



**HAL**  
open science

## Empirical temporal networks of face-to-face human interactions

Alain Barrat, C. Cattuto, V. Colizza, F. Gesualdo, L. Isella, E. Pandolfi, J.-F. Pinton, L. Ravà, C. Rizzo, M. Romano, et al.

► **To cite this version:**

Alain Barrat, C. Cattuto, V. Colizza, F. Gesualdo, L. Isella, et al.. Empirical temporal networks of face-to-face human interactions. *European Journal of Physics Special Topics*, 2013, 222, pp.1295. 10.1140/epjst/e2013-01927-7 . hal-00861907

**HAL Id: hal-00861907**

**<https://hal.science/hal-00861907v1>**

Submitted on 13 Sep 2013

**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.

# Empirical temporal networks of face-to-face human interactions

A. Barrat<sup>1,2</sup>, C. Cattuto<sup>2</sup>, V. Colizza<sup>3,4,5</sup>, F. Gesualdo<sup>6</sup>, L. Isella<sup>2</sup>, E. Pandolfi<sup>6</sup>, J.-F. Pinton<sup>7</sup>, L. Ravà<sup>6</sup>, C. Rizzo<sup>8</sup>, M. Romano<sup>6</sup>, J. Stehlé<sup>1,9</sup>, A. E. Tozzi<sup>6</sup>, and W. Van den Broeck<sup>2</sup>

<sup>1</sup> Centre de Physique Théorique, Aix-Marseille Univ, CNRS UMR 6207, Univ Sud Toulon Var, 13288 Marseille cedex 9, France

<sup>2</sup> Data Science Laboratory, ISI Foundation, Torino, Italy

<sup>3</sup> INSERM, U707, Paris F-75012, France

<sup>4</sup> UPMC Université Paris 06, Faculté de Médecine Pierre et Marie Curie, UMR S 707, Paris F75012, France

<sup>5</sup> Computational Epidemiology Laboratory, ISI Foundation, Torino, Italy

<sup>6</sup> Epidemiology Unit, Bambino Gesù Hospital, Rome, Italy

<sup>7</sup> Laboratoire de Physique de l'Ecole Normale Supérieure de Lyon, CNRS UMR 5672, Lyon, France

<sup>8</sup> National Centre for Epidemiology, Surveillance and Health Promotion, Istituto Superiore di Sanità Rome, Italy

<sup>9</sup> Centre de Recherche en Economie et Statistique, ENSAE, 92240 Malakoff, France

May 21, 2012

**Abstract.** The ever increasing adoption of mobile technologies and ubiquitous services allows to sense human behavior at unprecedented level of details and scale. Wearable sensors, in particular, open up a new window on human mobility and proximity in a variety of indoor environments. Here we review stylized facts on the structural and dynamical properties of empirical networks of human face-to-face proximity, measured in three different real-world contexts: an academic conference, a hospital ward, and a museum exhibition. First, we discuss the structure of the aggregated contact networks, that project out the detailed ordering of contact events while preserving temporal heterogeneities in their weights. We show that the structural properties of aggregated networks highlight important differences and unexpected similarities across contexts, and discuss the additional complexity that arises from attributes that are typically associated with nodes in real-world interaction networks, such as role classes in hospitals. We then consider the empirical data at the finest level of detail, i.e., we consider time-dependent networks of face-to-face proximity between individuals. To gain insights on the effects that causal constraints have on spreading processes, we simulate the dynamics of a simple susceptible-infected model over the empirical time-resolved contact data. We show that the spreading pathways for the epidemic process are strongly affected by the temporal structure of the network data, and that the mere knowledge of static aggregated networks leads to erroneous conclusions about the transmission paths on the corresponding dynamical networks.

## 1 Introduction

Due to the development of sensors of various types and the use of digital media and computational devices, we increasingly leave digital traces of our daily activities. The scale at which such data can be gathered and analyzed affords a novel, data-driven approach in the investigation of various aspects of human behavior at various scales, from mobility patterns [1–7] to instant messaging or email exchange [8–13]. Proximity patterns [14–18] can also be detected using Bluetooth and Wifi technologies, and even face-to-face copresence of individuals can be resolved with high spatial and temporal resolution [19–22].

In many cases, the corresponding information finds a convenient representation in terms of complex networks and the corresponding “network science” approach has led to many interesting analysis and results [23–30]. In this context, the combination of technological advances and of

heterogeneous data sources opens new challenges and opportunities, ranging from theoretical to applied issues. In particular, longitudinal data, which have been traditionally scarce in social network analysis [31, 32], are becoming more accessible. Thus, a dynamical perspective on interaction networks [33] becomes possible, and many new issues can be investigated, such as the interplay of the network dynamics with dynamical processes taking place on these networks.

In this perspective, we review here stylized facts on the structural and dynamical properties of empirical networks of human face-to-face proximity, measured in three different real-world contexts: an academic conference, a hospital ward, and a museum exhibition. We first describe the corresponding data sets in Section 2. In Section 3, we describe the structure of the aggregated contact networks, that project out the detailed ordering of contact events

while preserving the information about temporal heterogeneities in the duration of contacts. We expose the similarities and differences of these networks across contexts, and discuss the case in which individuals are divided into a priori categories or roles. Section 4 is devoted to the study of a “toy” spreading process acting as a probe of causality constraints of a dynamical network. Finally, we discuss some directions of future work in Section 5.

## 2 Data sets

We consider three data sets gathered in the context of the SocioPatterns project [19]. These data describe the face-to-face proximity of individuals in various contexts, with a temporal resolution of 20 seconds. More precisely, the data collection infrastructure developed by the SocioPatterns project is based on active Radio-Frequency Identification Devices (RFID) that exchange ultra-low power radio packets in a peer-to-peer fashion, as described in Refs. [19–22]. As the human body acts as a RF shield at the carrier frequency used for communication, exchange of radio packets between badges is only possible when two persons are at close range (1 to 1.5m) and facing each other. Relations of face-to-face proximity (or “contact”, as we will refer to it in the following) between the individuals wearing the RFID tags are thus detected, and the operating parameters are such that contacts can be assessed with a probability in excess of 99% over an interval of 20 seconds. Once a contact has been established, it is considered ongoing as long as the involved devices continue to exchange at least one radio packet for every subsequent interval of 20 seconds. Conversely, a contact is considered terminated if an interval of 20 seconds elapses with no packets exchanged. For a detailed description of the sensing platform and some of its deployments, see Refs. [19–22].

