Chapter 3
Face-to-Face Interactions

Alain Barrat and Ciro Cattuto

Abstract Face-to-face interactions of humans play a crucial role in their social relationships as well as in the potential transmission of infectious diseases. Here we discuss recent research efforts and advances concerning the measure, analysis and modelling of such interactions measured using strategies ranging from surveys to decentralised infrastructures based on wearable sensors. We present a number of empirical characteristics of face-to-face interaction patterns and novel techniques aimed at uncovering mesoscopic structures in these patterns. We also mention recent modelling efforts and conclude with some open questions and challenges.

3.1 Introduction

Our modern interconnected societies make many channels available for communications and social interactions, such as phone calls, email, virtual conferences, micromessaging, or online social networks. Despite this wealth of alternatives, direct face-to-face interactions between individuals remain an essential element of human behaviour and of human societies. Mining and analysing face-to-face interaction patterns between individuals therefore has a clear impact towards the fundamental knowledge and understanding of human behaviour and social networks. Most crucially, contact patterns among individuals play an important role in determining the potential transmission routes of infectious diseases, in particular of respiratory pathogens. An accurate description of these patterns represents therefore a crucial tool for identifying contagion pathways, for informing models

A. Barrat (✉)
Aix Marseille Université, Université de Toulon, CNRS, CPT, UMR 7332,
13288 Marseille, France

Data Science Laboratory, ISI Foundation, Torino, Italy
e-mail: alain.barrat@cpt.univ-mrs.fr

C. Cattuto
Data Science Laboratory, ISI Foundation, Torino, Italy
e-mail: ciro.cattuto@isi.it

B. Gonçalves, N. Perra (eds.), *Social Phenomena*, Computational
Social Sciences, DOI 10.1007/978-3-319-14011-7_3

of epidemic spread, and for the design and evaluation of control measures such as the targeting of specific groups of individuals with appropriate prevention strategies or interventions.

Empirical data describing direct interactions between individuals are however by nature difficult to gather as, contrarily to online interactions, phone calls or electronic communications, they do not leave any digital trace. Various data collection strategies have therefore been used, in particular in the epidemiological context and at different scales: surveys and diaries, synthetic population models, and, thanks to the increase in the availability and use of novel technologies, wearable sensors (see [1] for a review).

Here we first briefly review the measurement strategies and some of their advantages and intrinsic limitations (Sect. 3.2). We then discuss in Sects. 3.3 and 3.4 a number of empirical characteristics of face-to-face interactions as obtained by recent projects using wearable sensors, and review in Sect. 3.5 some recent attempts at modelling these processes. We conclude in Sect. 3.6 by presenting a number of open questions.

3.2 Proxies of Face-to-Face Interactions and Measurement Strategies

Face-to-face interactions between individuals occur in a variety of contexts and situations, contributing to phenomena as diverse as social coordination, information propagation, disease spread and more. Gathering data and understanding the patterns of these direct contacts is therefore of interest to fields of research ranging from the fundamental understanding of human behaviour to the epidemiology of transmissible diseases, and many efforts have been devoted to these tasks. We refer the reader to [1] for a recent review of the methods and technologies that have been used in various projects and provide here only a brief discussion on the range of methods available and non-exhaustive references to the corresponding research efforts.

A commonly employed method consists in asking individuals about their contacts, using surveys and diaries. Volunteer participants are asked to record their social interactions during a certain time period, for instance on a specific day, or on consecutive days. While social science studies can be interested in all such interactions (face-to-face, by email or by phone), the strongest focus on obtaining information on face-to-face interactions has emerged in epidemiological studies of infectious diseases, as such direct interactions are considered as relevant for transmission events. Many efforts have therefore been deployed to use contact diaries under various forms (using both paper and web-based questionnaires), either in specific contexts ranging from hospitals to schools or among the general population [2–9], and sometimes at very large scale with thousands of respondents [5, 7, 8]. Surveys have both advantages and limitations. One of the main advantages is that well-studied questionnaires allow to gather information not only on the existence of contacts but also on additional characteristics, such as their context (home, work, travel), estimates of their durations, existence of repeated contacts with the same individual, or even the distance from home at which the contacts take place [8]. Metadata such as the age, gender and occupation of the respondent can also be correlated with his/her contact numbers and durations. Questionnaires can even ask to specify for each contact if it involved physical contact and distinguish periods of well-being and illness of the respondent [10]. Surveys have also important limitations. First, questionnaires are costly and it is notoriously difficult to recruit participants [8, 9]. Second, self-reporting procedures entail biases that are difficult to estimate [4, 11, 12], as participants might not recall all their contacts or might make incorrect estimates of their durations. As surveys give access to ego-networks, the fraction of triangles in contact networks is also difficult to estimate and typically relies on each individual estimating if two of his/her contacts have themselves been in contact [8]. Finally, subtleties in questionnaire design might also influence the results, as discussed in [8]: for instance, the distribution of the number of reported contacts varies significantly whether individuals have to report the name of each contact or not.

Alternative approaches to the use of surveys have emerged in the recent years, giving usually access to proxies of face-to-face interactions. For instance, the availability of large-scale computing facilities and of detailed socio-demographic data have made it possible to recreate in silica synthetic populations at the scale of a whole city or country. These synthetic populations are typically used to generate contact networks to simulate the spread of infectious diseases [13–15]. Interestingly, the contact patterns obtained within such synthetic populations have been shown to match those obtained in large-scale surveys [14, 15].

Another approach takes advantage of the development of various types of sensors which can in particular measure the proximity of other similar devices, using technologies ranging from Bluetooth, WiFi or RFID [16–24]. Depending on the range of the signals used, such methods might yield information only on proximity at a range that might not imply face-to-face interaction (e.g., Bluetooth signals between devices can typically be received through a wall), or can be tuned to specifically detect close-range face-to-face proximity [19, 21, 22]. We will here mostly report on results obtained with the latter technology. Wearable sensors are nowadays simple to use and come at reasonable costs. They also afford an objective definition of contact and can report even short encounters. Their main limitation comes from the fact that they do not register contacts with individuals not participating to the data collection (and therefore not wearing any sensor) and therefore provide data on the contacts among a closed population. Sampling issues can thus arise if not all the members of the population of interest agree to wear the sensors [22].

Given the respective advantages and limitations of methods based on surveys and wearable sensors, a comparison of data collected by both types of methods in a given population is of great interest. To our knowledge, only one such study has been performed to date, namely in a high school context, showing in particular that many contacts registered by sensors are not reported in surveys, especially for short contacts, while long contacts are better reported [12]. More such studies in various contexts would be highly welcome in the future.

3.3 Face-to-Face Interactions as Temporal Networks

One of the advantages of decentralised sensing infrastructures based on wearable sensors is that they not only yield information on the existence of a face-to-face interaction between two individuals but also give access to the starting and ending time of each such interaction, with a certain time resolution (typically of the order of 20 s to a couple of minutes). The collected data can therefore adequately be represented as a time-varying social network of contacts within the monitored community, i.e., an instance of a "temporal network" [25].

