REGULAR ARTICLE

Open Access

CrossMark

Contact activity and dynamics of the social core

Enys Mones[1]* , Arkadiusz Stopczynski[2] and Sune Lehmann[1,3]

*Correspondence: enmo@dtu.dk
[1]Department of Applied
Mathematics and Computer
Science, Technical University of
Denmark, Kgs. Lyngby, Denmark
Full list of author information is
available at the end of the article

Abstract

Humans interact through numerous communication channels to build and maintain social connections: they meet face-to-face, make phone calls or send text messages, and interact via social media. Although it is known that the network of physical contacts, for example, is distinct from the network arising from communication events via phone calls and instant messages, the extent to which these networks differ is not clear. We show here that the network structure of these channels show large structural variations. The various channels account for diverse relationships between pairs of individuals and the corresponding interaction patterns across channels differ to an extent that social ties cannot easily be reduced to a single layer. Each network of interactions, however, contains both central and peripheral individuals: central members are characterized by higher connectivity and can reach a large fraction of the network within a low number of steps, in contrast to the nodes on the periphery. The origin and purpose of each communication network also determine the role of their respective central members: highly connected individuals in the person-to-person networks interact with their environment in a regular manner, while members central in the social communication networks display irregular behavior with respect to their physical contacts and are more active through irregular social events. Our results suggest that due to the inherently different functions of communication channels, each one favors different social behaviors and different strategies for interacting with the environment. These findings can facilitate the understanding of the varying roles and impact individuals have on the population, which can further shed light on the prediction and prevention of epidemic outbreaks, or information propagation.

Keywords: social networks; online behavior; contact patterns; physical proximity

1 Introduction

In modern society, an increasing number of communication channels are available, often relating to different aspects of our lives: we meet others face-to-face to build and maintain social ties [1–4]; we make phone calls for various reasons [5, 6] (as a replacement for physical contacts or simply arranging future appointments); we interact with others on social media [7–9]. Each channel requires different levels of time commitment as well as effort to participate, and may correspond to social ties of different strength [10–12]. Understanding the function of and interplay between these channels has been the subject of increased research interest over the past few years [13–16], along with a growing number of studies focusing on quantitative analysis of multilayer networks [17–19]. On one hand,

the central research questions focus on how the networks corresponding to each channel interact and how the concurrent application of them affects our communication and the dynamics of our social environment. On the other hand, as an increasing fraction of communication takes place via the digital channels, digital traces can provide unprecedented levels of information regarding human behavior and social interactions [10, 20–23].

A central challenge in the analysis of social networks is to identify the central individuals within a certain context solely based on their position in the global structure of the interactions [24, 25]. While it has been shown that there are differences regarding how people position themselves with respect to digital networks [26, 27], it remains unclear whether these differences materialize in aspect of their physical contacts. Specifically, do central members of a social network show specific behavioral patterns in their physical proximity networks as well?

Here we analyze the interplay between digital networks and real-world physical contacts by analyzing multi-channel data from more than 500 university students. First, we show that the frequency of interaction on social networks and by phone calls is not trivially correlated with physical contacts, indicating a fundamental difference between these networks. Secondly, we discuss how, depending on the physical distance of the proximity contacts (which is related to the strength of the social tie between the actors [11]), communication networks show varying levels of structural similarity with proximity networks. As a result, we find that the physical distance in proximity interactions provides information about the nature of the contacts, that is, short-range interactions resemble the communication networks more closely [11]. Finally, we quantify the behavioral differences of central individuals with respect to the communication and proximity networks. By measuring the intensity and regularity of physical engagement with the entire population, we reveal that students central in the digital communication networks exhibit high relative activity during evenings and in the weekend, and are less predictable compared to the population average.

2 Results
2.1 Strength of ties

Different interaction channels can represent fundamentally different aspects of a relationship and correspond to different strengths of a social tie. Phone calls and text messages are known to occur primarily between family members and acquaintances, with high call duration and frequency indicating a strong relationship [11]. On the contrary, social network sites, such as Facebook or Google+, serve as a platform to maintain a wide range of social interactions from instant messaging to posts, or quick responses to events in an individual's the ego-network. Due to the absence of substantial effort and time commitment needed to engage, social network channels constitute a weaker form of direct communication and may suggest a weaker social link. We distinguish between a link and a contact: a *link* represents a relationship between two individuals, such as a social tie (friends, acquaintances), whereas a *contact* is a single physical or online realization of interaction on the link. In other words, links form the backbones on which contacts take place. Here we consider the *functional* network of Facebook, that is, each time a student interacts with any other (via posting on wall, tagging, commenting, etc.), an activity contact is formed, irrespective of the interaction type. This is in contrast to the network of Facebook friendships, which does not involve active participation once the relationship is

established. Physical proximity plays an essential part in maintaining relationships. The presence of face-to-face contacts is frequently reported as having the strongest impact on emotional connection as well as representing the strongest ties [28–31]. Nevertheless, the mere presence of proximity between individuals does not imply a social interaction (as proximity can occur without active communication between participants), and thus physical proximity cannot trivially be used for inferring social connections. To emphasize the importance of physical distance, here we make the distinction between two types of proximity interactions [32]: *ambient*, corresponding to a physical distance of up to 10-15 m between the participants; and *intimate* that requires a distance of 1 m or less. The ambient and intimate networks do not perfectly correspond to their names. That is, every intimate interaction does not correspond to a social interaction (e.g. it is possible to be physically proximate without engaging in a conversation), and the intimate network is a strict subset of the ambient network, so this network includes all social interactions. Additionally, due to environmental noise, the mapping from bluetooth strength to distance is not perfect and a small fraction of links from the ambient network will erroneously appear in the intimate network (see Methods for the details on the construction of these interactions).

Although it has been shown that social ties can be inferred from online activity (or vice versa) [33–36], the interplay between online and offline communication channels is inherently complex and therefore estimating the strength of social ties by reducing them to a single channel (or aggregating the respective networks) may have non-trivial implications. Additionally, we compare the basic structural properties of the proximity and communication networks in the Appendix, where we report degree distributions along with the distribution of link weights. In agreement with the findings of Mastrandrea et al. [37], we found that weights are distributed over multiple orders of magnitude.

Here, we focus on understanding what differences in the level of engagement across channels can reveal about certain behavioral patterns. After an overview of the structural differences and underlying correlations found in these networks, we focus on the physical activity of the central individuals, selected based on their aggregated intensity of physical contacts (strength) throughout a week.

