not depend on p_{times}. The results clearly show that the weight distributions measured in the partial and complete networks are consistent, meaning that, even in the case in which we miss almost half of the links in the partial network, we are able to compute a weight distribution similar to the one of the complete network.

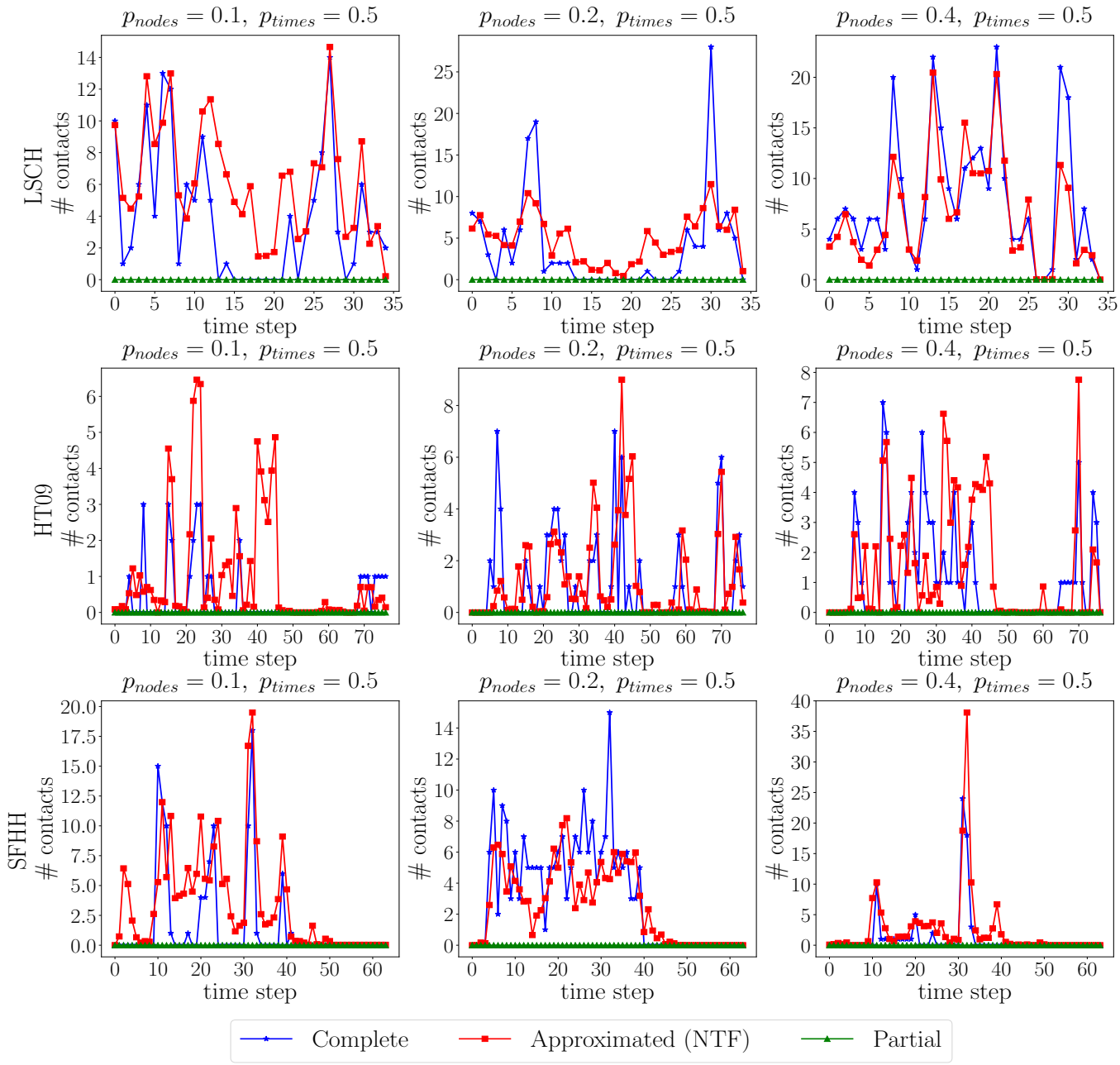

Figure 6. **Temporal activity**: example of the total number of contacts in time of a node for each dataset and case study. We report the temporal activity in the complete network, in the one with partial information, and in the one obtained approximated by the NTF method. We show here the temporal activity only for the time steps in which we lost the information about the contacts of the node, corresponding in our scenario to the first half of the timeline.