The first data set we consider (“HT09” in the following) was collected at the ACM Hypertext 2009 conference hosted by the ISI Foundation in Turin, Italy, from June 29<sup>th</sup> to July 1<sup>st</sup>, 2009. The data collection involved about 100 conference participants (75% of the conference attendees), and logged their mobility and respective proximity in the public spaces of the conference venue, as well as in the conference rooms [34, 35]. The sensed information on face-to-face proximity of attendees was used to deliver social networking services for the participants [21, 22]. In the context of such applications, the project also collected rich personal metadata about the participants and their social connections. The anonymized data set of the list of contacts between participants is publicly available on the SocioPatterns website [36].

The second data set studied (“SG” in the following) was gathered at the INFECTIOUS exhibition held at the Science Gallery museum in Dublin, Ireland, from April 17<sup>th</sup> to July 17<sup>th</sup>, 2009 [35, 37]. As part of an interactive exhibit, visitors were equipped with a RFID tag upon entering the venue, and therefore almost the totality of them were tracked as they streamed through the museum. Over 3 months, the individual mobility and proximity relations of more than 14,000 persons were recorded. In this case

participants were anonymous and no attributes are available about the nodes and the links of the measured networks. The data set corresponding to this deployment is also publicly available [38].

Finally, we will discuss a third data set (“OBG” in the following) that was collected in a hospital ward [39], the pediatric ward of the Bambino Gesù Hospital in Rome, Italy. The ward under study has 44 beds arranged in 22 rooms with 2 beds each, and mostly admits children with acute respiratory diseases who do not require intensive care or surgery. Patients admitted into the ward are accompanied by one parent or caregiver who spends the night in the same room. The data collection campaign ran from the 9<sup>th</sup> to the 16<sup>th</sup> of November 2009. In this case, the association of each RFID badge with the category of the person wearing it was recorded, according to the following classes: ward assistants (A), physicians (D), nurses (N), patients (P), and caregivers (C).

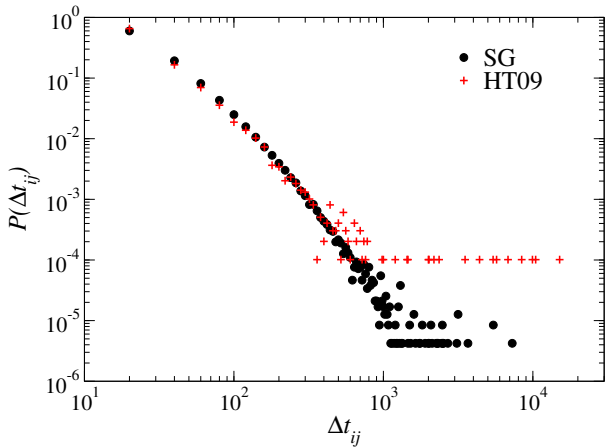
## 3 Aggregated contact networks

The whole temporal resolution of the data sets gives access to the temporal network of interactions as a dynamical object encoding the temporal structure of events. While we will consider this very detailed information in the next section, we first consider contact networks aggregated over a given time window, as very useful summaries of the contact dynamics: in such networks, nodes represent individuals and a weighted link between two nodes represents the fact that the two corresponding individuals have been in interaction. The weight of a link is given by the cumulated duration of the contacts that took place between the two corresponding individuals during the considered time window.

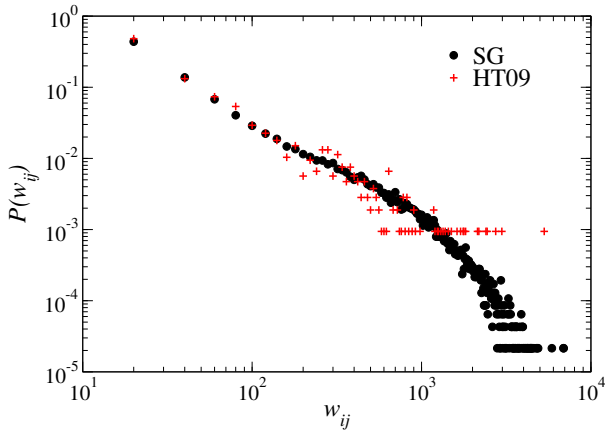
In this context, we note here that we are not dealing with bona fide social networks, as the measurement infrastructure does not provide information about what happens between two individuals while they are in face-to-face proximity. The data described here thus correspond to *behavioral social networks*. As often encountered in other types of real-world networks [28], they can be expected to display many non-trivial features, such as a high level of clustering, community structures and heterogeneous weights. All these features have an important impact on the dynamical processes that unfold on these networks [28].

### 3.1 Similar statistics, different network structures

We first consider the SG and HT09 data sets, in which individuals are not classified into distinct categories. As analyzed in [20, 35] and recalled in Fig. 1, the distribution of contact durations show a very strong similarity despite the differences in the contexts. This similarity is also observed at the aggregated level of the distribution of link weights (i.e., of daily cumulated contact durations), as shown in Fig. 2. The degree distributions of the daily aggregated contact networks, shown in Fig. 3, are short-tailed in both



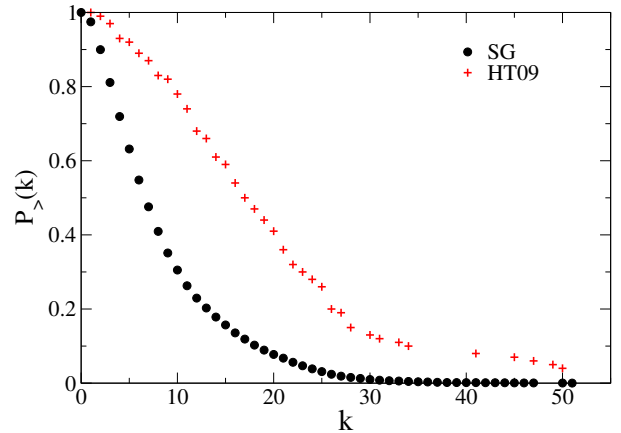
**Fig. 1.** Distributions of the contact durations for the HT09 (pluses) and SG (circles) deployments, averaged over all days. Despite the differences in the measurement contexts, the distributions are superimposed.