Figure 1 summarizes the usage activity across channels by investigating how ties are expressed across the networks. As we can see in Figure 1a, there are remarkable variations in how the various channels exhibit links (and thus contacts) between pairs of individuals. The Venn diagrams show the number of links in the different networks as well as those present in two (enclosed in the overlap of two circles) or three networks (enclosed within three circles). In other words, the numbers represent the edge set overlap between the various channels. The proximity networks contain a vast majority of all links (only 172 and 356 of all recorded links are not present in the ambient and intimate networks, respectively) and a dominant fraction of all links are exclusively represented as physical links (67,812 and 19,631, accounting for 48% and 14% of all potential links). Compared to the links observed in the proximity networks, a moderate number of interactions (1.81% and 5.96% of all interactions considering ambient or intimate network, respectively) are covered by Facebook activity and a negligible fraction of links (0.51% and 1.68%) are present in the phone call network as well. The presence of phone call-only relationships is, in part, due to the fact that we use a single month time window; by increasing the period of observation, those interactions diminish. Also note that the number of links present in all three channels remains around 180, irrespective the proximity channel considered. The
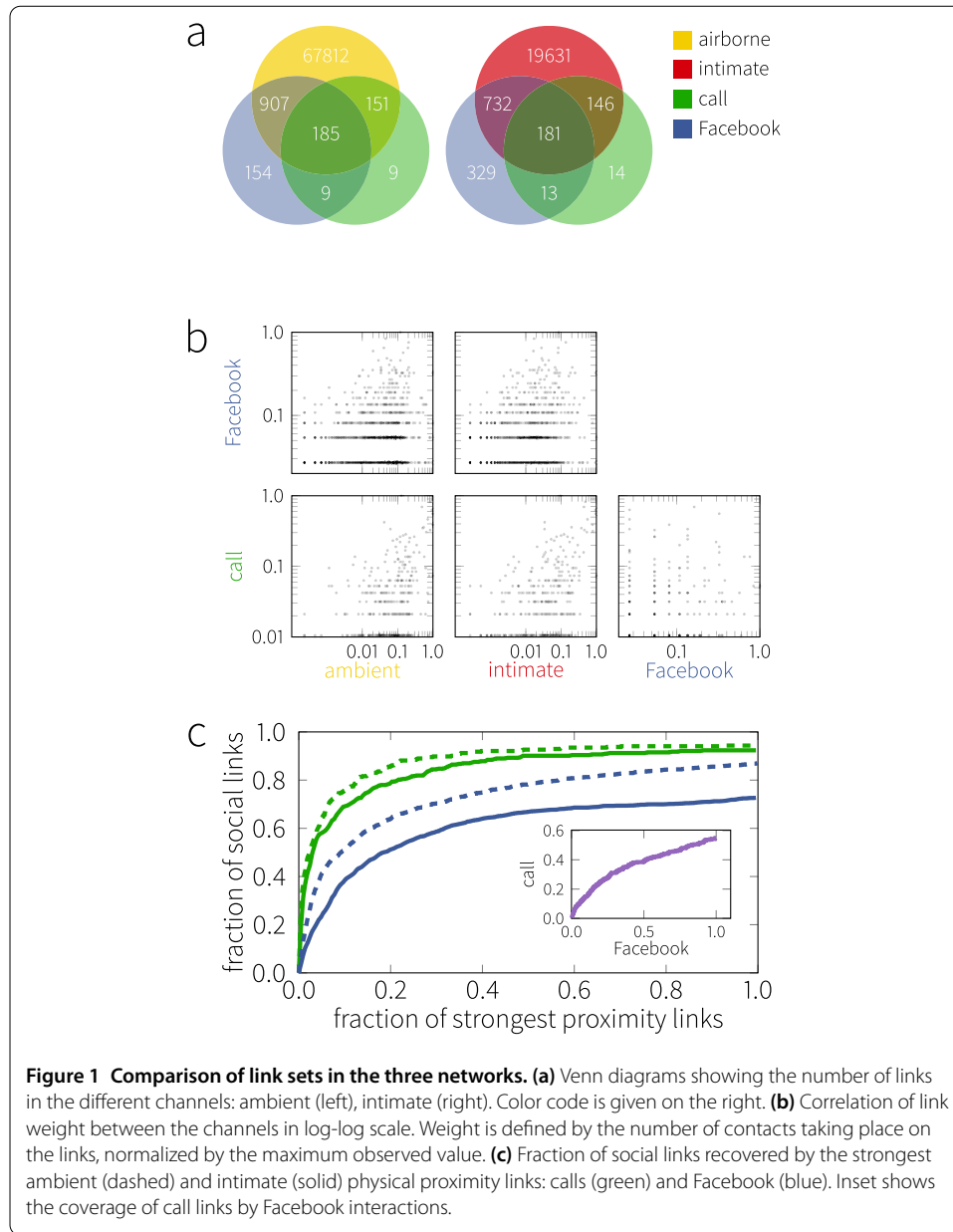
Figure 1 Comparison of link sets in the three networks. (a) Venn diagrams showing the number of links in the different channels: ambient (left), intimate (right). Color code is given on the right. **(b)** Correlation of link weight between the channels in log-log scale. Weight is defined by the number of contacts taking place on the links, normalized by the maximum observed value. **(c)** Fraction of social links recovered by the strongest ambient (dashed) and intimate (solid) physical proximity links: calls (green) and Facebook (blue). Inset shows the coverage of call links by Facebook interactions.

fact that these sets of links consist of the same pairs of users, suggests that even though the structure of the ambient network is blurred by spurious encounters, after removing those links that are not present in other channels, it is still possible to recover the strong links represented in the intimate network.