The amount of activity, quantified as the number of observed face-to-face contacts in a given time-window, varies substantially over time and can be very different in different contexts. For instance, children in a primary school interact much more than adults in offices. Despite these differences, some generic statistical properties of the temporal networks of human interactions have emerged through the various data collection efforts. First, the time intervals between successive contacts are broadly distributed, spanning several orders of magnitude: most intercontact durations are short, but very long durations are also observed, and no characteristic timescale emerges [16, 19, 22, 26–28]. This bursty behaviour is a well-known feature of human dynamics and has been observed in a variety of systems driven by human actions [29]. Moreover, the distributions of the durations of single contacts are also broad, spanning several orders of magnitude, and their functional form displays a remarkable robustness across contexts [22, 28], measurement periods, and measurement methods [30], as illustrated in Fig. 3.1.

Overall, temporal networks of face-to-face contacts between individuals exhibit strongly heterogeneous dynamics, with robust statistical features. This implies two important facts for modelers, in particular when dealing with processes depending on contact durations between individuals, such as epidemic spreading. First, the broadness of the distributions means that taking into account only average contact durations and assuming that all contacts are equivalent might be a too coarse representation of the reality. Indeed, different contacts might yield very different transmission probabilities: many contacts are very short and correspond to a small transmission probability, but some are much longer than others and could therefore play a crucial role in disease dynamics, Second, the robustness of the distributions found in different contexts means that these distributions can be assumed to depend negligibly on the specifics of the situation being modeled and thus directly plugged into the models.

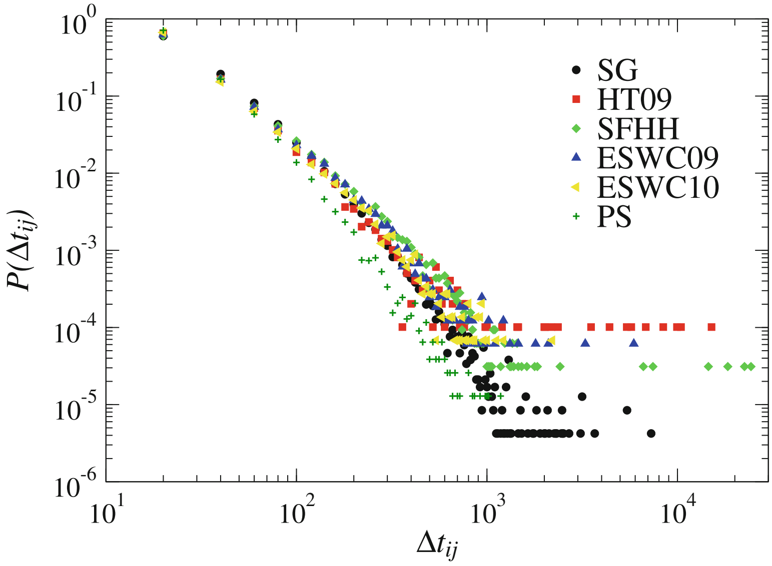

Fig. 3.1 Distributions of the face-to-face contact durations measured in different environments ranging from a museum (SG) to a school (PS) and several scientific conferences

3.4 Structures and Structure Discovery

3.4.1 Structures in Aggregated Data

3.4.1.1 Contact Networks

It is often useful to aggregate the temporal network of contacts between individuals over a given time window, in order to obtain static summaries of the contact sequence. In the obtained aggregated network, each node represents an individual, and a link between two nodes i and j denotes the fact that the corresponding individuals have been in contact at least once during the time window under consideration. Each such link is weighted by a summary of the temporal contact activity that took place between i and j, such as the number of contact events or the cumulative duration w of the contact events between the corresponding individuals.

The time window considered for aggregation can range from the finest time resolution of the recording system (that can be of the order of seconds or minutes) up to the entire duration of the data collection (e.g., days or weeks). In the case of surveys, the detailed temporal information on the contacts' timing is often not available, and the natural aggregation time scale is 1 day. Surveys thus typically give access to daily aggregated networks. Overall, different aggregation levels typically provide complementary views of the network dynamics at different scales.

The obtained aggregated networks unveil important information about the contact patterns of the population under study. A first characterisation is provided by the statistical distributions of nodes' degrees: in a contact network, the degree of a node (individual) is given by the number of distinct individuals with whom that individual has been in contact. In datasets collected through wearable sensors, the observed degree distributions are typically narrow, with an exponential decay at large degrees and characteristic average values that depend on the particular context [30, 31]. Interestingly, contact data obtained through surveys can lead either to narrow or broad degree distributions, as discussed in [8], and the result might be influenced by the way in which the survey is designed. When individuals are asked to report a precise list of persons encountered during the day, the obtained degree distributions are typically narrow [4, 6, 7], and the data of [8] actually show a good agreement with sensor data. However, if the respondents have in addition the possibility to report encounters with "groups" of individuals without specifying the identity of each group member, the distribution becomes broad [8].

While the number of distinct individuals met is certainly important when discussing behavioural patterns of humans, the durations and cumulated durations of face-to-face contacts also carry crucial information, in particular with respect to social or epidemiological contexts. The distributions of links' weights is thus a very relevant characteristic of these networks. Such distributions have been found to be broad in many different datasets collected either through sensors [19, 22, 24, 28] or surveys [8]: most pairs of interacting individuals have been in face-to-face proximity for a short total amount of time, but some cumulated contact durations are very long. No characteristic interaction timescale can be naturally defined, except for obvious temporal cutoffs due to the finite duration of the measurements. Strikingly, and similarly to the case of the durations of single contacts, recent studies have shown a strong robustness of the functional shape of these distributions in different contexts and even different data collection methods [22, 28, 30]. The empirically found distributions seem therefore to be a robust property of human behaviour and can be used directly for modelling purposes in various contexts.

While statistical distributions of node and link features display a strong robustness, the detailed structures of aggregated networks of contacts are much more diverse depending on the context. For instance, aggregated networks of interactions during a typical day at a small conference are rather "compact" with a close-knit structure [31], as participants are typically engaged in interacting with known individuals as well as in meeting new persons. Networks of contacts among children in a primary school or students in a high school display on the contrary a strong community structure, shown in Fig. 3.2, as a consequence of the grouping of individuals in classes [30, 33]. A similar structure has been observed in an office building, where workers from the same department have more contacts than with workers from other departments, even during lunch hours [34]. In hospitals, different structures emerge due to the different roles of individuals: as shown in [35, 36], nurses tend to form a rather dense group of nodes in the aggregated network. The network of links involving patients and caregivers has, on the other hand, a particular structure linking each patient to a specific caregiver, with very few links among caregivers or among patients (see [35, 37] for graphical illustrations).