**Fig. 2.** Weight distributions for the daily aggregated networks of one HT09 conference day (pluses) and for the SG aggregated networks (circles), averaged over all daily aggregated networks. The weight of a link represents the total time spent in face-to-face proximity by the two linked individuals during the aggregation interval (here one day).

data sets, with exponential decreases at large degrees; the different behavioral patterns of conference attendees and museum visitors are only reflected in the daily average degrees, close to 8 for SG and much larger (close to 20) for HT09, which seems natural given that one of the goals of a conference attendee is precisely to encounter colleagues.

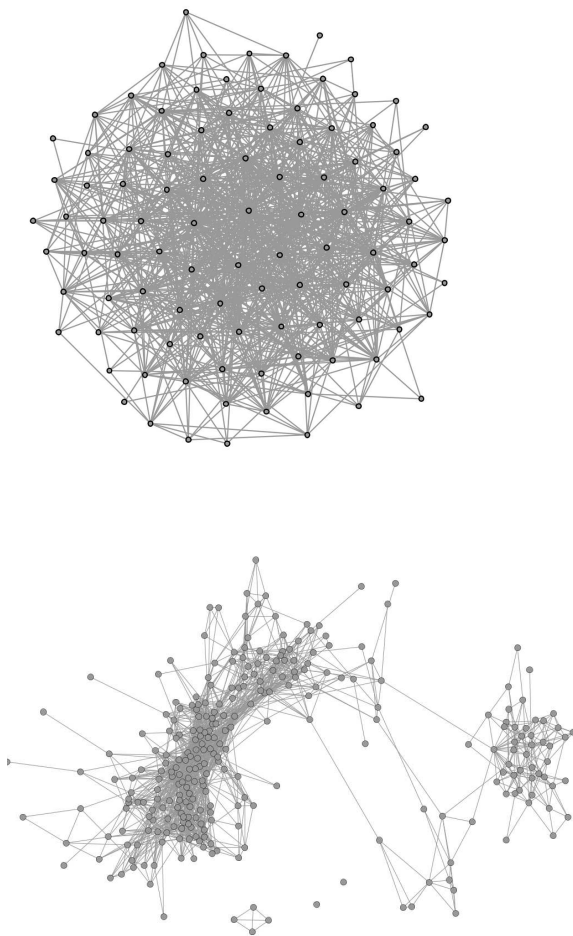
The statistical similarities of the aggregated contact networks hide however important differences in their structures, as the visual inspection of Figure 4 reveals: the aggregated network of interactions during a HT09 conference day is much more “compact” than the ones describing the interactions between museum visitors. Several of the SG daily aggregated networks are even composed of various connected components. A more quantitative analysis shows that a typical daily aggregated network is a small-world network in a conference context, while it lacks this property in the museum case [35]. This difference can



**Fig. 3.** Cumulative degree distributions for the daily aggregated networks of one HT09 conference day (pluses) and for the SG aggregated networks (circles), averaged over all daily aggregated networks.

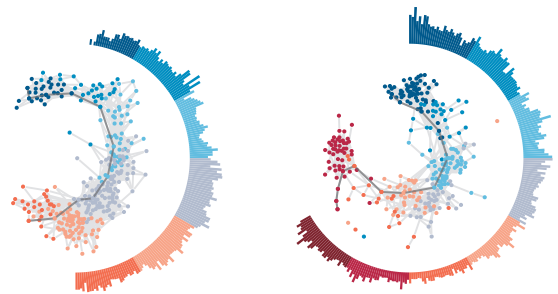
be explained by the different patterns of presence of the attendees on the premises: in the conference case, most participants are present during the whole conference duration (or at least during entire days). In the museum case on the contrary, the visit duration distribution is close to a log-normal, with geometric mean around 35 minutes. As a consequence, museum visitors are unlikely to interact directly with other visitors entering the venue more than one hour after or before them, thus preventing the aggregated network from exhibiting small-world properties. Figure 5 reports the SG aggregated networks for two different days, where the network diameter is highlighted and each node is colored according to the arrival time of the corresponding visitor (see also the visualizations of [40]): there is limited interaction among visitors entering the museum at different times, and the network diameter clearly defines a path connecting visitors that enter the venue at subsequent times, mirroring the longitudinal dimension of the network. These findings show that aggregated network topology and longitudinal/temporal properties are deeply interwoven.

Beyond the topological analysis and the statistical distributions of node’s and link’s properties, the investigation of *correlations* between temporal and topological aspects reveals also some interesting differences between the conference and museum contexts. Let us consider the strength  $s$  of each node, defined as the sum of the weights of all links starting from it [4]: it corresponds, for each individual, to the cumulated time of interaction with other individuals. Correlations between the strength and the degree are quantified by computing the average strength  $\langle s(k) \rangle$  of nodes of degree  $k$ : a random assignment of weights would yield a linear dependency with  $\langle s(k) \rangle \sim \langle w \rangle k$ , where  $\langle w \rangle$  is the average link weight. Other behaviors have been observed in various contexts [4, 20, 42]. A super-linear dependence such as the one observed in some conference settings [20] hints at the presence of super-spreader nodes that play a prominent role in processes such as information diffusion [43, 44]. On the other hand, the sub-linear depen-