While it has been argued that intensity of digital communication does not necessarily correlate strongly with the strength of the corresponding social links [38], we expect ties expressed in the phone network or Facebook interaction network to correspond to real social ties and therefore we expect these ties to be stronger (i.e., active with high frequency) in the physical proximity network [11, 32]. Surprisingly, this is only partly true, as shown in Figure 1b, where weight of the links (number of interactions on the link) is plotted. On one hand, the absence of structure in the plot of Facebook and call weights indicates that these two communication channels are used interchangeably (with a Pearson correlation
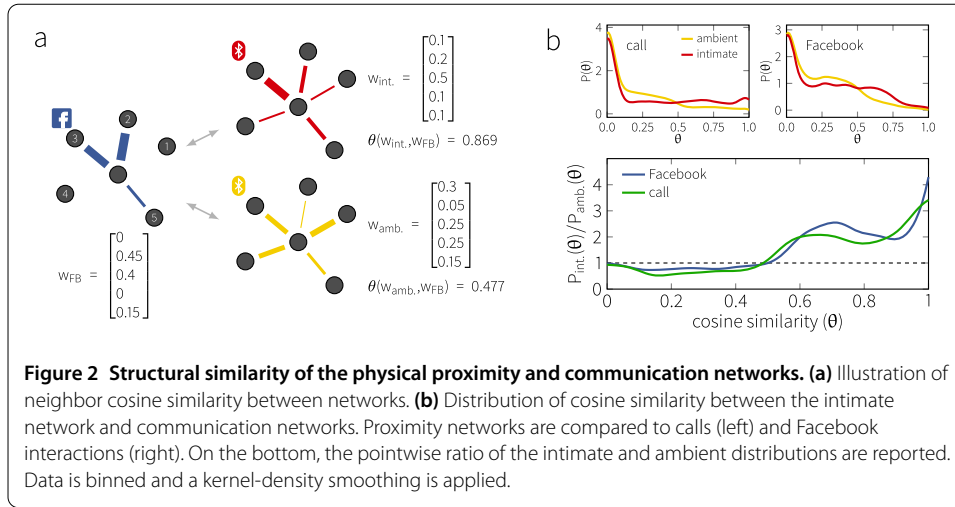
of $r_{\text{Facebook,call}} = 0.007$). On the other hand, the communication networks show moderate positive correlation with the physical links: $r_{\text{Facebook,ambient}} = 0.130$, $r_{\text{Facebook,intimate}} = 0.146$ and $r_{\text{call,ambient}} = 0.510$, $r_{\text{call,intimate}} = 0.554$. In general, the call network shows higher correlation with respect to link weight with the proximity networks compared to Facebook activity, and the intimate network seems to be better predictor of strong social ties having a consistently higher correlation with communication channels than that of the ambient network. The highest correlation is found between calls and intimate interactions.

Assuming that calls and Facebook activity (i.e., online communication) correspond to strong social ties, the capability of proximity links to predict those ties can be assessed by calculating the fraction of links in the former two networks that are covered by the strongest physical links. This is shown in Figure 1c. The most striking observation is that once around 5,000 of the strongest proximity links are considered ($\approx 7\%$ and 24% of links in the ambient and intimate networks, respectively), the contribution of additional links is comparably small (even negligible in case of call links). That is, almost none of the remaining links correspond to links found in the Facebook or call networks. Again, we see that call links can be captured more efficiently by proximity links than Facebook interactions. However, even in the case of the intimate network, the strongest 1,000 links (around 5% of the intimate proximity links) cover only 58% of all call links, meaning that although a large fraction of digital communication links are also included in the proximity networks (see Figure 1a), these links are not the necessarily the strongest physical links, and the ordered set of social ties are separated by many strong proximity links that are not represented by phone calls. In other words, many high-frequency physical links correspond to passive and socially less significant interactions. As the inset of Figure 1c illustrates, the strongest call links are distinct from the strongest Facebook interactions, indicating that the links characterized by the most intense communication on Facebook constitute separate group from that of the most frequent mobile calls. In other words, individuals tend to avoid mixing the two channels and limit the maintenance of relationships to one of them. In the Appendix, we include further evaluation of how well proximity links can predict those of communication links.

It is important to realize that a majority of the proximity links (both ambient and intimate) are due to the co-location of students attending the same classes, which explains the low correlations seen in the data. It should be noted, however, that due to the nature of ambient and intimate networks, the latter exhibits nevertheless higher agreement with the call and Facebook networks. In the next section, we further elaborate on this observation and show that the intimate network also shows higher level of structural similarity with the communication networks.

2.2 Structural similarity

Besides single links, we can compare the local structure of the networks, that is, the ego-networks. To this end, we calculate the similarity between the contacts of an individual in the different networks, as shown in Figure 2. For a given participant u, we first consider their generalized neighbor-set, which consists of all other participants, and construct the weighted degree vector w_c^u that corresponds to the distribution of interactions with u's alters in channel c. In other words, the weighted degree describes how the user distributes their time over their links (that is, the number of contacts over the links) in a given channel (Figure 2a). Similarity between the weighted degree of a specific individual in two different

Figure 2 Structural similarity of the physical proximity and communication networks. (a) Illustration of neighbor cosine similarity between networks. **(b)** Distribution of cosine similarity between the intimate network and communication networks. Proximity networks are compared to calls (left) and Facebook interactions (right). On the bottom, the pointwise ratio of the intimate and ambient distributions are reported. Data is binned and a kernel-density smoothing is applied.

channels c and c' is calculated using the cosine similarity:

$$\theta\left(w_c^u, w_{c'}^u\right) = \frac{w_c^u \cdot w_{c'}^u}{\|w_c^u\| \|w_{c'}^u\|},$$

where $x \cdot y$ denotes the scalar-product of vectors x and y, while $\|x\|$ is the ℓ_2 norm of a vector x. When compared to the communication networks, the distributions $P_{\text{int.}}(\theta)$ and $P_{\text{amb.}}(\theta)$ characterize how similar the intimate and ambient networks are to the call and Facebook activity networks. To quantify how different each of the proximity networks are from the communication networks, we consider the distribution of cosine similarity values, shown in Figure 2b. In the top two plots we report the distribution of the similarity between the proximity networks and the digital communication networks. The plots show that both ambient and intimate networks exhibit low similarity with the communication networks at the level of ego networks (distributions peak around zero, see top of Figure 2b). However, in case of both call and Facebook networks, we observe that the distribution of similarity between intimate and the communication network is consistently lower for values $\theta < 0.5$ and higher for values of $\theta > 0.5$, indicating a higher correspondence between ego networks. The fact that this difference is persistent for both digital communication networks supports the notion that intimate network is more similar to the call or Facebook networks (see Figure 6 of the Appendix for detailed analysis of the overlap between proximity and digital communication networks). In the following, we limit our analyses to the intimate network. We note, however, that the qualitative results are the same for both ambient and intimate contacts, although the phenomena introduced in this paper is more pronounced for the intimate network.

2.3 Contact patterns
In each network, a small set of individuals can be considered as central members of their respective communities. Proximity networks are characterized by high link density and therefore measures based on geodesic distance fail to distinguish among individuals. In other words, due to the high number of links, the distribution of any shortest-path based centrality is narrow, and the value of the centrality measure is not descriptive of the individual's status in the network. Therefore, in the physical proximity network we rank

students according to their total time spent in the proximity with others and choose the ones with the highest time as central, that is, by their strength. We argue that the strength of a node is a meaningful measure of centrality. In the context of epidemic monitoring, Smieszek and Salathé have shown, that strength is able to locate potential candidates for the monitoring problem with efficiency close to the optimal solution [39]

In case of phone call and Facebook networks, however, the networks are sparser and higher order centrality measures can be meaningfully applied. Nevertheless, in order to compare groups of the same selection process, here we first identify central members of the communication networks based on their degree. Due to the high level of heterogeneity in the degree distribution in these networks, degree already differentiates among individuals of low and high activity (and centrality). In addition, as we show in the Appendix, results obtained by considering higher level structural features such as closeness centrality in the communication networks lead to the same qualitative conclusions. Central individuals in the communication networks - we call these the *social core* - within each different network show distinct activity patterns with respect to their physical contacts.