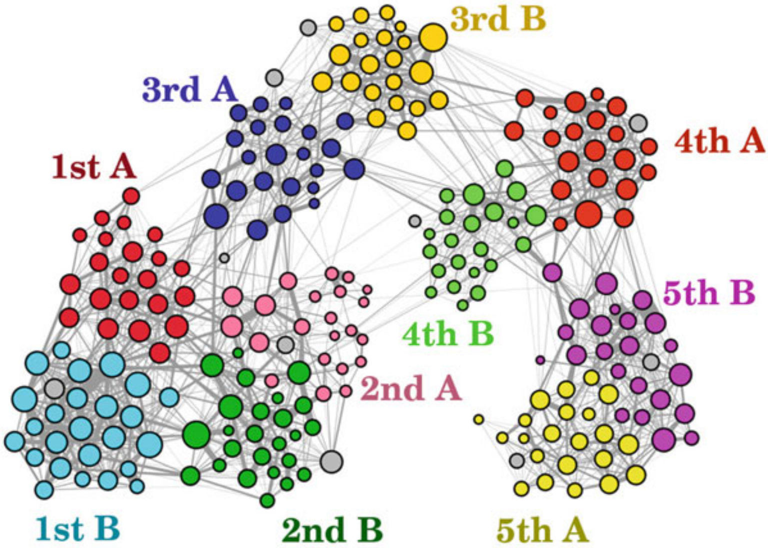

Fig. 3.2 Primary school contact network, aggregated over 1 day. Only links that correspond to cumulated face-to-face proximity in excess of 5 min are shown. The *color* of nodes indicates the grade and class of students. *Grey* nodes represent teachers. The network layout was generated by using the force atlas graph layout implementation available in Gephi [32]

3.4.1.2 Contact Matrices

It is often convenient to go one step further in the aggregation of contact data when the population under study is structured, i.e., when individuals can be classified according to specific characteristics or role (e.g., according to their age class or professional activity). A convenient summary of their contact patterns is then provided by contact matrices whose elements give the average number (or duration) of the contacts that individuals in one given class have with individuals of another class. Such a representation can be used at different scales: to describe, e.g., the contact patterns between individuals having different roles in a hospital ward (e.g., nurses, doctors, patients) [35, 36], or between children or students of different classes in schools [30, 33], but also to account for the mixing patterns between individuals of different age classes in the population of a country, as obtained by surveys [5].

Of note, the use of contact matrices for modeling contact patterns relies on a set of restricted homogeneous mixing assumptions within each class and on the representativeness of the average mixing behaviour between classes. Such an approach neglects the strong fluctuations observed in the distributions of the numbers and durations of contacts between two individuals of given classes [5, 6, 28, 38]. It neglects also the fact that contact networks are typically sparse, and that the density of links connecting individuals in given classes depends on the specific classes and is sometimes very small: many pairs of individuals never have any contact.

In order to provide a data representation that is not as specific as a high-resolution temporal network of contacts but does not discard relevant heterogeneities in the contact patterns, the contact matrix of distributions (CMD) has been introduced in [38]. This representation, instead of considering only the average of the contact time between individuals of specific classes, considers the whole distribution of contact times, typically fitted by a negative binomial distribution. Similarly to the customary contact matrices, the CMD is not an individual-based representation, and does not retain the detailed structure of the empirical contact network. It thus keeps the simplicity of a contact matrix formulation by grouping the individuals into role classes, but takes into account the heterogeneity of contact durations between individuals and the sparseness of the contact network. Such a representation is useful for designing interventions as it can suggest easily generalizable strategies that target specific classes of individuals [38].

3.4.1.3 Different Types of Contacts

Let us finally note that we have here mostly discussed aggregated networks of contacts as registered by wearable sensors in different contexts. Contacts are then gathered only between individuals participating to the data collection, and within the considered environment. Individuals however have contacts in different situations, ranging from home to workplaces and transportation means. In this respect, surveys can help understand and quantify how contacts depend on context. For instance, the large-scale survey analysed in [8] highlights how contact time decreases with age and how contacts involving touch tend be of longer duration. It also shows that home contacts account for the majority of contact hours, while work corresponds to more numerous but shorter contacts. Different occupations correspond also to different average daily numbers of contacts. Finally, the survey answers show that the time in contact decreases when the distance from home increases [8].

3.4.2 Longitudinal Structures

Human activity and contact patterns are highly non-stationary. In particular, the number of contacts among a given population varies strongly in time, obeying typical circadian rhythms and possibly modulated by the unfolding of scheduled activities [22]. It is therefore important to assess how statistical properties of contacts are impacted by and possibly coupled with these activity variations. Moreover, high-resolution datasets on contacts between individuals are typically gathered during few days or weeks in a certain context, and assessing the long-term stability of the data characteristics across different periods is also crucial.

3.4.2.1 Short-Term Stability

Despite the strong variations in activity, i.e., in the numbers of registered contacts, the main statistical properties of the contacts have been empirically shown to be stable [22, 28]. In particular, the contact duration distributions measured over different time windows coincide, as well as the structure of contact matrices across different workdays [30, 34, 36]. In fact, even the activity timelines can be remarkably stable across days when they depend on schedules either externally imposed as in schools [30] or driven by the organisation of work as in hospitals [36], as illustrated in Fig. 3.3.

On the other hand, surveys have shown that important differences between contact matrices describing contact patterns in the population are observed between work and non-work days [6, 8, 39], as well as, for a given individual, between periods of well-being and periods of illness [10].

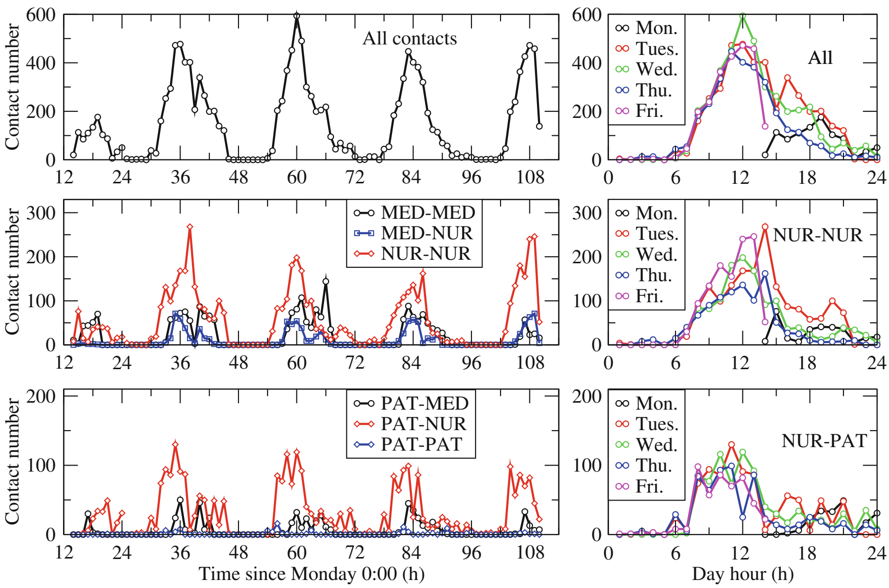

Fig. 3.3 Number of contacts per 1-h periods in a hospital ward. *Top row*: global number of contacts. *Middle and bottom rows*: number of contacts involving patients, healthcare workers and medical doctors. The *left* plots give the number of contacts as a function of the time since the start of the week (Monday, 0:00 AM). The *right* plots display the number of contacts of several types in each day, as a function of the hour of the day, to show the similarity of the curves in different days. Abbreviations: *NUR* paramedical staff (nurses and nurses' aides), *PAT* patient, *MED* medical doctor. From [36]

3.4.2.2 Long-Term Stability

Few datasets afford a comparison between contact patterns observed in a given context or in similar contexts during different periods. In particular, the comparison of the data gathered in different hospital wards [23, 35, 36] shows the robustness of stylised facts such as the central role of nurses and the small number of contacts between patients. Very few studies report and compare high-resolution contact networks measured in the same context at different points in time. Fournet and Barrat [30] compares contact data gathered in the same high school in two different years, and reports a very strong qualitative and quantitative similarity between contact matrices for different years.