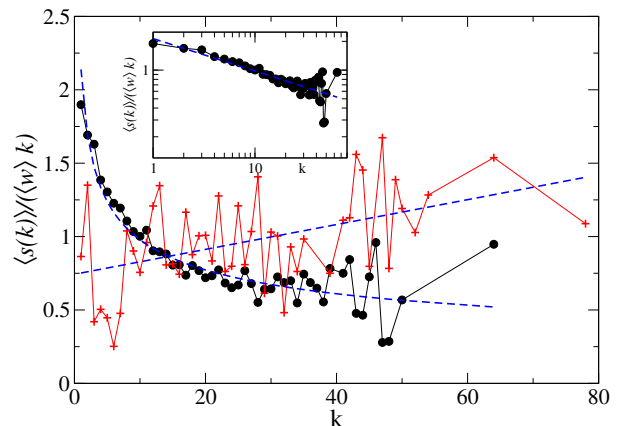


**Fig. 4.** Daily aggregated networks in the HT09 and SG deployments. Nodes represent individuals and edges are drawn between nodes if at least one contact event was detected during the aggregation interval. Top: aggregated network for one day of the HT09 conference. Bottom: one representative day at the SG deployment. The network visualizations were produced using the Gephi software [41].

dence observed for large-scale phone call networks [42] corresponds to the fact that more active individuals spend on average less time in each call. Figure 6 displays the ratio  $\langle s(k) \rangle / (\langle w \rangle k)$  for the SG and HT09 daily aggregated networks, showing two opposite trends: an increasing trend in the conference setting, and a decreasing one in the museum setting. In particular,  $\langle s(k) \rangle / (\langle w \rangle k)$  can be fitted linearly for the HT09 case, with a positive slope, while the SG data can be fitted by a power law with a negative exponent. These results indicate that individuals who encountered the same number of distinct persons can have different spreading potentials, depending on the setting. It also gives a warning about characterizing the spreading potential of a node by only measuring the number of distinct encounters, which can yield a misleading view.



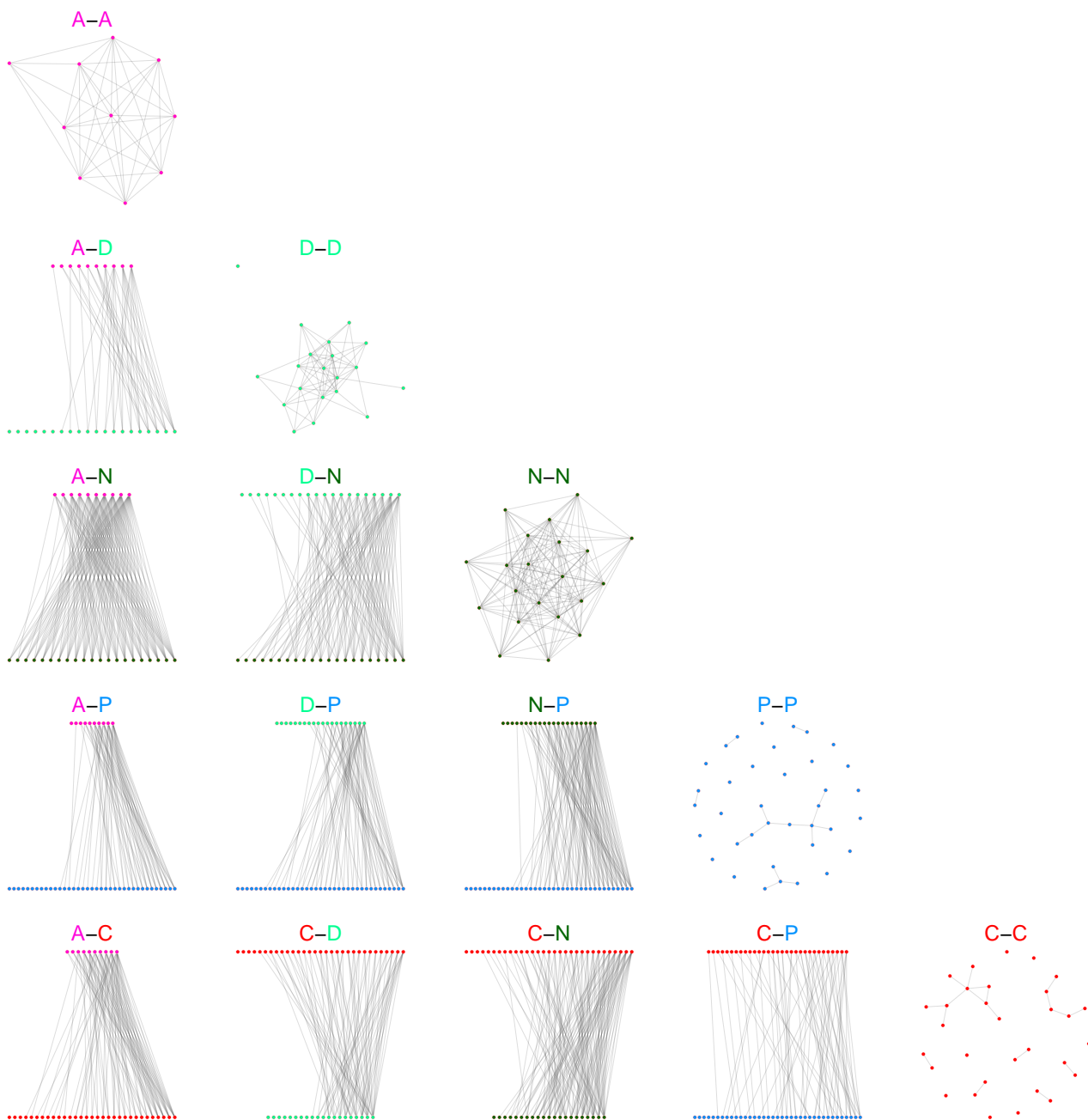
**Fig. 5.** (Color online) Aggregated networks for two days of the SG museum deployment (From [40]). Nodes are colored according to the corresponding visitor's entry time slot. The network diameter is highlighted. The circular bar chart displays the recorded number of social encounters over two-minute intervals. The position and angle of the bars are those of the hour hand of a 12-hour clock at the corresponding time. The colors of the bars match those used for the arrival time of nodes in the graph, so that the bar chart also serves as the legend for the color-coding of time. See [40] for a visualization of all the SG museum deployment days.