To show the different physical behavior of the central individuals selected by the physical or online networks, for each student we calculate the intensity of physical contacts defined by the number of proximity interactions in each hour of the week for a fixed period of time (between February and May, 2014). After each individual weekly proximity interaction intensity is calculated (which is expected to characterize the time of the week these individuals are more active in their physical contacts), for each group of most central individuals, we average their intensity patterns. For reference, we also calculate the same average intensity for randomly selected individuals. Results are shown in Figure 3a, depicting an important difference when compared to the population average. Central members of the proximity network (i.e., the proximity driven group) relative to the social core engage actively with the population as a whole according to the circadian rhythm and weekly schedules: most contacts take place during the day while students attend classes, with decreased intensity in the night. Furthermore, the activity pattern of these individuals is not only consistent with the population average, but they also display a periodic intensity limited to weekdays. However, the social core shows high contact activity during the afternoon, night and during the weekend, irrespective of the communication channel by which they are selected.

Figure 3b depicts a more detailed comparison of the activity of the social core and the proximity driven group, illustrated by the difference in the relative frequency that a social core or proximity driven group member interacts with any other individual. In the plots, each tile shows the relative frequency of physical interactions by the top ten members of the social core during a specific hour of the week, minus the relative frequency of interactions including the proximity driven group in the same hour. Small white circles indicate a difference that is outside the error bars defined by the standard error of the mean. We refer to the outlined hours in the working days as *working hours*, to distinguish that period from the rest of the week, that is, from hours where most of the voluntary and social activities are expected to take place (*social hours*). More precisely, the interval of working hours is defined by the period between 8am and 16pm (inclusive) on weekdays (from Monday to Friday). The periods of social hours are defined by the rest of the 168 hours of the week. The social core shows decreased activity during working hours compared to the proximity driven group, and they are more active in the evening and nights as well as during the
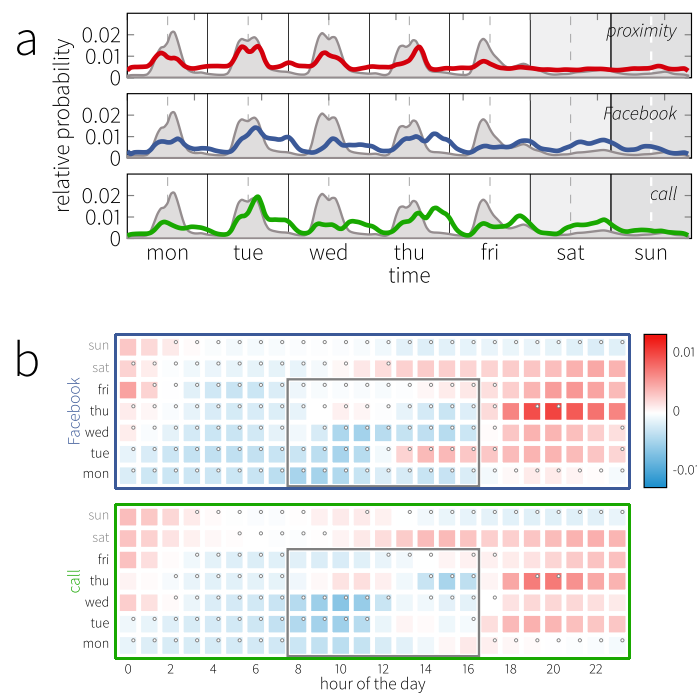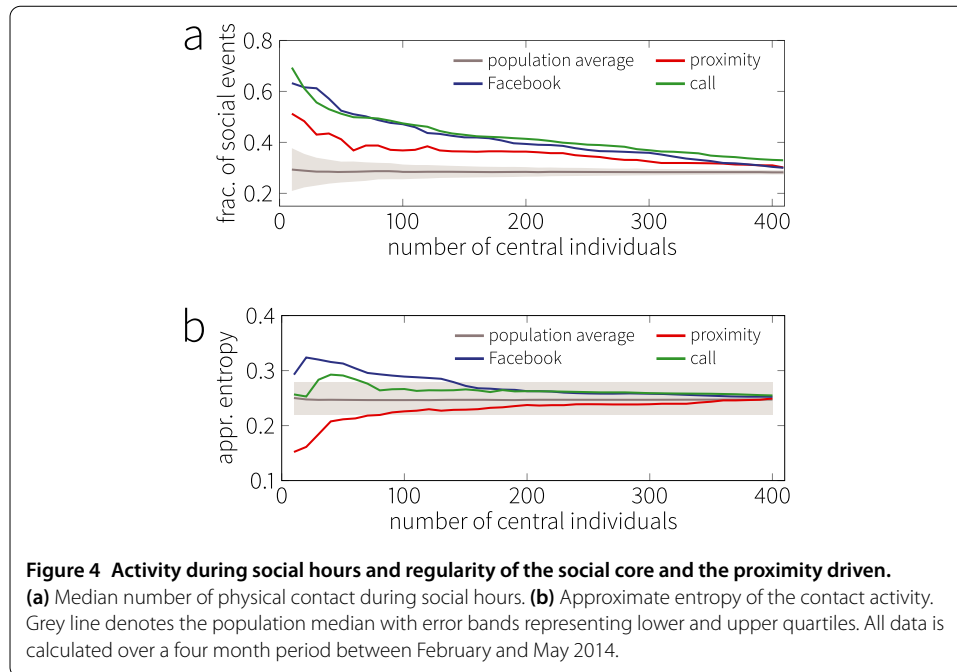
Figure 3 Contact activity of central individuals in the proximity network. (a) Relative probability of physical contacts, compared to the distribution of the population average (grey area). Dark regions indicate weekends. **(b)** Heat maps showing the difference in the contact activity between the social core and the proximity driven for each hour in a week. The periods for working hours are surrounded by the grey frame. In all cases, the 10 most central individuals are considered, and results are aggregated over a four month period between February and May 2014 (inclusive). Hours of significant difference are marked by the white circles.

weekend (especially Monday night and Thursday evening). Also note that for most of the days (from Monday to Saturday), hours in the early morning display significant differences from the proximity driven group, suggesting that the behavior of the social core deviates from the rest of the population predominantly during working hours and nights.