3.4.3 Mesoscopic Structures and Latent Factor Analysis

In the previous sections we discussed the short-term and long-term stability of some statistical distributions of interest. The aggregation over time or over node attributes that is required to compute such distributions projects away many specificities, structures, and correlations of the original data. Depending on the problem at hand, these aggregated representations may overlook or confound important structural features of the network.

For example, a node or group of nodes may belong to different communities at different points in time: aggregating the network over time will artificially merge the communities and create a cluster that does not represent the network at any point in time. Similarly, groups of nodes may exist that share similar activity patterns over time. This is a common occurrence in environments such as schools, where an externally imposed schedule of social activities (e.g., class and lunch breaks) drives and constrains the interactions that are possible at a given time. In this case, temporal aggregation of the network may retain the topology of interactions but loses the information on correlated activity patterns, which may play an important role for, e.g., epidemic processes unfolding over the temporal network [40]. In general, correlated topological and temporal features of the network may give rise to structures that are neither local features of individual nodes or edges nor global structures, such as, for instance, a suitably defined network backbone. Hence, in the following we will refer to these structures as "mesoscopic structures". It is important to remark that meso-scale structures are not limited to the (possibly hierarchical) community structure of the network: communities are usually defined as cohesive clusters, whereas the structures under study may also comprise two-mode communities [41], groups of links with correlated activity patterns, and more.

Detecting mesoscopic structures in high-resolution social network data is an outstanding challenge that calls for principled approaches and efficient computational techniques. Recent work focuses on extending well-known community detection techniques to the case of temporal networks. A common approach is to detect communities in static networks snapshots obtained by aggregating the temporal network over consecutive time intervals. The changes of the community structure over time are then analysed to relate communities found at different times and track their evolution. Simple approaches to mine the temporal community structure of a system are based on a continuity assumption for the (static) community structure detected at successive time intervals [42–44]. These approaches may prove useful in specific cases, but fail in the presence of discontinuous activity patterns, abrupt community formation or dissolution, and in general they cannot deal with temporal correlations over extended periods of time. Instead of separately treating the community structure and the temporal evolution of the network, some studies [45–47] pioneered global approaches to the problem of community detection in temporal networks.

More recently, we have investigated the use of techniques for latent factor analysis to simultaneously identify mesoscopic network structures and track their activity over time, without the assumption that the sought structures should be cohesive clusters. The starting point for this analysis is a mathematical representation of time-varying network data that treats topology and time on an equal footing: A temporal network can be naturally represented as a time-ordered sequence of adjacency matrices, each describing the state of the network at a discrete point in time. The adjacency matrices can be combined into a three-way tensor $\mathscr{T} \in \mathbb{R}^{N \times N \times S}$, where N is the number of nodes of the network and S the number of network snapshots. The tensor $\mathscr{T}$ encodes the entire information about the temporal network and has been recognized as a convenient representation both for multi-layer networks and temporal networks [48].

Once the network and its evolution are represented in a tensor form, we can use a variety of methods from data mining and machine learning to identify latent structures. We focused on tensor decomposition techniques that were developed in diverse domains like signal processing, psychometrics and brain science [49, 50]. In particular, we investigated the use of non-negative tensor factorization [50, 51] because, like non-negative matrix factorization [52], it is recognized as a powerful tool for learning parts-based representations. The basic idea is to approximate the tensor $\mathscr{T}$ by a sum of products of lower-dimensional factors, each of which can be interpreted in terms of groups of nodes and temporal activity patterns. Formally, $\mathscr{T}$ can be approximated by a sum $\tilde{\mathscr{T}}$ of rank-1 tensors:

$$\tilde{\mathscr{T}} = \sum_{r=1}^{R} \mathbf{a_r} \circ \mathbf{b_r} \circ \mathbf{c_r} \ , \tag{3.1}$$

subjected to non-negativity constraints on $\mathbf{a_r}$, $\mathbf{b_r}$ and $\mathbf{c_r}$. The number R of terms in the decomposition controls the complexity of the model: for small values of R, $\tilde{\mathscr{T}}$ is a crude approximation of $\mathscr{T}$, whereas for high values of R the decomposition yields a good approximation but eventually overfits $\mathscr{T}$. The choice of R is usually set by means of heuristics or quality metrics for the decomposition [53]. The vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_R$, $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_R$ and $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_R$ can be arranged into matrices $\mathbf{A} \in \mathbb{R}^{N \times R}$, $\mathbf{B} \in \mathbb{R}^{N \times R}$ and $\mathbf{C} \in \mathbb{R}^{S \times R}$. Rows correspond to the nodes of the network,