**Fig. 6.** (Color online) Correlation between node's strength and degree, as measured by the average strength  $\langle s(k) \rangle$  of nodes of degree  $k$ . The figure shows  $\langle s(k) \rangle / (\langle w \rangle k)$ , for the SG (circles) and the HT09 (pluses). The dashed lines stand for a linear fit and a power law fit to the data for the HT09 and SG deployments, respectively. Distinct increasing and decreasing trends are respectively observed. The inset shows the data for the SG deployment on a doubly logarithmic scale (circles) together with the power law fit to the data (dashed line).

### 3.2 Population categories and contact network subgraphs

In the case of the OBG data set, the fact that the individuals are divided into several categories implies a rich underlying structure of the aggregated networks. The global aggregated network can indeed be divided into two different types of subgraphs: (i) the subgraphs of the interactions taking place between individuals of a given category (e.g., nurse-nurse interactions), and (ii) the bipartite networks of the interactions between members of a category  $A$

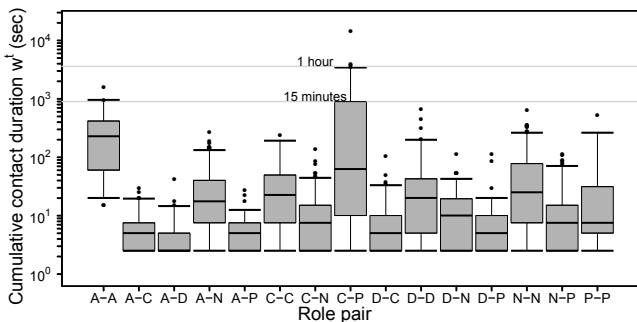


**Fig. 7.** Subgraphs of the cumulative contact networks, for all pairs of categories and within each category. Nodes represent individuals, and edges between nodes represent the occurrence of at least one face-to-face proximity event over the whole monitoring period. In the off-diagonal layouts, nodes are positioned from left to right in increasing order of number of edges.

with members of a different category  $B$ . The corresponding networks are displayed for the OBG data in Figure 7, unveiling important differences in the contact patterns of different categories. The subgraphs describing the interactions between physicians, between nurses or between ward assistants are rather dense, indicating a large diversity of contacts: within a given role, each health care worker interacts with many others. The picture is completely different for patients and caregivers: not only are there very few contacts among caregivers or among patients, but, as

expected, the patient-caregiver contacts are very specific. Each patient has contacts with essentially one caregiver, and vice-versa, which corresponds to the fact that each patient was accompanied by one caregiver.

As in other contexts, the durations of individual contact and the cumulated time spent in face-to-face proximity by pairs of individuals are broadly distributed, spanning several orders of magnitude. Figure 8 shows the distributions of the weights of the subgraphs of the global aggregated network corresponding to the various role pairs



**Fig. 8.** Boxplots for the distributions of cumulative contact durations  $w^t$  between individuals belonging to given role pairs (horizontal axis), given the occurrence of a contact. We only consider non-zero values of  $w^t$ , and contact durations are expressed in seconds and normalized to a 24-hour interval. On normalizing, the experimental resolution of 20 seconds yields the lowest value of 2.5 seconds visible in the figure. The bottom and top of the boxes correspond to the 25<sup>th</sup> and 75<sup>th</sup> percentiles, and the horizontal segment indicates the median. The ends of the whiskers correspond to the 5<sup>th</sup> and 95<sup>th</sup> percentiles. The dots are outliers located outside the 90% confidence interval, i.e., events falling below the 5<sup>th</sup> percentile or above the 95<sup>th</sup> percentile.

(i.e., the subgraphs shown in Fig. 7). Interestingly, all the distributions are broad, showing that strong fluctuations are observed for each category pair. However, the distributions are different and cover distinct ranges, with the strongest fluctuations for the pairs involving one caregiver and one patient.

The aggregated networks encode a rich information about the contact patterns within a category and between categories of individuals, that can be used to simulate dynamical processes unfolding on the network, as discussed in the next section. Moreover, some interesting open questions emerge, regarding for instance the possibility to automatically infer the category of each individual, from the global aggregated network. As the interaction patterns seem to be assortative in some cases (nurses) and disassortative in other cases (patients and caregivers), usual community detection algorithms that focus on finding groups of cohesively connected nodes would be unable to perform such classification, and more advanced machine learning algorithms would most probably be needed.

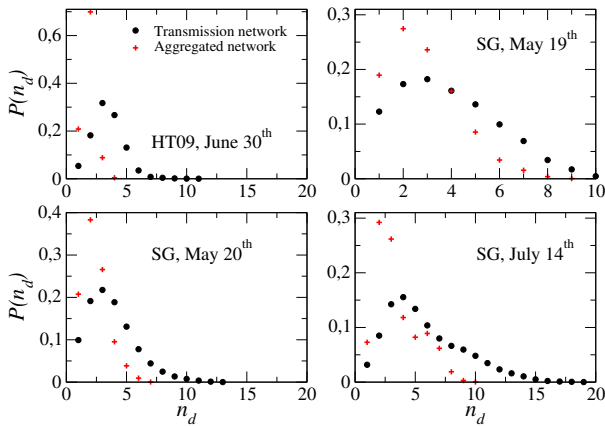
## 4 Dynamical processes on dynamical networks

The measurement infrastructure yields in each deployment very detailed informations on the face-to-face interactions between individuals, including not only the duration, but also the starting time of each contact. While aggregated networks are static representations in which the order of appearance of links is unknown, the gathered data sets contain the time-order of the contacts, i.e. it encodes the causality of the events. These data sets represent therefore instances of dynamical networks, whose study

implies new challenges and opportunities [33]. In particular, some efforts have started to be devoted to the development of new characterization tools specific to temporally evolving networks [45–47]. In this context, the study of dynamical phenomena on dynamical networks can be an interesting way of exploring the role of causality in such networks. To this aim, it is possible to consider a dynamical process taking place (i) either on the dynamical network describing the real temporal sequence of events, or (ii) on an aggregated network in which the information about the precise order of events is discarded. We focus here on spreading processes, that can range from very schematic and simplistic models used as theoretical investigation tools to increasingly complex and detailed models of practical relevance.