2.4 Regularity

The interaction frequency patterns in Figure 3 indicate distinct behaviors in terms of activity as well as regularity in the case of the social core and the proximity driven group, respectively. We quantify active periods and regularity in Figure 4 for groups as a function of increasing number of central individuals. In case of the population average, curves represent the average of median values over a sample of 1,000 randomly chosen groups. First, we compare the fraction of contact events that take place during social hours to the population average for all three channels. The social core is characterized by a large fraction of contact events during social hours, and the difference does not vanish even for a group of 300 individuals, that is, almost 60% of the population (Figure 4a). Proximity-central individuals are also more active during social hours, however, they show less deviation from the population average. Note that although the period of social hours is longer than that of the working hours, and therefore contacts have comparably higher probability to fall in social hours than to working hours, we merely focus on the relative behavior of the social core and the proximity driven group.

Figure 4 Activity during social hours and regularity of the social core and the proximity driven.
(a) Median number of physical contact during social hours. **(b)** Approximate entropy of the contact activity.
Grey line denotes the population median with error bands representing lower and upper quartiles. All data is
calculated over a four month period between February and May 2014.

To measure regularity of the activity patterns, we calculate the approximate entropy of
the relative frequency of contact events through a four month period. Approximate entropy (ApEn) quantifies the level of irregularity in time series, comparing it to a completely
periodic signal [40, 41]. We chose ApEn due to its robustness against noise and because it
can be efficiently computed from limited data. Results are shown in Figure 4b with sampling length of $m = 2$ and filter level of $r = 0.25$, however, results are robust with respect
to the choice of m or r. Here we observe a strong effect: proximity based central individuals have an ApEn value that is even below the population average meaning that these
individuals are more regular than the average. On the other hand, the social core shows
sign of high irregularity for a large range of group sizes, starting with an ApEn that is 25%
higher than the population average. The difference in the regularity measure of the social core and the proximity driven group vanishes only above the size of 200 individuals
(approximately 40% of the population).

3 Discussion

With advances in technology, humans have begun using a variety of new communication
channels. It is known that the different networks of interactions (physical contacts, online social media, phone calls, etc.) correspond to different types of communication and
can be the proxy for the strength of the social ties. Due to the varying function of these
communication modes, it is interesting to ask the question: Are the same individuals central in all networks? Here we studied differences among individuals that are central in two
fundamentally diverse environments: the network of physical proximity contacts and in
digital communication networks (Facebook interactions and phone calls). By locating central members within all three networks for a single coherent population, we find that the
central members are described by qualitatively different presence and activity patterns in
the physical contact networks. We note however, that our population of students is not

representative of a general population. Therefore, further research is needed to validate the generality of our findings.

The most central members of the population with respect to physical contacts, interact with others in a regular manner: they are most active during official schedule of a week (working hours) and follow a rather periodic activity pattern. Therefore, their interactions can be easily predicted as they are limited by circadian rhythm and weekly schedules. On the contrary, those central in the communication network (the social core), display increased activity during periods of time outside working hours, that is, during events not restricted by work-day schedules. The social core also shows more irregular interaction activity and their interactions are therefore more difficult to predict.

Thus, if one were designing predictive or preventive strategies for counteracting infectious disease based on data from digital communications alone, it could prove useful to incorporate knowledge of the different behaviors of the social core and the proximity driven group, respectively. More generally, while digital communication networks are quite different from the networks formed by person-to-person interactions, the idiosyncratic behavioral patterns of the social core illustrate some of the ways in which digital communication channels can be used to understand aspects of our off-line social interactions.
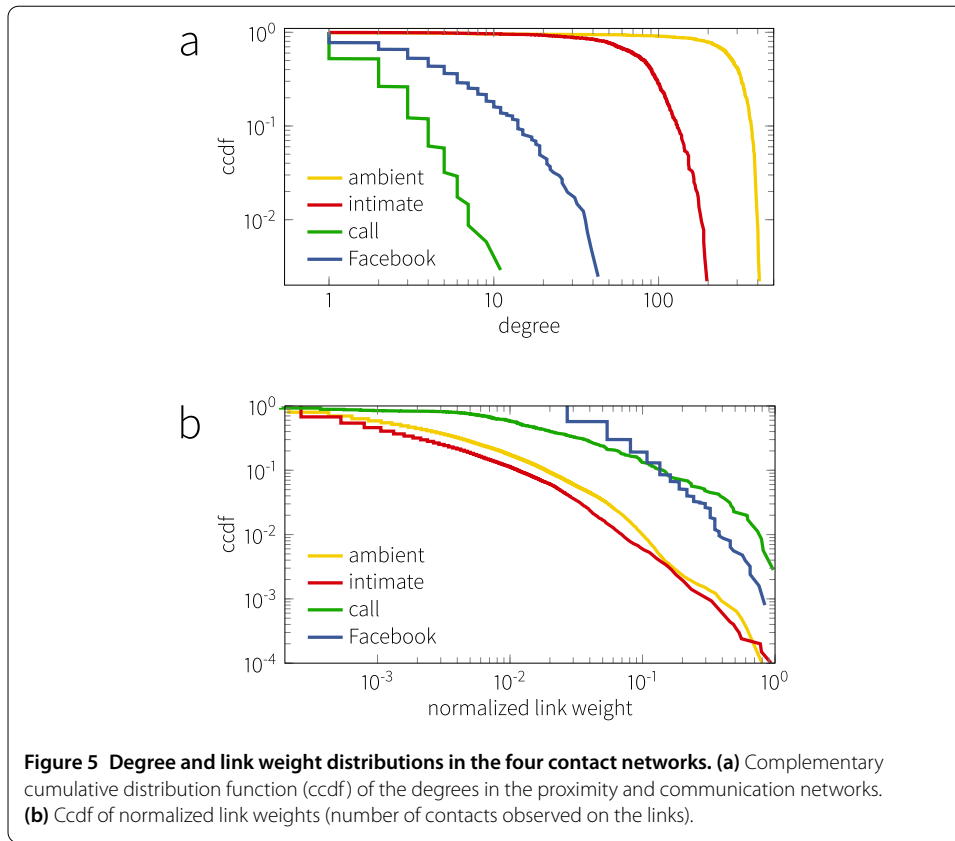
4 Methods
4.1 Data
Data was collected during the Copenhagen Network Study (CNS) between 2012 and 2014 [42], and the results presented in this paper are obtained by analyzing the period from February to May 2014. During the experiment, various data was collected from 1,000 smartphones handed out to students of the Danish Technical University. Among others, the channels included in the data collection are GPS, Bluetooth scans, Wifi scans, call detail records and Facebook communication activity. All data types have different temporal resolution, but in each case a minimal resolution of 5 minute is ensured. Details of the data collection and basic characteristics are presented in Stopczynski et al. [42].