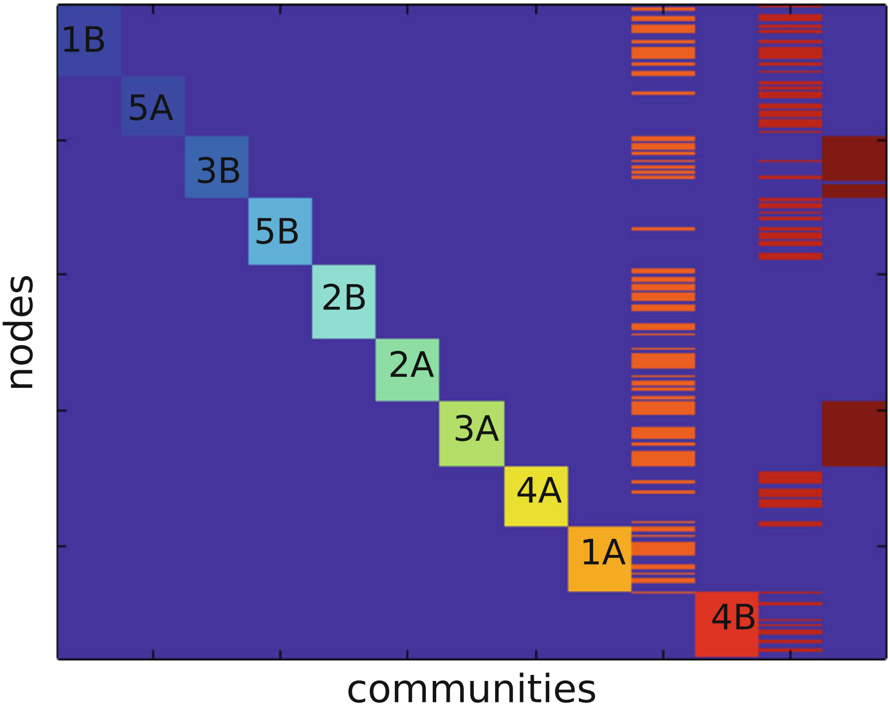

Fig. 3.4 Component-node matrix obtained via non-negative tensor factorization, for $R = 13$. Rows correspond to network nodes and columns to components, here regarded as mesoscopic structural features of the network. The matrix is obtained from the factor **A** by classifying each node as belonging (*light colours*) or not belonging (*dark blue*) to a given component. Node order has been rearranged to expose the block structure of the matrix. Colours identify components, and the structures that correspond to school classes are annotated with the corresponding class name. From [53]

while columns correspond to terms of the decomposition: specifically, the elements a_{ir} and b_{ir} relate individual nodes to components, while the elements c_{kr} associate each component with the times k it spans and can be regarded as an activity level of that component over time. As an illustration, in Fig. 3.4 we display matrix **A** for a sample decomposition of the high-resolution school social network of [33], obtained via non-negative tensor factorization. The method detects both cohesive structures corresponding to school classes and components that describe mixing patterns of the classes induced by scheduled social events such as lunch breaks [53].

Overall, this decomposition model can accommodate the description of mesoscale network structures that mix topological and temporal features in complex fashions: cohesive communities, overlapping communities, groups of links that are only active at specific times, abrupt transitions of the community structure, similar connectivity patterns at distant times, and more. The non-negativity constraints make the decomposition purely additive, and hence yield terms that are more interpretable [54] in relation to contextual information or other background knowledge about the network at hand. We notice that non-negative factorization, because of the properties summarized above, has been already proposed for community detection in static networks [55, 56] when dealing with densely overlapping communities.

We finally remark that a central challenge in designing techniques for detecting mesoscopic structures is the ability to validate the obtained results either by running the decomposition on synthetic benchmark networks or by using empirical data for which a ground truth is independently available, (e.g., the case of [53]).

3.5 Modelling Face-to-Face Interactions

The modelling activity concerning time-varying networks of contacts between individuals is quite recent, mainly because it has followed the availability of time-resolved datasets.[1] For instance, Scherrer et al. [27] have proposed a model of Markovian graph dynamics, in which each link can appear or disappear with probabilities depending on the graph state at each time: This model was tuned to reproduce detailed features of a specific dataset. Another approach consists in considering a set of agents, defining rules of interactions between these agents, and studying the statistical properties of the contact network that emerge from these "microscopic" rules. In particular, the model developed in [58, 59] considers N agents who can either be isolated or form groups. Each agent is characterized by his/her coordination number indicating the number of agents interacting with him/her, and the time at which this coordination number last changed. At each time step, an isolated agent can create a link with another isolated agent, and an agent who is part of a group can leave the group or invite an isolated agent to join it. Each such creation or deletion of links occurs with probabilities that can depend on the concerned agents' status. Interestingly, the introduction of memory effects in the definition of these probabilities is able to generate dynamical contact networks with properties similar to the ones of empirical data sets [58, 59]. In particular, a reinforcement principle can be implemented by considering that the probability that an agent changes his/her state decreases with the time elapsed since his/her last change of state: This is equivalent to the assumption that the longer an agent is interacting in a group, the smaller is the probability that s/he will leave the group, and that the longer an agent is isolated, the smaller is the probability that s/he will form a new group. As a result, the distributions of contact durations and of time intervals between successive contacts of an individual are power-law distributed, and the aggregated contact networks display features similar to the empirically observed ones [58, 59].

Vestergaard et al. [60] consider a similar model in which, for each pair of agents, the probabilities of creation and deletion of links between agents depend on the

[1] See also [57] for more abstract modelling of adaptive networks.

time elapsed since the last evolution of the involved agents. The model considers four different "memory mechanisms" inspired by empirical evidence showing that long-term memory effects akin to self-reinforcing effects are present in the creation and disappearance of links in contact networks. For instance, more active agents tend to create more new contacts and are more attractive to other agents initiating new contacts; moreover, one of the mechanisms captures the fact that one tends to interact more often with close acquaintances. While all these memory effects are combined in empirical data, the modeling framework of [60] makes it possible to explore their individual roles both analytically and numerically. The model analysis shows how each memory mechanism by itself can lead to the emergence of some heterogeneity in the temporal characteristics of the contact networks, as quantified by broad distributions of, e.g., contact durations or inter-contact times. Interestingly however, the whole empirical phenomenology is retrieved only when all four memory mechanisms are introduced into the model.

Another model of interacting agents is put forward in [61, 62]: agents perform here random walks in two dimensions, and two agents are considered as in contact if they are within a certain distance d of each other. The main ingredient of the model is that each agent i is characterised by an intrinsic "attractiveness" $a_i \in [0, 1]$ that can be interpreted as due, for instance, to social status. When an agent is in contact with other agents, s/he can either perform a random walk step or keep the interaction by staying immobile, and the probability to maintain the contact is proportional to the attractiveness of the most attractive neighbour. Agents can also be active (i.e., can have contacts) or inactive with certain probabilities, to mimic the fact that in empirical datasets, individuals can leave the premises and stop having contacts, or come back and start again interacting. The mechanism is illustrated in Fig. 3.5 and leads to heterogeneous distributions of contact durations, of inter contact times and of aggregated contact durations very similar to empirical data (see Fig. 3.6).

3.6 Conclusions and Open Problems

Face-to-face interactions are a crucial element in the fabric of social connectivity. Their properties and their dynamics entangle many complex aspects that comprise the free agency of individuals, social coordination, human mobility and dynamics under spatial constraints, the interplay of stochasticity and deterministic activity patterns, social network structure, organizational structure, multi-layer and time-varying social networks, and more. On top of this, face-to-face interactions mediate and constrain important dynamical processes, such as information diffusion and epidemic spread of infectious agents that are transmissible during a face-to-face interaction. The research agenda on face-to-face interactions, of course, cannot be fully decoupled from domain-specific aspects, but—as it is usually the case for many complex systems—it is possible to discover and exploit summarized data representations, statistical regularities, stylised facts, and minimal models that reproduce a set of observations across diverse contexts.