Let us first focus on a simple deterministic snowball SI model of epidemic or information diffusion [44]. In this model, individuals are divided into two categories, susceptible (S) who have not yet been reached by the “infection” (or information), and infected ones (I), who carry the disease (or have received the information) and can propagate it to other individuals. In the simplest, deterministic version, every contact between a susceptible individual and an infected one results in a transmission event in which the susceptible becomes infected, according to  $S + I \rightarrow 2I$ . Moreover, infected individuals do not recover. The process is initiated by a single infected individual (“seed”) selected at random. Through its simplicity, such a schematic model gives interesting illustrations on the impact of causality constraints in dynamical networks, and on how different temporal contact patterns can lead to different outcomes.

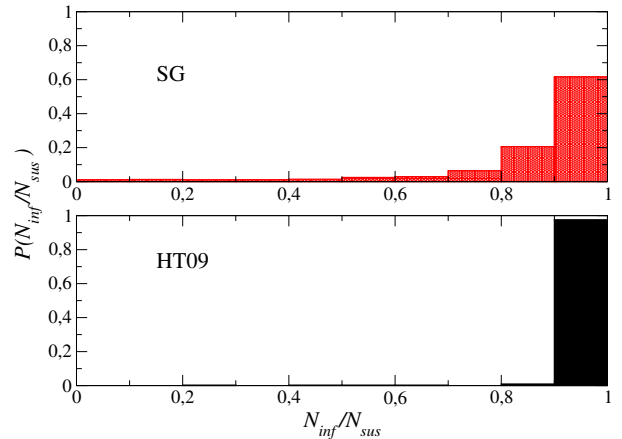
Let us first focus on the difference in the spreading paths on the dynamical network and on the corresponding aggregated network. More precisely, we define, for each seed, two distinct networks: the *transmission network* is the network along which the infection spreads (i.e., the network whose edges are given by  $S \leftrightarrow I$  contacts); the *partially aggregated network* on the other hand is defined as the network aggregated from the time the seed appears in the data set to the end of the day (as the infection can only reach individuals present at the venue after the entry of the seed). The distance between the seed and another node  $i$  in the transmission network gives therefore the actual number of transmission events that occurred in order to reach  $i$ , and is therefore the length of the *fastest* seed-to-infected-individual path which respects causality. On the other hand, it is also possible to compute the *shortest* seed-to-infected-individual path on the partially aggregated network. As shown in Figure 9, the distribution of fastest paths turns out to be broader and shifted toward higher values than the corresponding shortest path distribution, both for a conference day or a museum setting, as it is also observed in other cases [48]. This difference can be understood through a simple example: if  $i$  is contaminated and encounters  $j$ , who then encounters  $k$ , before  $i$  encounters  $k$ , the actual (fastest) spreading path between  $i$  and  $k$  is of length 2, while the shortest path is of length 1.



**Fig. 9.** Distribution of the path lengths  $n_d$  from the seed to all the infected individuals calculated over the transmission network (black circles) and the partially aggregated networks (red pluses). The distributions are computed, for each day, by varying the choice of the seed over all individuals.

Simple processes such as the SI model also provide a comparison tool between different data sets or different dynamical networks. We illustrate from this point of view the differences between the dynamical networks of face-to-face interactions in a museum (SG data set) and during a conference (HT09 data set) in Figures 10 and 11.

As mentioned previously, the spreading process cannot reach individuals who have left the venue before the seed enters. Therefore, we consider the ratio of the final number of infected individuals,  $N_{\text{inf}}$ , to the number  $N_{\text{sus}}$  of individuals who can be potentially reached through causal transmission paths starting at the seed. The distributions of this ratio are reported in Fig. 10. In the case of HT09, almost all the potentially infected individuals will be infected by the end of the day, whereas the distribution of  $N_{\text{inf}}/N_{\text{sus}}$  is broader in the SG case. Within a static network description on the other hand,  $N_{\text{inf}}$  would always be equal to  $N_{\text{sus}}$ , a misleading approximation of reality. Figure 11 gives additional insight into this issue, by showing, for each seed, the cumulative number of infected individuals as a function of time (i.e., the incidence curve), for a selected day of the HT09 conference and for one day of the SG data. In the case of the HT09 conference, the earliest possible seeds are the conference organizers, but few spreading events occur until conference participants gather for the coffee break and/or meet up at the end of the first talk, between 10:00 and 11:00. A strong increase in the number of infected individuals is then observed, and a second strong increase occurs during the lunch break. As a result of the concentration in time of transmission events, spreading processes reach very similar (and high) incidence levels after a few hours, regardless of the initial seed or its arriving time. Even processes started after 15:00 can reach about 80% of the conference participants. Thus, the crucial point for the spreading process does not consist in knowing where and when the epidemic trajectory has started, but whether the seed or any other subsequently infected individual attend the coffee break or