Due to the nature of the data and methodological choices, these results are subject to various limitations that we discuss in the following. First, there is a fraction of students with missing data resulting in low data quality. To avoid working with structurally biased networks due to data loss, we selected a subset of students based on their coverage of proximity data: during the period of February-May 2014, we considered participants with signals in at least 60% of the total time. After the above filtering of the data, the size of the population considered in this paper is 532. We note that there is a possibility that data quality is correlated with sociodemographic traits, an issue which might result in the final study population differing from the student population as a whole.

4.2 Networks
From the CNS data, we built three types of networks: physical proximity networks are based on the Bluetooth scans of the devices. These networks can be thresholded by the received signal strength index (RSSI) to obtain proximity networks with a distance of 1 m (by setting RSSI > -75 dBm). Note that after thresholding based on the RSSI, some of the ambient (full range) contacts disappear due to their low signal strength. Also, due to environmental noise, any fixed threshold on RSSI will result in a mix of short- (<1 m) and long-range (>1 m) contacts. These two distances are not fully linearly separable based on

Figure 5 Degree and link weight distributions in the four contact networks. (a) Complementary cumulative distribution function (ccdf) of the degrees in the proximity and communication networks. **(b)** Ccdf of normalized link weights (number of contacts observed on the links).
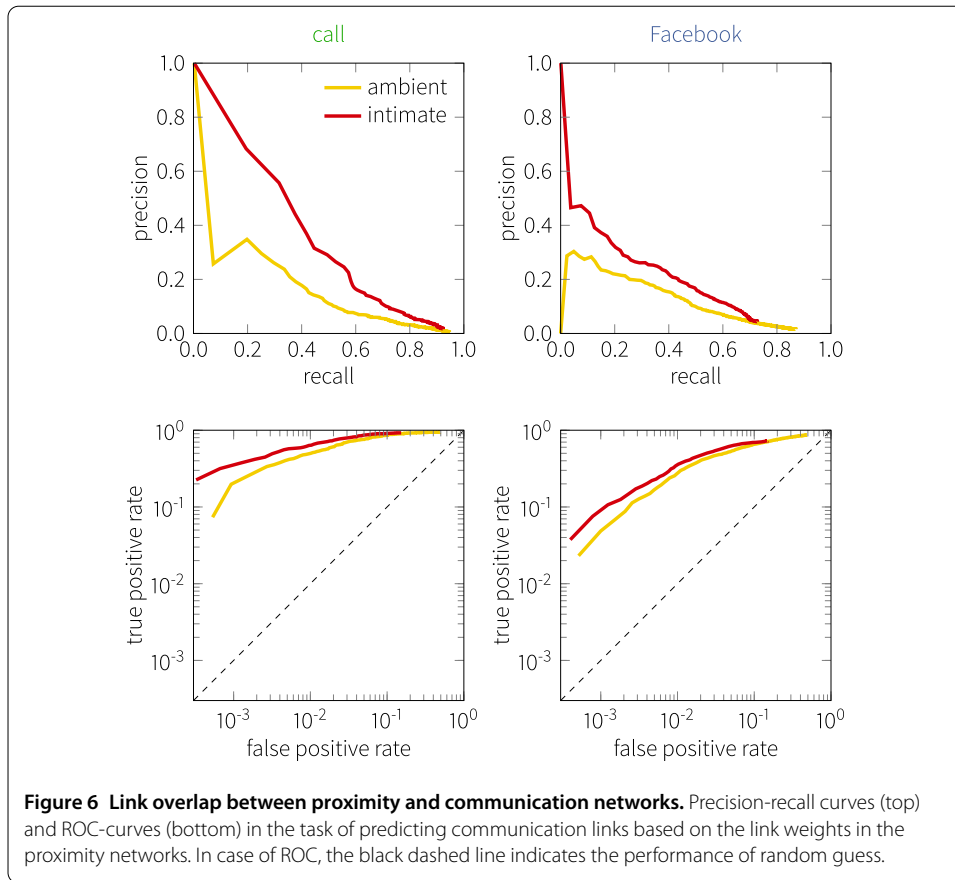
Bluetooth RSSI even in a lab setting (as discussed in Ref. [11]). However, it has been shown in Ref [32] that using a fixed threshold results in significantly different network structures, capturing notions of intimate and ambient networks. Finally, note that the ambient network is a superset of the intimate network, as we allow for short range interactions as well in the former. In other words, each network sets an upper limit on the contact distance.

Facebook feed and phone calls are used to create the communication networks: all interactions inside the population of 532 individuals are aggregated and static weighted networks are constructed. Weights of the links are defined by the relative number of occurrence of the specific link across the time period we consider. In the proximity network, weights represent the total time two individuals spent in the proximity of each other. However, in the activity analysis based on the call and Facebook networks, we did not consider link weight in these networks because unusual communication patterns within romantic couples (extremely high link weight) form a strong bias towards those links and their presence renders other contacts negligible. Therefore, each link in the digital communication networks is set to unit weight, and we consider the embeddedness of the students in these networks rather than their intensity of communication.

4.3 Central groups

In each network, we select central individuals, i.e., central groups of size *n*, by ranking the participants by a centrality measure and considering the n ones with the lowest rank. In case of proximity network, students are ranked by the total time spent in the proximity of others, while target groups in the communication networks are selected by their degree. In

Figure 6 Link overlap between proximity and communication networks. Precision-recall curves (top) and ROC-curves (bottom) in the task of predicting communication links based on the link weights in the proximity networks. In case of ROC, the black dashed line indicates the performance of random guess.

the Appendix, we show results using closeness centrality in the communication networks, which is defined by

$$C_C(i) = \frac{N-1}{\sum_{j \neq i} d_{ij}},$$

where N is the number of participants and d_{ij} denotes the geodesic distance between participant i and j, i.e., the lowest number of steps to reach j from i. In case of a disconnected graph, d_{ij} is defined to be N.