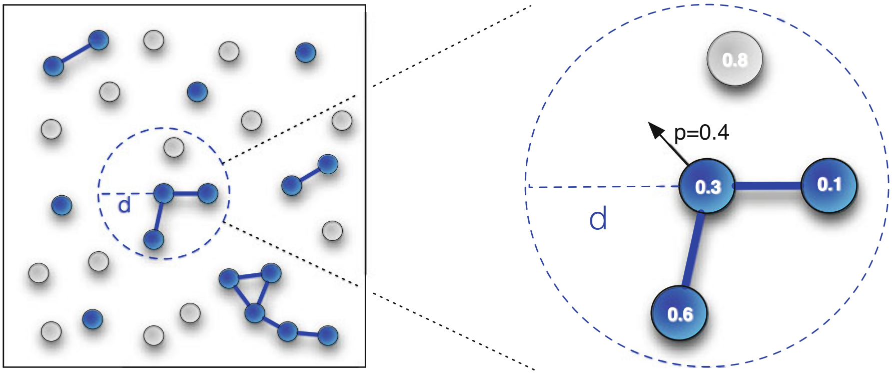

Fig. 3.5 Illustration of the mechanism of interaction of [61]. Each *circle* represents an agent. *Left*: *Dark* agents are active, *grey* (light) agents do not move nor interact. Agents interact if they are within a distance d, and are then connected by a link. *Right*: Each agent is characterized by a number representing attractiveness. The probability for the central individual to move is $p = 1 - \max(0.1, 0.6) = 0.4$ since the attractiveness of the inactive agent is not taken into account. Reprinted figure with permission from Michele Starnini, Andrea Baronchelli, and Romualdo Pastor-Satorras, Phys. Rev. Lett. 110, 168701 (2013). Copyright 2013 by the American Physical Society [61]

The research agenda we envision, thus, starts by building an "atlas" of human contacts, which is incrementally assembled by adding map after map of human encounters, obtained by measuring face-to-face interactions in a variety of social contexts, at different points in time, at different scales, and using different proxies to assess individual interactions. The availability of these empirical datasets allows to make progress in the direction of the following goals:

- To learn which proxy is best suited to measure a given type of close-range interaction in a given context, and how different proxies relate to one another when used to quantify the same face-to-face interactions. We illustrated some of these points in Sect. 3.2.
- To uncover statistical regularities, as discussed in Sect. 3.3. The ultimate goal is not to empirically quantify all interactions in any given environment, but rather to learn what should be measured and what we do not need to measure every time.
- To design summarised data representations such as the contact matrices and aggregated networks discussed in Sect. 3.4 that, ideally, retain only the essential information and generalise well to other environments or social contexts.
- To devise minimal dynamical models, like those described in Sect. 3.5, that reproduce a set of important stylised facts and observed statistical properties under minimal assumptions. Models like these are precious to generate synthetic but realistic interaction networks, and to gain insight into the deep mechanisms that are responsible for the observed behaviors.

All of the above points are aimed at achieving parsimonious representations of the empirical data and parsimonious mathematical models for selected observables. However, it is important to remark that whereas simple generative models can

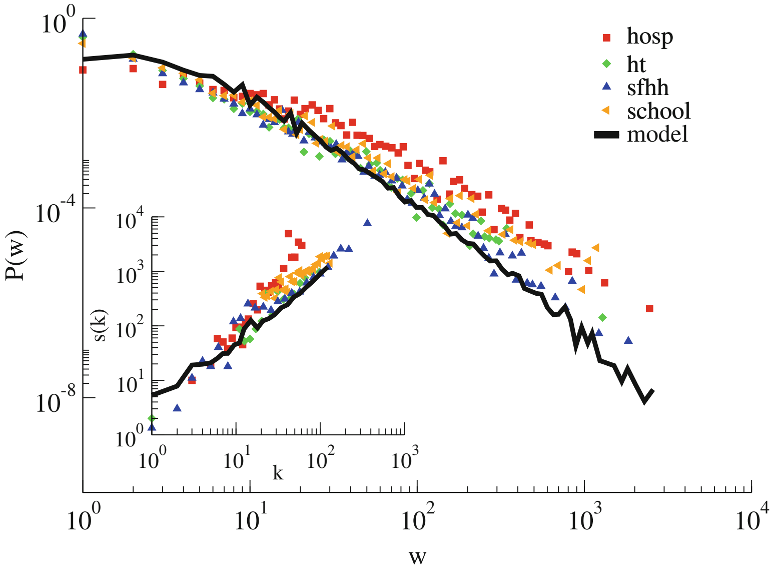

Fig. 3.6 Comparison of the model of [61] with empirical data. *Main figure*: distribution $P(w)$ of links' weights (i.e., aggregated contact durations between pairs of individuals) in the aggregated contact network. *Inset*: average strength s of nodes of degree k in the aggregated network, i.e., average total time in contact (s) of agents having had contacts with k other agents. The datasets correspond to contacts gathered by the SocioPatterns collaboration [21] in a hospital ("hosp"), conferences ("ht" and "sfhh") and in a primary school ("school"). Reprinted figure with permission from Michele Starnini, Andrea Baronchelli, and Romualdo Pastor-Satorras, Phys. Rev. Lett. 110, 168701 (2013). Copyright 2013 by the American Physical Society [61]

reproduce some or even many of the observed statistical distributions, the rich multi-level structure that is visible in face-to-face interaction data cannot emerge from such models: when aiming at realistic scenarios, both in a data mining perspective and in a mathematical modeling perspective, there are specificities of the system at hand that we cannot ignore. Because of this, is it important to develop and validate techniques for detecting structures and correlated activity patterns of face-to-face interactions, as discussed in Sect. 3.4.3. Many highly relevant ideas and methods rooted in the domains of data mining and machine learning can be brought to bear on network science. The design of mathematical models that naturally incorporate the observed longitudinal structures, mesoscopic structures, and correlated activity patterns is an outstanding problem.

In this chapter we often discussed, explicitly or implicitly, epidemic processes over face-to-face interaction networks. This disciplinary focus arises from two reasons:

- A need for simplicity. Biological contagion processes unfold over face-to-face interaction in a mechanistic fashion. To describe their dynamics we need not take into account complex cultural attributes of the individuals that may play a crucial role when dealing with, e.g., information spreading over face-to-face encounters. Moreover, when dealing with airborne infectious agents, the network of close-range encounters in space is generally regarded to be the relevant network for the epidemic process. The same does not hold for, e.g., information diffusion, as face-to-face encounters are just one of many information exchange modalities among humans and the relevant network structure for this process is likely to be a multi-layer network.

- Moving from understanding to control. Controlling and mitigating epidemic processes on face-to-face interaction networks is a challenge that combines data-intensive approaches and mathematical models, with a potentially huge real-world impact. Nosocomial infections alone are a huge burden, both financially and in terms of individual health outcomes, and they occur in a context, the hospital, where it is comparatively easy to measure face-to-face interactions and put them in relation with infection surveillance and microbiological data. In general, there is an opportunity to use knowledge on high-resolution social networks to design mitigation strategies and targeted interventions. In [63], for example, we investigated targeted class-closure strategies for mitigating the epidemic of a flu-like disease in schools.