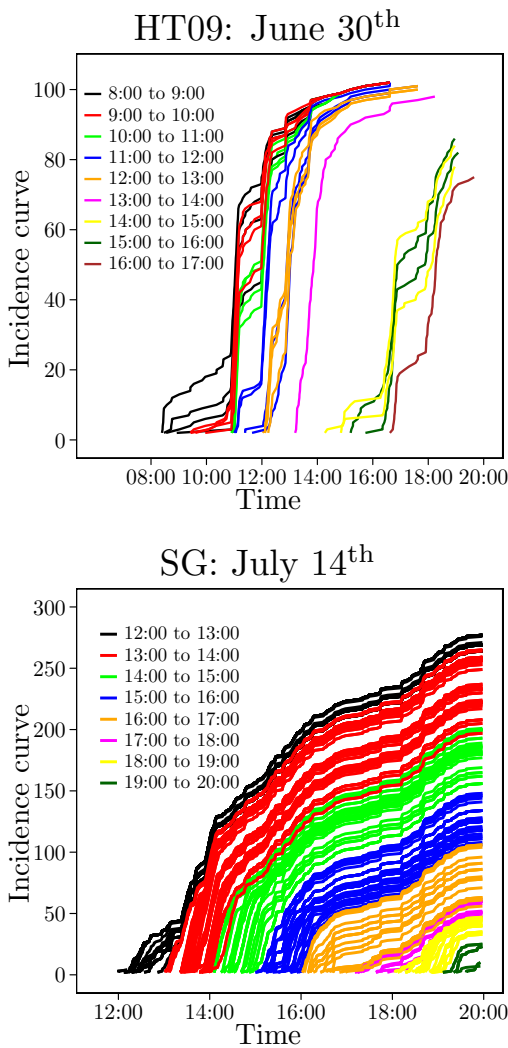
**Fig. 10.** Distribution of the ratio  $N_{\text{inf}}/N_{\text{sus}}$  for the HT09 and the SG data, averaged over all potential seeds.  $N_{\text{inf}}$  is the final number of infected individuals at the end of one day, while  $N_{\text{sus}}$  is the number of individuals that could potentially be reached by a causal transmission path starting at the seed.  $N_{\text{sus}}$  is given by the number of individuals visiting the premises in the same day, from the time the seed enters the premises, and belonging to the same connected component as the seed.

not. A different picture is obtained in the SG case, even in the case shown in the figure, which is a favorable setting for spreading, as it corresponds to a day with many visitors and a globally connected aggregated network: the incidence curves do not present sharp gradients, and later epidemics are unable to infect a large fraction of daily visitors. Once again, such differences in the spreading temporal patterns could not be detected from the study of aggregated networks.

The quantitative description of face-to-face interactions between individuals can also be used to simulate and study the propagation of communicable diseases, using more realistic models of illnesses [49]. In particular, it is possible to compare the simulated spread on the real dynamical contact network with the spreading on aggregated static networks. It turns out that, as the spreading and recovery timescales of realistic disease models is usually of the order of days, an aggregated network taking into account the daily duration of contacts is a good approximation to the full resolution network, whereas a homogeneous representation which retains only the topology of the contact network fails in reproducing the size of the epidemic.

In summary, the role and importance of the causality constraints at a temporal resolution of 20s depends on the timescale of the dynamical process that unfolds on top of the dynamical network. On the other hand, the heterogeneity in the contact durations remains an essential information in order to correctly model and simulate propagation processes.





**Fig. 11.** (Color online) Incidence curves, giving the number of infected versus time for a spreading phenomenon simulated in the HT09 and SG data. Top: HT09, June 30<sup>th</sup> (aggregated network consisting of  $N = 102$  individuals); Bottom: SG network for July 14<sup>th</sup> ( $N = 282$  individuals). Each curve corresponds to a different seed, and is color-coded according to the starting time of the spreading.

## 5 Conclusion

The availability of longitudinal data describing the interactions of individuals in different contexts opens new issues and opportunities in the analysis of complex networks. We have reviewed here recently gathered data sets on behavioral networks of face-to-face proximity in three very different contexts and environments: in a conference, a group of individuals gathers and interacts in a repeated fashion; in a museum on the contrary, a flux of individuals streams through the premises, following a predefined path; finally, in a hospital, individuals with very different roles and mobility patterns interact repeatedly. We have shown how unexpected similar statistical properties emerge, sometimes hiding very important structural prop-

erties of the contact networks. The contact patterns are very heterogeneous, as could be expected, and, in the hospital case, this heterogeneity is observed inside each category of individuals and in the duration of contacts for each category pair, although differences are observed between different category pairs. From this point of view, it would be very interesting to develop automated way of detecting behavioral categories in the interaction patterns and to understand how they match the a priori roles of individuals.

The interplay of “toy” dynamical models such as a deterministic spreading process and of the network temporal dimension has been shown to be an interesting investigation tool for longitudinal data, as it uncovers important differences in the data sets. In this respect, an open issue regards the role of sampling: in many cases one could have access to only a subset of the interactions taking place between the individuals, or even to the interactions of a fraction of the population under study, and it would be interesting to understand how this sampling would affect the results. How the results of section 4 depend on the respective time scales of the dynamical process and of the network is also an interesting direction for future work.

This study was partially supported by the FET-Open project DYNANETS (grant no. 233847) funded by the European Commission.

## References

1. G. Chowell, J. M. Hyman, S. Eubank, C. Castillo-Chavez, Scaling laws for the movement of people between locations in a large city, *Phys. Rev. E* 68 (2003) 066102.
2. A. De Montis, M. Barthélemy, A. Chessa, A. Vespignani, The structure of inter-urban traffic: a weighted network analysis, *Environmental Planning Journal B* 34 (2007) 905–924.
3. D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, *Nature* 439 (2006) 462–465.
4. A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, The architecture of complex weighted networks, *Proc. Natl. Acad. Sci. USA* 101 (2004) 3747–3752.
5. D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, A. Vespignani, Multiscale mobility networks and the spatial spreading of infectious diseases, *Proc. Natl. Acad. Sci. USA* 106 (2009) 21484–21489.
6. M. C. González, C. A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns, *Nature* 453 (2008) 779–782.
7. C. Song, Z. Qu, N. Blumm, A.-L. Barabási, Limits of Predictability in Human Mobility, *Science* 327 (2010) 1018–1021.
8. J.-P. Eckmann, E. Moses, D. Sergi, Entropy of dialogues creates coherent structures in e-mail traffic, *Proc. Natl. Acad. Sci. USA* 101 (2004) 14333–14337.
9. G. Kossinets, D. Watts, Empirical analysis of an evolving social network, *Science* 311 (2006) 88–90.
10. S. Golder, D. Wilkinson, B. Huberman, Rhythms of social interaction: messaging within a massive online network, *Communities and technologies 2007: proceedings of the*