Appendix
A.1 Comparison of networks
Figure 5 depicts the complementary cumulative distribution function (ccdf) of the degrees and link weights (number of contacts on a specific link) for the four networks in question. Weights are normalized by the largest value to obtain comparable distributions. In the proximity networks, weights are proportional to the total time two individuals spent in each other's proximity, whereas in the communication networks weight is the number of contacts between the individuals (irrespective of the length of the contact). The two distributions show opposite behavior: communication networks display the well-known heavy-tailed degree distributions, with a narrower link weight distribution. On the contrary, proximity networks have narrow degree distribution (with most of the individuals
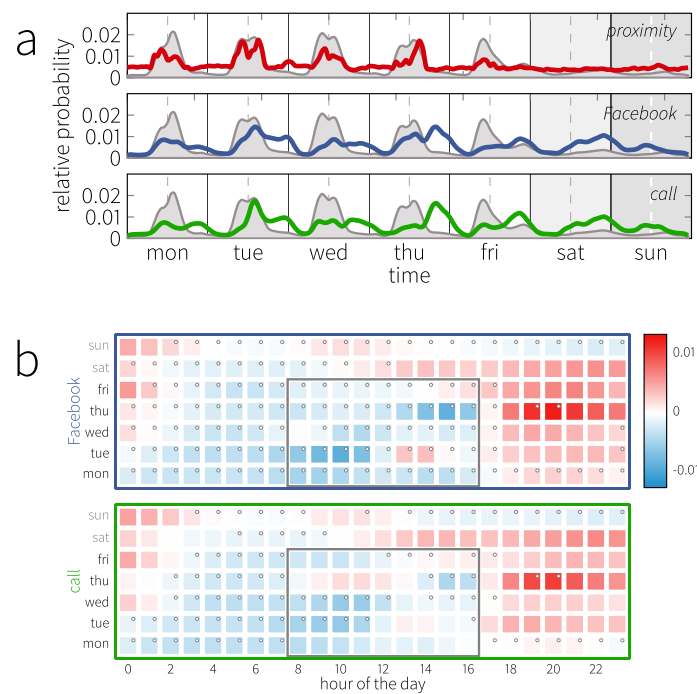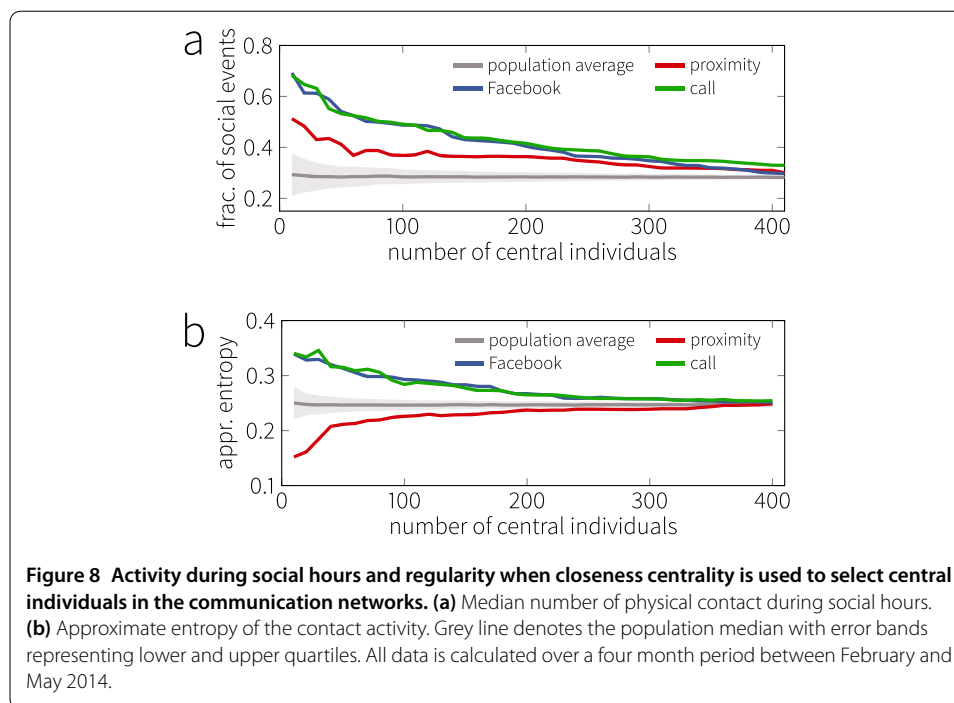
Figure 7 Physical activity of central individuals when closeness centrality is used in the communication networks. (a) Relative probability of physical contacts, compared to the distribution of the population average (grey area). Dark regions indicate weekends. **(b)** Heat maps showing the difference in the contact activity between the social core and the proximity driven for each hour in a week. The periods for working hours are surrounded by the grey frame. In all cases, the 10 most central individuals are considered, and results are aggregated over a four month period between February and May 2014 (inclusive). Hours of significant difference are marked by the white circles.

having more than 100 distinct links), but the link weights are more heterogeneously distributed, spanning from single occasional contacts to links displaying long interactions (i.e., couples or people living in the same dormitory). Results are in agreement with the findings of Mastrandrea et al. [37].

As further assessment of how well the set of links in different networks can be mapped to each other, we have calculated the precision-recall and ROC curves between the proximity and communication networks, shown in Figure 6. In each case, we have sorted the proximity links according to their weight in descending order and calculated the confusion matrix after including more and more proximity links. In other words, we considered the problem of predicting communication links using the proximity link weight as a predictor. Due to the nature of the problem, there is a large imbalance represented by the vast amount of non-existent links with a true positive ratio of 0.0025 and 0.0089 in the phone call and Facebook networks respectively). Therefore, the precision-recall curve is more informative of the differences observed in the link overlaps.

As the top figures in Figure 6 show, the intimate network outperforms the ambient network in predicting links in both digital communication networks (phone call and Facebook), although the difference is more pronounced in the phone call network. The bottom figures show the ROC curve (true positive ratio vs false positive ratio), compared to a random guess (dashed line). However, as the ROC curve is sensitive to class imbalance, the relative position of the ambient and intimate networks is more relevant than the actual

Figure 8 Activity during social hours and regularity when closeness centrality is used to select central individuals in the communication networks. (a) Median number of physical contact during social hours. **(b)** Approximate entropy of the contact activity. Grey line denotes the population median with error bands representing lower and upper quartiles. All data is calculated over a four month period between February and May 2014.

shape of the curves. Nevertheless, we note that both networks outperform random guess, suggesting some level of correspondence between proximity and communication links. Note that even if all physical proximity links are considered, we are not able to account for some of the links observed in the digital communication networks, which explains the absence of some precision-recall and ROC values in the interval $[0, 1]$.

A.2 Central individuals

In the main text we have shown the comparison of physical activity among central individuals selected based on their communication and proximity network degree. Here we show that the difference is present and even more pronounced if higher level structural properties are taken into account. Figure 7 shows the relative contact activity of the three groups compared to the population average in case central individuals selected by their closeness centrality in the communication networks. As Figure 7b indicates, differences are even more significant. Similarly, Figure 8 shows the observed fraction of interactions during social hours as well as the approximate entropy among the central individuals of the proximity and communication networks. Deviations are larger than what we find for degree, especially in case of small groups.