Despite the recent important advances that we have in part described in this chapter, many other open problems and challenges remain [64]. They include further measurements of face-to-face interactions at different scales and in different contexts, with in particular the comparison and integration of different measurement strategies and the development of means to compensate for missing data due to sampling issues and to the finiteness of the population studied. A crucial challenge, in the context of understanding infectious disease dynamics over face-to-face contact networks, regards also the combination of contact data with virological data to better understand the links between contacts and infection events and to better assess the relative importance of different routes of transmission of various infectious diseases. Another important open problem lies in measuring, understanding, and modeling the reactive aspects of social contact in relation to disease status. This entangles the biological contagion dynamics and behavioural/cultural aspects, greatly increasing the complexity of the dynamics [10, 65, 66].

Let us finally note that, in order to make progress in the research agenda described in this section, continued data gathering efforts using various strategies and in contexts as diverse as possible remain essential, as well as the availability of the corresponding datasets for the research community [21].

Acknowledgements This work is partially supported by the Lagrange Project of the ISI Foundation funded by the CRT Foundation to AB, and CC, by the Q-ARACNE project funded by the Fondazione Compagnia di San Paolo to CC, by the HarMS-flu project (ANR-12-MONU-0018) funded by the French ANR to AB, and by the FET Multiplex Project (EU-FET-317532) funded by the European Commission to AB and CC.

References

1. Read, J. M., Edmunds, W. J., Riley, S., Lessler, J., & Cummings, D. A. T. (2012). Close encounters of the infectious kind: Methods to measure social mixing behaviour. *Epidemiology and Infection, 140*, 2117–2130.
2. Bernard, H., Fischer, R., Mikolajczyk, R. T., Kretzschmar, M., & Wildner, M. (2005). Nurses contacts and potential for infectious disease transmission. *Emerging Infectious Diseases, 15*(9), 1438–1444.
3. Edmunds, W. J., O'callaghan, C. J., & Nokes, D. J. (1997). Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proceedings of the Royal Society of London B, 264*, 949–957.
4. Read, J. M., Eames, K. T. D., & Edmunds, W. J. (2008). Dynamic social networks and the implications for the spread of infectious disease. *Journal of the Royal Society Interface, 5*, 1001–1007.
5. Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., et al. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine, 5*, e74.
6. Mikolajczyk, R. T., Akmatov, M. K., Rastin. S., & Kretzschmar, M. (2008). Social contacts of school children and the transmission of respiratory-spread pathogens. *Epidemiology and Infection, 136*(6), 813–822.
7. Danon, L., House, T. A., Read, J. M., & Keeling, M. J. (2012). Social encounter networks: Collective properties and disease transmission. *Journal of the Royal Society Interface 9*, 2826–2833.
8. Danon, L., Read, J. M., House, T. A., Vernon, M. C., & Keeling, M. J. (2013). Social encounter networks: Characterizing Great Britain. *Proceedings of the Royal Society B, 280*, 20131037.
9. Conlan, A. J. K., Eames, K. T. D., Gage, J. A., von Kirchbach, J. C., Ross, J. V., Saenz, R. A., et al. (2011). Measuring social networks in British primary schools through scientific engagement. *Proceedings of the Biological Sciences, 278*, 1467–1475.
10. Van Kerckhove, K., Hens, N., Edmunds, W. J., & Eames, K. T. D. (2013). The impact of illness on social networks: implications for transmission and control of influenza. *American Journal of Epidemiology, 178*, 1655–1662.
11. Smieszek, T., Burri, E. U., Scherzinger, R., & Scholz, R. W. (2012). Collecting close-contact social mixing data with contact diaries: Reporting errors and biases. *Epidemiology and Infection, 140*, 744–752.
12. Smieszek, T., Barclay, V. C., Seeni, I., Rainey, J. J., Gao, H., Uzicanin, A., et al. (2014). How should social mixing be measured? Comparing survey- and sensor-based methods. *BMC Infectious Diseases, 14*, 136.
13. Eubank, S., Guclu, H., Kumar, V. S. A., Marathe, M. V., Srinivasan, A., Toroczkai, Z., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature, 429*, 180–184.
14. Iozzi, F., Trusiano, F., Chinazzi, M., Billari, F. C., Zagheni, E., Merler, S., et al. (2010). Little Italy: An agent-based approach to the estimation of contact patterns-fitting predicted matrices to serological data. *PLoS Computational Biology, 6*(12), e1001021.
15. Fumanelli, L., Ajelli, M., Manfredi, P., Vespignani, A., & Merler, S. (2012). Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Computational Biology, 8*(9), e1002673.
16. Hui, P., Chaintreau, A., Scott, J., Gass, R., Crowcroft, J., & Diot, C. (2005). Pocket switched networks and human mobility in conference environments. In *Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-Tolerant Networking* (pp. 244–251), Philadelphia, Pennsylvania, USA.
17. O'Neill, E., Kostakos, V., Kindberg, T., Fatah gen Schieck, A., Penn, A., Fraser, D. S., et al. (2006). Instrumenting the city: Developing methods for observing and understanding the digital cityscape. *Lecture Notes in Computer Science, 4206*, 315–322.
18. Pentland, A. (2008). *Honest signals: How they shape our world*. Cambridge, MA: MIT Press.