- Third Communities and Technologies Conference*, Michigan State University, 2007.
11. J. Leskovec, E. Horvitz, Planetary-scale views on a large instant-messaging network, *Proceeding of the 17th international conference on World Wide Web (2008)* p 915-924, ACM New York, NY, USA.
  12. D. Rybski, S.V. Buldyrev, S. Havlin, F. Liljeros, H.A. Makse, Scaling laws of human interaction activity, *Proc. Natl. Acad. Sci. USA* 106 (2009) 12640-12645.
  13. R.D. Malmgren, D.B. Stouffer, A.S.L.O. Campanharo, L.A. Nunes Amaral, On Universality in Human Correspondence Activity, *Science* 325 (2009) 1696-1700.
  14. P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Pocket switched networks and human mobility in conference environments, *Proceedings of the 2005 ACM SIGCOMM workshop on Delay-tolerant networking*, Philadelphia, Pennsylvania, USA (2005) 244 - 251.
  15. N. Eagle, A. Pentland, Reality mining: Sensing complex social systems, *Personal and Ubiquitous Computing* 10 (2006) 255-268.
  16. E. O'Neill, V. Kostakos, T. Kindberg, A. Fatah gen. Schieck, A. Penn, D. Stanton Fraser, T. Jones, Instrumenting the city: developing methods for observing and understanding the digital cityscape, *Lecture Notes in Computer Science* 4206 (2006) 315-332.
  17. A. Pentland, *Honest Signals: how they shape our world*, MIT Press, Cambridge MA, 2008.
  18. A. Clauset and N. Eagle, Persistence and periodicity in a dynamic proximity network, *DIMACS Workshop on Computational Methods for Dynamic Interaction Networks*, 2007.
  19. <http://www.sociopatterns.org>
  20. C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, A. Vespignani, Dynamics of person-to-person interactions from distributed RFID sensor networks, *PLoS ONE* 5(7) (2010) e11596.
  21. H. Alani, M. Szomsor, C. Cattuto, W. Van den Broeck, G. Correndo, A. Barrat, Live Social Semantics, 8<sup>th</sup> International Semantic Web Conference ISWC2009, *Lecture Notes in Computer Science* 5823 (2009) 698-714, [http://dx.doi.org/10.1007/978-3-642-04930-9\\_44](http://dx.doi.org/10.1007/978-3-642-04930-9_44).
  22. W. Van den Broeck, C. Cattuto, A. Barrat, M. Szomsor, G. Correndo, H. Alani, The Live Social Semantics application: a platform for integrating face-to-face presence with on-line social networking, *First International Workshop on Communication, Collaboration and Social Networking in Pervasive Computing Environments (PerCol 2010)*, *Proceedings of the 8th Annual IEEE International Conference on Pervasive Computing and Communications*, Mannheim, Germany (2010) 226-231.
  23. Special issue of *Science* on Complex networks and systems. *Science* 325 (2009) 357-504.
  24. S.N. Dorogovtsev, J.F.F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*, Oxford University Press, Oxford 2003.
  25. M.E.J. Newman, The structure and function of complex networks, *SIAM Review* 45 (2003) 167-256.
  26. R. Pastor-Satorras, A. Vespignani, *Evolution and structure of the Internet: A statistical physics approach*, Cambridge University Press, Cambridge, 2004.
  27. G. Caldarelli, *Scale-Free Networks*, Oxford University Press, Oxford, 2007.
  28. A. Barrat, M. Barthélemy, A. Vespignani, *Dynamical processes on complex networks*, Cambridge University Press, Cambridge, 2008.
  29. A. Wasserman, K. Faust, *Social Network Analysis: Methods and applications*, Cambridge University Press, Cambridge, 1994.
  30. D. Watts, *Connections A twenty-first century science*, *Nature* 445 (2007) 489.
  31. J.F. Padgett, C.K. Ansell, Robust action and the rise of the Medici. *Am J Sociol* 98 (1993) 1259-1319.
  32. M.J. Lubbers, J.L. Molina, J. Lerner, U. Brandes, J. Avila, C. McCarty, Longitudinal analysis of personal networks. The case of argentinean migrants in Spain, *Social Networks*, 32 (2010) 91-104.
  33. P. Holme, C. Saramäki, Temporal networks. arXiv:1108.1780.
  34. <http://www.ht2009.org/>
  35. L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, W. Van den Broeck, What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theor. Biol.* 271 166-180 (2011).
  36. <http://www.sociopatterns.org/datasets/hypertext-2009-dynamic-contact-network/>
  37. <http://www.sciencegallery.com/infectious>
  38. <http://www.sociopatterns.org/datasets/infectious-sociopatterns/>
  39. L. Isella, M. Romano, A. Barrat, C. Cattuto, V. Colizza, W. Van den Broeck, F. Gesualdo, E. Pandolfi, L. Ravà, C. Rizzo, A.E. Tozzi, Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS ONE* 6(2): e17144 (2011).
  40. <http://www.sociopatterns.org/gallery/infectious-sociopatterns-poster/>
  41. <http://www.gephi.org>
  42. J.-P. Onnela, J. Saramäki, J. Hyvonen, G. Szabó, M. Argollo de Menezes, K. Kaski, A.-L. Barabási, J. Kertész, Analysis of a large-scale weighted network of one-to-one human communication, *New J. Phys* 9 (2007) 179.
  43. R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, *Phys. Rev. Lett.* 86 (2001) 3200-3203.
  44. R. Anderson, R. May, *Infectious Diseases of Humans: Dynamics and Control*. Oxford: Oxford University Press (1991).
  45. V. Nicosia, J. Tang, M. Musolesi, G. Russo, C. Mascolo, V. Latora. Components in time-varying graphs. arXiv:1106.2134
  46. P. Bajardi, A. Barrat, F. Natale, L. Savini, V. Colizza, Dynamical patterns of cattle trade movements. *PLoS ONE* 6(5):e19869 (2011).
  47. L. Kovanen, M. Karsai, K. Kaski, J. Kertész, J. Saramäki, Temporal motifs in time-dependent networks. *J. Stat. Mech.* P11005 (2011).
  48. G. Kossinets, J. Kleinberg, D. Watts. The Structure of Information Pathways in a Social Communication Network, *Proceedings of the 14<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2008)*.
  49. J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, P. Vanhems. Simulation of a SEIR infectious disease model on the dynamic contact network of conference attendees. *BMC Medicine* 9:87 (2011).