Author details
[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark. [2] Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [3] The Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
1. Eubank S, Guclu H, Kumar VSA, Marathe MV, Srinivasan A, Toroczkai Z, Wang N (2004) Modelling disease outbreaks in realistic urban social networks. Nature 429(6988):180-184
2. 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(7):11596
3. Salathé M, Kazandjieva M, Lee JW, Levis P, Feldman MW, Jones JH (2010) A high-resolution human contact network for infectious disease transmission. Proc Natl Acad Sci 107(51):22020-22025
4. Starnini M, Machens A, Cattuto C, Barrat A, Pastor-Satorras R (2013) Immunization strategies for epidemic processes in time-varying contact networks. J Theor Biol 337:89-100
5. Onnela J-P, Saramäki J, Hyvönen J, Szabó G, Lazer D, Kaski K, Kertész J, Barabási A-L (2007) Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci 104(18):7332-7336
6. Aledavood T, Lehmann S, Saramäki J (2015) On the digital daily cycles of individuals. Front Phys 3:73
7. Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B (2007) Measurement and analysis of online social networks. In: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement. IMC'07. ACM, New York, pp 29-42
8. Ellison NB, Steinfield C, Lampe C (2007) The benefits of Facebook "friends:" social capital and college students' use of online social network sites. J Comput-Mediat Commun 12(4):1143-1168
9. Grabowicz PA, Ramasco JJ, Moro E, Pujol JM, Eguiluz VM (2012) Social features of online networks: the strength of intermediary ties in online social media. PLoS ONE 7(1):1-9
10. Eagle N, Pentland AS, Lazer D (2009) Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci 106(36):15274-15278
11. Sekara V, Lehmann S (2014) The strength of friendship ties in proximity sensor data. PLoS ONE 9(7):100915
12. Gilbert E, Karahalios K (2009) Predicting tie strength with social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. CHI'09. ACM, New York, pp 211-220
13. Eagle N, Pentland AS (2006) Reality mining: sensing complex social systems. Pers Ubiquitous Comput 10(4):255-268
14. Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Van Alstyne M (2009) Computational social science. Science 323(5915):721-723
15. Raento M, Oulasvirta A, Eagle N (2009) Smartphones: an emerging tool for social scientists. Sociol Methods Res 37(3):426-454
16. Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton J-F, Quaggiotto M, Van den Broeck W, Regis C, Lina B et al (2011) High-resolution measurements of face-to-face contact patterns in a primary school. PLoS ONE 6:23176
17. De Domenico M, Solé-Ribalta A, Cozzo E, Kivelä M, Moreno Y, Porter MA, Gómez S, Arenas A (2013) Mathematical formulation of multilayer networks. Phys Rev X 3:041022
18. De Domenico M, Nicosia V, Arenas A, Latora V (2015) Structural reducibility of multilayer networks. Nat Commun 6:6864
19. Nicosia V, Latora V (2015) Measuring and modeling correlations in multiplex networks. Phys Rev E 92:032805
20. Aharony N, Pan W, Ip C, Khayal I, Pentland A (2011) Social fmri: investigating and shaping social mechanisms in the real world. Pervasive Mob Comput 7(6):643-659
21. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, Van den Broeck W, Gesualdo F, Pandolfi E, Ravà L, Rizzo C et al (2011) Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS ONE 6(2):17144
22. Miller G (2012) The smartphone psychology manifesto. Perspect Psychol Sci 7(3):221-237
23. Staiano J, Pianesi F, Lepri B, Sebe N, Aharony N, Pentland A (2012) Friends don't lie: inferring personality traits from social network structure. In: Proceedings of the 2012 ACM conference on ubiquitous computing. UbiComp'12. ACM, New York
24. Freeman LC (1978) Centrality in social networks conceptual clarification. Soc Netw 1:215
25. Garcia-Herranz M, Moro E, Cebrian M, Christakis NA, Fowler JH (2014) Using friends as sensors to detect global-scale contagious outbreaks. PLoS ONE 9(4):92413
26. Christakis NA, Fowler JH (2010) Social network sensors for early detection of contagious outbreaks. PLoS ONE 5(9):12948
27. Manrique P, Cao Z, Gabriel A, Horgan J, Gill P, Qi H, Restrepo EM, Johnson D, Wuchty S, Song C, Johnson N (2016) Women's connectivity in extreme networks. Sci Adv 2(6):e1501742
28. Sherman LE, Michikyan M, Greenfield PM (2013) The effects of text, audio, video and in-person communication on bonding between friends. Cyberpsychol J Psychol Res Cyberspace 7(2):3
29. Antheunis ML, Valkenburg PM, Peter J (2013) The quality of online, offline and mixed-mode friendships among users of a social networking site. Cyberpsychol J Psychol Res Cyberspace 6(3):1
30. Marsden PV, Campbell KE (1984) Measuring tie strength. Soc Forces 63(2):482-501
31. Mesch G, Talmud I (2006) The quality of online and offline relationships: the role of multiplexity and duration of social relationships. Inf Soc 22(3):137-148
32. Stopczynski A, Pentland AS, Lehmann S (2015) Physical proximity and spreading in dynamic social networks. ArXiv preprint. arXiv:1509.06530

33. Scellato S, Noulas A, Mascolo C (2011) Exploiting place features in link prediction on location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, New York, pp 1046-1054

34. Jones JJ, Settle JE, Bond RM, Fariss CJ, Marlow C, Fowler JH (2013) Inferring tie strength from online directed behavior. PLoS ONE 8(1):1-6

35. Sapiezynski P, Stopczynski A, Wind DK, Leskovec J, Lehmann S (2016) Inferring person-to-person interactions using Wifi signals

36. Sapiezynski P, Stopczynski A, Wind DK, Leskovec J, Lehmann S (2016) Offline behaviors of online friends

37. Mastrandrea R, Fournet J, Barrat A (2015) Contact patterns in a high school: a comparison between data collected using wearable sensors, contact diaries and friendship surveys. PLoS ONE 10(9):1-26

38. Wiese J, Min J-K, Hong JI, Zimmerman J (2015) Call and sms logs do not always indicate tie strength. In: Proceedings of the 18th ACM conference on computer supported cooperative work & social computing. ACM, New York, pp 765-774

39. Smieszek T, Salathé M (2013) A low-cost method to asses the epidemilogical importance of individuals in controlling infecitous disease outbreaks. BMC Med 11(35):1-8

40. Pincus SM (1991) Approximate entropy as a measure of system complexity. Proc Natl Acad Sci 88(6):2297-2301

41. Pincus SM, Gladstone IM, Ehrenkranz RA (1991) A regularity statistic for medical data analysis. J Clin Monit 7(4):335-345

42. Stopczynski A, Sekara V, Sapiezynski P, Cuttone A, Madsen MM, Larsen JE, Lehmann S (2014) Measuring large-scale social networks with high resolution. PLoS ONE 9(4):95978