19. Salathé, M., Kazandjieva, M., Lee, J. W., Levis, P., Feldman, M. W., & Jones, J. H. (2010). A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences, 107*(51), 22020–22025.
20. Hashemian, M., Stanley, K., & Osgood, N. (2010). Flunet: Automated tracking of contacts during flu season. In *Proceedings of the 6th International Workshop on Wireless Network Measurements* (pp. 557–562), Avignon, 1–3 June 2010.
21. http://www.sociopatterns.org/.
22. Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J. -F., & Vespignani, A. (2010). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PLoS One, 5*, e11596.
23. Hornbeck, T., Naylor, D., Segre, A. M., Thomas, G., Herman, T., & Polgreen, P. M. (2012). Using sensor networks to study the effect of peripatetic healthcare workers on the spread of hospital-associated infections. *The Journal of Infectious Diseases, 206*, 1549.
24. Stopczynski, A., Sekara, V., Sapiezynski, P., Cuttone, A., Madsen, M. M., Larsen, J. E., et al. (2014). Measuring large-scale social networks with high resolution. *PLoS One, 9*(4), e95978.
25. Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports, 519*, 97–125.
26. Karagiannis, T., Le Boudec, J. -Y., & Vojnovic, M. (2007). Power law and exponential decay of inter contact times between mobile devices. In *Mobicom 07* (p. 183).
27. Scherrer, A., Borgnat, P., Fleury, E., Guillaume, J. -L., & Robardet, C. (2008). Description and simulation of dynamic mobility networks. *Computer Networks, 52*, 2842.
28. Barrat, A., Cattuto, C., Tozzi, A. E., Vanhems, P., & Voirin, N. (2014). Measuring contact patterns with wearable sensors: Methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clinical Microbiology and Infection, 20*, 10–16.
29. Barabási, A. -L. (2010). *Bursts: The hidden pattern behind everything we do*. London: Dutton Adult.
30. Fournet, J., & Barrat, A. (2014). Contact patterns among high-school students. *PLoS One, 9*(9), e107878.
31. Isella, I., Stehlé, J., Barrat, A., Cattuto, C., Pinton, J. F., & Van den Broeck, W. (2011). What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology, 271*, 166–180.
32. www.gephi.org.
33. Stehlé, J., Voirin, N., Barrat, A., Cattuto, C., Isella, L., Pinton, J.-F., et al. (2011). High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS One, 6*, e23176.
34. Génois, M., Vestergaard, C., Fournet, J., Panisson, A., Bonmarin, I., & Barrat, A. (2015). Data on face-to-face contacts in an office building suggest a low-cost vaccination strategy based on community linkers. Network Science. arXiv:1409.7017.
35. Isella, L., Romano, M., Barrat, A., Cattuto, C., Colizza, V., Van den Broeck, W., et al. (2011). Close encounters in a pediatric ward: Measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS One, 6*, e17144.
36. Vanhems, P., Barrat, A., Cattuto, C., Pinton, J. -F., Khanafer, N., Régis, C., et al. (2013). Estimating potential infection transmission routes in hospital wards using wearable proximity sensors. *PLoS One, 8*(9), 73970.
37. Matson, J. (2012). RFID tags track possible outbreak pathways in the hospital. *Scientific American, 307*. http://www.scientificamerican.com/article/graphic-science-rfids-tags-track-possible-outbreak-pathways-in-hospital/
38. Machens, A., Gesualdo, F., Rizzo, C., Tozzi, A. E., Barrat, A., & Cattuto, C. (2013). An infectious disease model on empirical networks of human contact: Bridging the gap between dynamic network data and contact matrices. *BMC Infectious Diseases, 13*, 185.
39. Eames, K. T. D., Tilston, N. D., & Edmunds, W. J. (2011). The impact of school holidays on the social mixing patterns of school children. *Epidemics, 3*, 103.
40. Gauvin, L., Panisson, P., Cattuto, C., & Barrat, A. (2013). Activity clocks: Spreading dynamics on temporal networks of human contact. *Scientific Reports, 3*, 3099.

41. Yang, J., McAuley, J., & Leskovec, J. (2014). Detecting cohesive and 2-mode communities indirected and undirected networks. In *Proceedings of 7th ACM International Conference on Web Search and Data Mining (WSDM2014)* (pp. 323–332).
42. Chen, Y., Kawadia, V., & Urgaonkar, R. (2013). Detecting overlapping temporal community structure in time-evolving networks. arXiv:1303.7226.
43. Hopcroft, J., Khan, O., Kulis, B., & Selman, B. (2004). Tracking evolving communities in large linked networks. *Proceedings of the National Academy of Sciences, 101*(1), 5249–5253.
44. Greene, D., Doyle, D., & Cunningham, P. (2010). Tracking the evolution of communities in dynamic social networks. In *Proceedings of 2010 International Conference on Advances in Social Networks Analysis and Mining (ASONAM2010)* (pp. 176–183).
45. Bassett, D. S., Porter, M. A., Wymbs, N. F., Grafton, S. T., Carlson, J. M., & Mucha, P. J. (2013). Robust detection of dynamic community structure in networks. *Chaos, 23*(1), 013142.
46. Mucha, P. J., Richardson, T., Macon, K., Porter, M. A., & Onnela, J. -P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science, 328*, 876–878.
47. Ronhovde, P., Chakrabarty, S., Hu, D., Sahu, M., Sahu, K. K., Kelton, K. F., et al. (2012). Detection of hidden structures for arbitrary scales in complex physical systems. *Scientific Reports, 2*, 329.
48. De Domenico, M., Solé-Ribalta, A., Cozzo, E., Kivelä, M., Moreno, Y., Porter, M. A., et al. (2013). Mathematical formulation of multi-layer networks. *Physical Review X, 3*, 041022.
49. Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review, 51*(3), 455–500.
50. Cichocki, A., Phan, A. H., & Zdunek, R. (2009). *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation*. New York: Wiley.
51. Shashua, A., & Hazan, T. (2005). Non-negative tensor factorization with applications to statistics and computer vision. In *Proceedings of 22nd International Conference on Machine Learning (ICML2005)* (pp. 792–799).
52. Lee, D. D., & Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature, 401*, 788–791.
53. Gauvin, L., Panisson, A., & Cattuto, C. (2014). Detecting the community structure and activity patterns of temporal networks: A non-negative tensor factorization approach. *PLoS One, 9*(1), e86028.
54. Nickel, M., Tresp, V., & Kriegel, H. P. (2011). A three-way model for collective learning on multi-relational data. In *Proceedings of 28th International Conference on Machine Learning (ICML2011)* (pp. 809–816).
55. Yang, J., & Leskovec, J. (2013). Overlapping community detection at scale: A nonnegative matrix factorization approach. In *Proceedings of 6th ACM International Conference on Web Search and Data Mining (WSDM2013)* (pp. 587–596)
56. Wang, F., Li, T., Wang, X., Zhu, S., & Ding, C. (2011). Community discovery using nonnegative matrix factorization. *Data Mining and Knowledge Discovery, 22*(3), 493–521.
57. Gross, T., & Sayama, H. (Eds.). (2008). *Adaptive networks: Theory, models and applications*. Springer/NECSI Studies on Complexity Series. Berlin: Springer.
58. Stehlé, J., Barrat, A., & Bianconi, G. (2010). Dynamical and bursty interactions in social networks. *Physical Review E, 81*, 035101(R).
59. Zhao, K., Stehlé, J., Bianconi, G., & Barrat, A. (2011). Social network dynamics of face-to-face interactions. *Physical Review E, 83*, 056109.
60. Vestergaard, C. L., Génois, M., & Barrat, A. (2014). How memory generates heterogeneous dynamics in temporal networks. *Physical Review E*, *90*, 042805.
61. Starnini, M., Baronchelli, A., & Pastor-Satorras, R. (2013). Modeling human dynamics of face-to-face interaction networks. *Physical Review Letters, 110*, 168701.
62. Starnini, M., Baronchelli, A., & Pastor-Satorras, R. (2014). Model reproduces individual, group and collective dynamics of human contact networks. arXiv:1409.0507.
63. Gemmetto, V., Barrat, A., & Cattuto, C. (2014). Mitigation of infectious disease at school: Targeted class closure vs school closure. *BMC Infect. Dis. 14*, 695

64. Eames, K., Bansal, S., Frost, S., & Riley, S. (2015). Six challenges in measuring contact networks for use in modelling. *Epidemics, 10*, 72–77.
65. Meloni, S., Perra, N., Arenas, A., Gómez, S., Moreno, Y., & Vespignani, A. (2011). Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific Reports,1*, 62.
66. Perra, N., Balcan, D., Gonçalves, B., & Vespignani, A. (2011). Towards a characterization of behavior-disease models. *PLoS One, 6*(8), e23084.