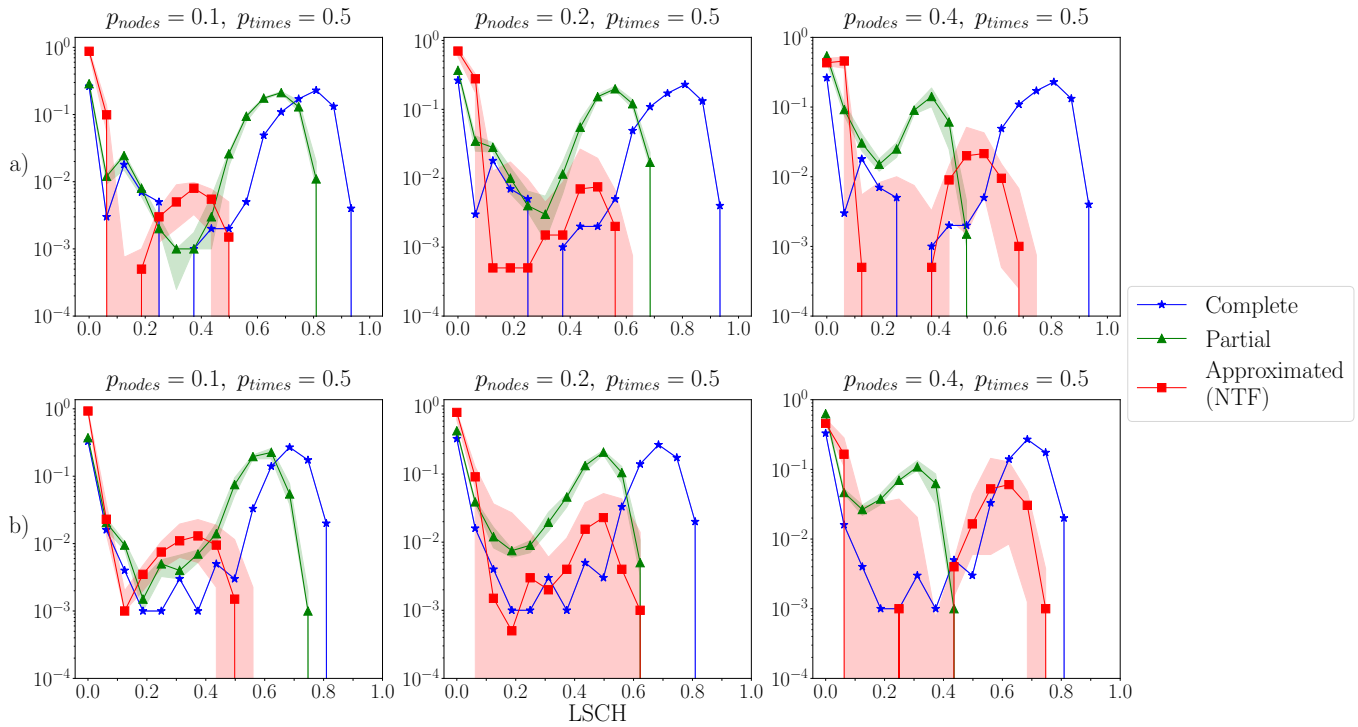

Figure 7. **Outcome of an SIR process**: over the complete network, the partial network and the network approximated through the NTF for the LSCH dataset. Each panel corresponds to one fraction p_{nodes} of nodes with partial information and one couple of spreading parameters: a) $\beta = 0.30$ and $\mu = 0.30$; b) $\beta = 0.15$ and $\mu = 0.25$. The lines show the median distribution of the epidemic size computed from the results relative to 10 different sets of nodes with missing information, while the shaded area is delimited by the 25-th and 75-th percentiles.

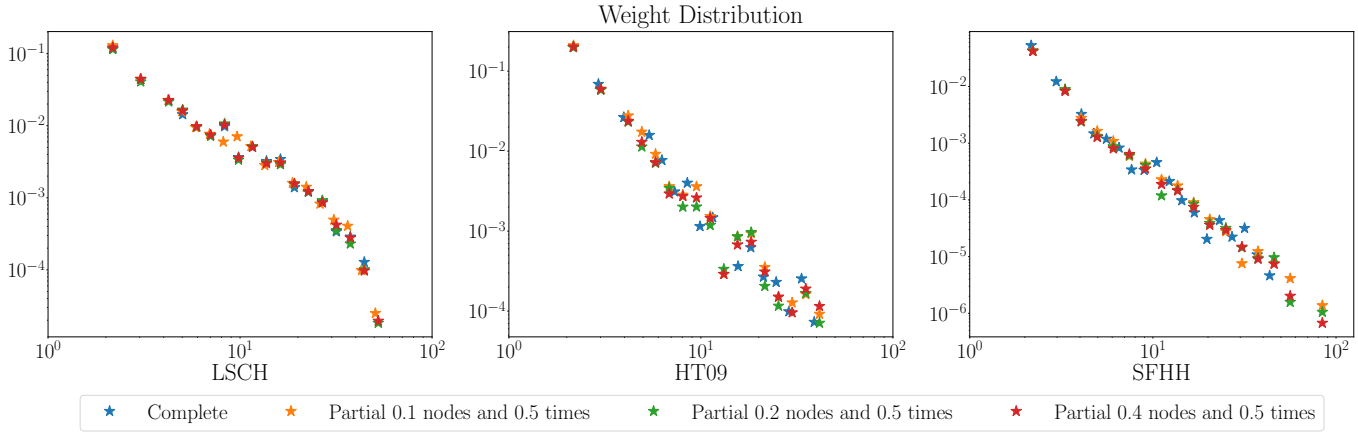

Figure 8. **Weight distributions** in the three datasets for the complete network and the partial network with increasing percentage of nodes with missing information: $p_{nodes} \in [0.1, 0.2, 0.4]$ and $p_{times} = 0.5$.